



Denilson Ricardo de Lucena Nunes

Gestão de suprimentos no varejo
Sistemas de reposição de estoques em duas camadas e análise de
esquemas de monitoramento da previsão de demanda

Tese de Doutorado

Tese apresentada ao programa de Pós-graduação em Engenharia de Produção da PUC-Rio como requisito parcial para obtenção do título de Doutor em Engenharia de Produção.

Orientador: Prof. Eugenio Kahn Epprecht

Rio de Janeiro

Abril de 2014



Denilson Ricardo de Lucena Nunes

Gestão de suprimentos no varejo

Sistemas de reposição de estoques em duas camadas e análise de esquemas de monitoramento da previsão de demanda

Tese apresentada como requisito parcial para obtenção do título de Doutor em Engenharia de Produção pelo Programa de Pós-Graduação em Engenharia Produção da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Eugenio Kahn Epprecht

Orientador

Departamento de Engenharia Industrial - PUC-Rio

Prof. Antônio Fernando Castro Vieira

PUC-Rio

Prof. André Cristiano Silva Melo

Universidade do Estado do Pará - UEPA

Prof^a. Fernanda Maria Pereira Raupp

Laboratório Nacional de Computação Científica - LNCC

Prof. Paulo Henrique Soto Costa

UERJ

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico da PUC-Rio

Rio de Janeiro, 11 de abril de 2014

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Denilson Ricardo de Lucena Nunes

Graduou-se em Engenharia Civil pela Universidade da Amazônia –UNAMA, em 2000. Obteve o título de Mestre em Engenharia Civil pela PUC-Rio em 2002. É Professor do curso de Engenharia de Produção da Universidade do Estado do Pará - UEPA desde 2004.

Ficha Catalográfica

Nunes, Denilson Ricardo de Lucena

Gestão de suprimentos no varejo.
Sistemas de reposição de estoques em duas camadas e análise de esquemas de monitoramento da previsão de demanda / Denilson Ricardo de Lucena Nunes ; orientador: Eugenio Kahn Epprecht. – 2014.

95 f. : il. (color.) ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2014.

Inclui bibliografia

CDD: 658.5

Para meus pais, Mário e Maria, minha esposa Emilia e minha filha Izabel,
pela confiança e companhia.

Agradecimentos

À Deus, por se fazer sempre presente em todos os momentos da minha vida.

À minha esposa Emilia, por me apoiar nos momentos difíceis e por compartilhar as alegrias.

À minha filha Izabel, por dar ainda mais significado a minha vida.

À minha mãe Maria e meu pai Mário por me darem os exemplos e os ensinamentos que norteiam o meu caráter.

Aos meus irmãos Glauton e Mayra pelo apoio.

Ao meu primo Bruno por ser meu irmão, meu amigo e por vezes minha força.

Ao meu orientador Eugenio Kahn Epprecht por ter acreditado no meu potencial e ter dado todo o apoio no desenvolvimento desta tese.

Aos amigos que fiz durante a elaboração desta tese, em especial: Felipe Jardim, Orivalde, Augusto, Ana Carla, Beatriz, David e Antônio Ricardo.

À PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Resumo

Nunes, Denilson Ricardo de Lucena; Epprecht, Eugenio Kahn. **Gestão de suprimentos no varejo: Sistemas de reposição de estoques em duas camadas e análise de esquemas de monitoramento da previsão de demanda**. Rio de Janeiro, 2014. 95p. Tese de Doutorado - Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

As empresas do varejo estão constantemente buscando aumentar sua competitividade. Dentre os elementos que influenciam nessa competitividade, esta tese se concentra nos custos com suprimentos e no nível de serviço ao cliente. O nível de serviço depende do planejamento dos estoques, que tem por objetivo balancear a oferta de produtos com a demanda. O ponto de partida desse planejamento são as previsões de demanda. Dessa forma, nesta tese é proposta uma rotina com esquemas de monitoramento do erro dessas previsões utilizando estatísticas denominadas *Tracking Signal*. Essas estatísticas são utilizadas para detectar quando o erro das previsões apresenta comportamento diferente do esperado, indicando assim um viés nas previsões. Para tanto, foi realizado um experimento que resultou na determinação de limites de detecção para as estatísticas *Tracking Signal*, bem como na identificação, dentre elas, daquelas estatísticas mais eficientes na detecção de viés em previsões por amortecimento exponencial simples. Além disso, a redução nos custos com suprimentos e a melhoria no nível de serviço também são contemplados nesta tese pela proposta de solução de um modelo de estoques. Esse modelo considera um item em um sistema com duas camadas e revisão periódica, composto por um Centro de Distribuição (CD) e N lojas no varejo. Nas lojas são aplicadas revisões em intervalos iguais e sincronizados, sendo permitido ao CD entregar parcialmente os pedidos para as lojas em caso de falta. Para esse modelo, foi desenvolvido um algoritmo de busca, baseado no método da seção áurea, que minimiza o custo total no sistema atendendo à meta de nível de serviço no varejo.

Palavras-chave

Varejo; Modelos de estoque; Monitoramento da previsão; Cadeia de suprimentos.

Abstract

Nunes, Denilson Ricardo de Lucena; Epprecht, Eugenio Kahn (advisor). **Supply management in retail: Two-echelon inventory replenishment and analysis of demand forecast monitoring schemes.** Rio de Janeiro, 2014. 95p. D. Sc. Thesis - Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

The retail companies are constantly seeking to increase their competitiveness. Among the factors that influence this competitiveness, this thesis focuses on the cost of supplies and the level of customer service. The level of service depends on the planning of inventories, which aims to balance the supply of products to the demand. The starting point of this planning is the demand forecast. Thus, this thesis proposes a routine of monitoring forecasting errors using statistics called Tracking Signal. These statistics are used to detect when the forecasting error out shows the expected behavior, thus indicating a bias in the forecasts. For this purpose, an experiment was conducted which resulted: determination of detection limits for these statistics; identification those most effective in detecting bias in forecasts by simple exponential smoothing. Moreover, the reduction in costs of supplies and improving the level of service are also covered by this thesis proposed an inventory model. This model considers an item on a system with two echelons and periodic review, consisting of a Warehouse (W) and N retail stores. The retails revisions apply in equal and synchronized interval; it is possible the W partially deliver the orders to the stores in case of shortages. For this model, a search algorithm based on the golden section method, which minimizes the total cost in the system given the target level of service in retail was developed.

Keywords

Retail; Inventory models; Monitoring of forecasting; Supply Chain

Sumário

1	Introdução	12
1.1.	Objetivos	15
1.2.	Contribuições Originais	16
1.3.	Organização do texto	17
1.4.	Uso de software de Programação	19
2.	Previsão de demanda e sistemas de controle de estoques	21
2.1.	Conceitos básicos de previsão de demanda	21
2.2.	Previsão da demanda	24
2.2.1.	Modelo constante de demanda e amortecimento Exponencial Simples	24
2.3.	Monitoramento da previsão	26
2.3.1.	Limites de detecção	29
2.3.2.	Medidas de desempenho	32
2.3.3.	Diferentes interpretações para NMA0	33
2.3.4	Outras medidas de desempenho	35
2.3.5.	Críticas ao uso de <i>TST</i> e <i>TSB</i> e oportunidades de contribuição da Pesquisa	35
2.4.	Sistemas de Controle de Estoques	36
2.4.1.	Modelos de estoque em camadas	38
2.4.2.	Hipóteses e a física do problema	40
2.4.3.	Sistemas com duas camadas e política (T, S)	41
2.4.4.	Técnica de desagregação do sistema	42
3.	Metodologia de monitoramento da previsão	45
3.1.	Variantes para as estatísticas <i>Tracking Signal</i>	45
3.2.	Experimento de simulação e determinação dos limites de detecção	46
3.3.	Apresentação dos resultados	47

4. Modelo de estoque proposto	53
4.1. Descrição do modelo	53
4.2. Notação	54
4.3. Determinação do estoque médio no CD e no varejo	55
4.4. Alocação das faltas no varejo	56
4.5. <i>Lead time</i> efetivo no varejo e desagregação do sistema	58
4.6. Formulação para demanda normal	61
4.6.1. Determinação do nível de estoque para demanda normal	61
4.6.2. Função de custo e o problema de otimização	63
4.6.3. Algoritmo de otimização	65
4.6.4 Ilustração numérica e simulação	66
5. Conclusões	70
6. Referências Bibliográficas	72
APÊNDICE A – Métodos de busca dos limites de detecção Correspondentes a um determinado NMA0	77
APÊNDICE B – Limites iniciais de busca para S_0 no Algoritmo da Seção 4.6.3.	78
APÊNDICE C – Rotinas e funções programadas em MATLAB	80

Lista de Figuras

Figura 1.1 - Vendas anuais nos Estados Unidos	12
Figura 1.2 - Receita líquida do varejo no Brasil	13
Figura 1.3 - Participação do varejo no PIB brasileiro	13
Figura 1.4 - Representação de uma rede de suprimentos	14
Figura 2.1- Estrutura de um sistema de previsão	22
Figura 2.2 – Sistema centralizado com duas camadas	40
Figura 3.1 – comparação de desempenhos com o $e_{padrão}$	49
Figura 3.2 – NMA1 para variações Δa e $\alpha\{0,045,0,135,0,21,0,3\}$	51
Figura 3.3 – Estatísticas com melhor desempenho para cada α	51
Figura 4.1 – Nível de estoque no CD e no varejo i para $T_0 = 2T$	54
Figura 4.3 – Valores de S_1, S_2, S_3 e c para $S_0 \in \{1,2,\dots,729\}$	68

“Todo o trabalho é vazio a não ser que haja amor”

Khalil Gibran

1. Introdução

O conceito de varejo engloba as atividades de comércio e serviço, que por serem atividades distintas dificultam a elaboração de uma definição. Para o IBGE o varejo abrange, além do comércio e serviço, as atividades de transportes, comunicações, serviços pessoais e auxiliares, atividades financeiras e governamentais. Kotler & Armstrong (1993) vão além e afirmam que qualquer um que venda bens ou serviços ao consumidor final está praticando atividades de varejo, mesmo que seja fabricante ou atacadista.

Do ponto de vista econômico, o varejo é um dos maiores setores da economia mundial e encontra-se em crescimento. Nos Estados Unidos, por exemplo, as vendas somaram \$2,34 trilhões em 1995, excluindo a venda direta, e naquele mesmo ano quase 20% da força de trabalho norte-americana estava concentrada no varejo (Levy e Weitz, 2000). A tendência de crescimento apontada por esses autores pode ser observada nas vendas do varejo nos Estados Unidos nos anos de 1995 - 2010 (ver Figura 1.1).



Figura 1.1 - Vendas anuais nos Estados Unidos

Fonte: U.S. Department of Commerce (2011)

Esses números são significativos para a economia norte-americana, pois nos anos 2007-2010 as receitas no varejo contribuíram, em média, com mais de 26% do produto interno bruto (BEA, 2011).

No Brasil, em 1998, o setor apresentou mais de R\$100 bilhões em vendas, o que correspondeu a mais de 10% do PIB daquele ano (Parente, 2000). Nos

últimos anos esse setor apresentou crescimento nas vendas e na participação no PIB brasileiro (Figuras 1.2 e 1.3). Ao final de 2007, aproximadamente 18% da população considerada ocupada estava concentrada no varejo (IBGE, 2008b).

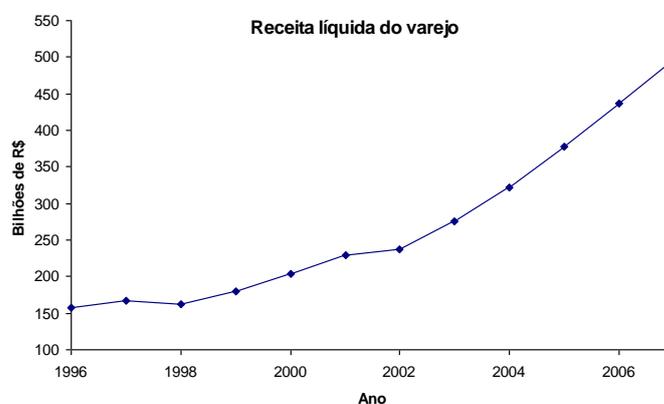


Figura 1.2 - Receita líquida do varejo no Brasil

Fonte: IBGE (2011)

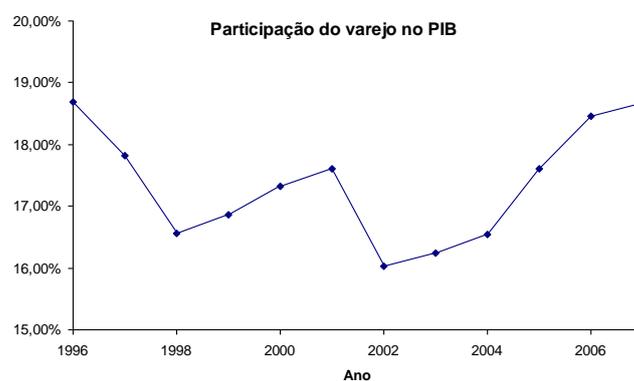


Figura 1.3 - Participação do varejo no PIB brasileiro

Fonte: IBGE (2011)

As atividades de varejo se caracterizam principalmente pela relação direta de venda de bens e serviços ao consumidor final. Isso é consequência de o varejo ser a última camada ou nível em uma rede de suprimentos. Segundo Ballou (2004), a cadeia de suprimentos abrange todas as atividades (empresas) envolvidas no processo de transformação da matéria-prima até seu usuário final.

Pode-se ainda dizer que, em geral, cadeias de suprimento são sistemas com múltiplos níveis, que têm como objetivo levar bens acabados a seus consumidores finais (ver Figura 1.4) e se caracterizam pelo fluxo de material e informação através de instalações dispostas em níveis ou camadas. Nas

instalações podem ocorrer diferentes estágios de um processo produtivo de maneira que, para cada estágio, por meio de montagem ou outro processo de fabricação, o produto ganha valor agregado até ser finalizado e entregue ao consumidor.

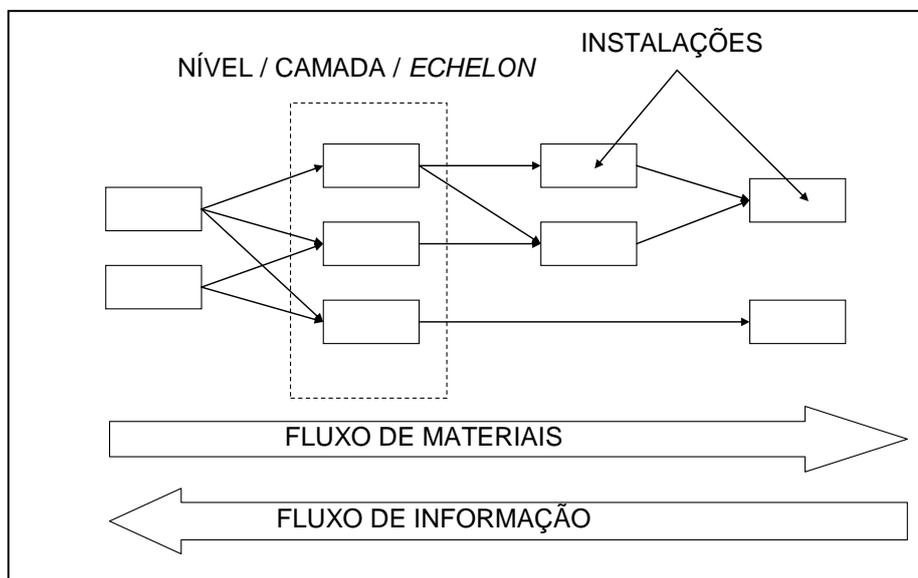


Figura 1.4 - Representação de uma rede de suprimentos

A concorrência entre as empresas do varejo ocorre principalmente através de seu “mix” de produtos, selecionado de acordo com seu público alvo. Levy e Weitz (2000) ressaltaram que é importante a estratégia de competição ter uma base sustentável, e nesse contexto, a fidelização do cliente é um bom exemplo disso. Vários fatores podem influenciar no processo de fidelização. A disponibilidade do produto na prateleira é um dos fatores mais importantes para conquistar o cliente (Levy e Weitz, 2000). Parente (2000) concorda com essa afirmação, e acrescenta que a insatisfação do cliente é certa quando o produto que compra rotineiramente não está disponível para pronta entrega.

Segundo ECRBRASIL (2008), diante da falta de diferentes produtos na prateleira existe 56% de chance do consumidor trocar de loja ou deixar de levar um produto substituto. Caso o cliente não encontre cerca de 20% dos itens em sua lista, o mesmo pode desistir de toda a compra. Esse inconveniente somente é tolerado no máximo duas vezes, sob o risco de o cliente não retornar mais ao estabelecimento (ECRBRASIL, 2008). Portanto, a disponibilidade para pronta entrega de qualquer produto é importante na estratégia de fidelização.

Na tentativa de melhorar essa disponibilidade de produtos, redes de varejo vêm recorrendo a centros de distribuição (CD) próprios ou terceirizados, ou seja,

o próprio varejo assume a responsabilidade de entregar os produtos nas lojas. As entregas são programadas com base nas informações de demanda e estoque das lojas passadas ao CD por meio eletrônico. Em muitos casos, para diminuir a probabilidade de falta de produtos, bem como reduzir os estoques nas lojas, as entregas são mais frequentes do que se fossem feitas diretamente pelo fornecedor.

A falta de um produto na prateleira é conhecida como ruptura e apesar dos esforços do varejo ainda são recorrentes. Segundo Rosa *et. al.* (2007), a falta de produtos na prateleira pode implicar em vendas perdidas e afetar a confiabilidade do cliente no estabelecimento, o que, em longo prazo, implica em perda de competitividade.

1.1. Objetivos

Segundo a Associação *Efficient Consumer Response* do Brasil (ECRBRASIL), um dos principais desafios do varejo é evitar a ruptura, que resulta de muitos fatores, por exemplo: falta de reposição na prateleira a partir do estoque da loja, erro na definição da frequência e quantidades de reposição do estoque, atrasos na entrega por parte do CD, atraso na negociação com fornecedores, etc (ECRBRASIL, 2008).

Os níveis de reposição dos estoques no CD e nas lojas influenciam a ocorrência das rupturas e os custos relacionados com os estoques. Definir esses níveis de reposição é um dos objetivos desta tese, o segundo objetivo é fornecer uma rotina para o monitoramento dos erros das previsões de demanda. Essas duas vertentes convergem para soluções mais eficientes na redução das rupturas no varejo e dos custos com estoques.

Portanto, os principais objetivos desta tese são: **propor um método heurístico para definir os níveis de reposição de estoques nas lojas e no CD de itens comercializados no varejo, que resultem no nível de serviço desejado nas lojas ao menor custo com estoques; e propor um sistema de monitoramento do erro de previsão, capaz de detectar um eventual viés na previsão de demanda.**

Esses objetivos podem ser desdobrados nos seguintes objetivos secundários:

- Propor um modelo de estoque que permita definir as quantidades de reposição do estoque, em um sistema com duas camadas (CD e lojas) e revisão periódica;
- Desenvolver algoritmo de otimização para o modelo de estoque proposto, visando um nível de satisfação da demanda nas lojas ao menor custo;
- Propor uma metodologia de monitoramento do erro da previsão que permita a automatização e seja eficiente na detecção de vieses na previsão;
- Apresentar uma metodologia de previsão de demanda;
- Estudar e compreender os aspectos mais importantes dos modelos de inventário voltados a sistemas com múltiplos níveis;
- Realizar uma revisão bibliográfica sobre monitoramento de erro da previsão.

1.2. Contribuições Originais

Uma vez atingidos os objetivos desta tese, podem ser ressaltadas suas contribuições originais nas duas áreas de conhecimento, as quais ela se concentra: modelos de estoque em camadas e monitoramento das previsões de demanda.

Apesar da vasta literatura sobre modelos de estoque em camadas, são encontradas poucas soluções exatas para esses modelos. No caso de sistemas com reposição contínua dos estoques, por exemplo, algumas referências são: Graves (1985), Chen e Zheng (1997), Cheung e Hausman (2000) e Axsäter (2000). Quando se trata de reposição periódica, essas soluções são ainda mais raras, como por exemplo: Chu e Shen (2010), Cachon (2001) e Liljenberg (1996). O método heurístico aqui proposto é direcionado a um sistema de estoque com duas camadas e reposição periódica dos estoques, com as seguintes características: intervalos entre reposições iguais e sincronizadas nas lojas; atendimento parcial dos pedidos das lojas ao CD; as demandas nas lojas não são necessariamente iguais, podendo seguir uma distribuição normal, gama ou Poisson. Segundo a pesquisa realizada nesta tese, estudos anteriores (Heijden,1992; Heijden, 2000; Wang 2013) não consideraram essas características de forma conjunta nesse tipo de sistema, diferente do que é proposto nesta tese caracterizando assim a contribuição.

No que tange o monitoramento das previsões de demanda, esta tese considerou o monitoramento feito com o uso das estatísticas *Tracking Signal*, que são calculadas a partir do erro das previsões de demanda. Foram estudadas

quatro estatísticas diferentes propostas por Brown (1962), Trigg (1964) e Gardner (1983). Ainda que essas estatísticas tenham sido amplamente estudadas até a década de 80, como pode ser observado em Batty (1969), McKenzie (1978), Gardner (1983, 1985) e (McClain, 1988), algumas questões ainda permaneceram, como por exemplo, qual a melhor metodologia na determinação dos limites de detecção, ou mesmo qual o melhor esquema de monitoramento que produz uma detecção de viés mais eficiente, além disso, qual dessas estatísticas tem o melhor desempenho na detecção. Essas questões continuam sendo investigadas como em Cohen *et. al* (2009). No desenvolvimento desta tese foi realizado um número de experimentos maior do que aqueles aplicados em trabalhos anteriores, o que resultou nas seguintes contribuições: metodologia de obtenção dos limites de detecção dessas estatísticas, para o monitoramento do erro das previsões por amortecimento exponencial simples; e mapeamento das situações em que cada uma das estatísticas estudadas apresentou melhor desempenho na detecção de viés nas previsões.

1.3. Organização do texto

O Capítulo 2 se inicia com a descrição de uma estrutura de um sistema de previsão de demanda. Em seguida, é apresentado o método de amortecimento exponencial simples e o modelo constante de demanda.

Conforme descrito no Capítulo 2, no monitoramento das previsões são utilizadas estatísticas *Tracking Signal* calculadas a partir dos erros das previsões. No centro da discussão sobre monitoramento com essas estatísticas está a determinação dos limites de detecção. Esse capítulo contém breve discussão sobre a metodologia usada em pesquisas anteriores para calcular esses limites. Além disso, outro elemento discutido nesse capítulo é a medida de desempenho NMA1 (número médio de períodos até um alarme verdadeiro), que é usada para avaliar a eficiência na detecção de um viés. São então apresentadas diferentes interpretações e formas de obtenção dessa medida.

Ainda no Capítulo 2 são descritos os quatro principais sistemas de controle de estoque, incluindo um comentário sobre como os custos são considerados na modelagem de estoques. A partir deste ponto são discutidos apenas os modelos voltados para os sistemas de estoques com duas camadas e revisão periódica, destacando aqueles que se aproximam do sistema estudado nesta tese.

O Capítulo 3 é dedicado a apresentar a metodologia proposta para determinar os limites de controle das estatísticas *Tracking Signal*. Inicialmente são apresentadas duas variantes das estatísticas mais conhecidas (descritas no Capítulo 2), que também foram testadas para avaliação de desempenho. Em seguida, são descritos os detalhes do experimento de simulação para a determinação dos limites de detecção para todas as estatísticas consideradas. Além disso, são descritos os detalhes do experimento de simulação para avaliar o desempenho dessas estatísticas na detecção do viés.

Dessa forma, quatro estatísticas foram testadas, o que resultou na identificação das mais eficientes e nas situações em que isso ocorre. Esses resultados estão reunidos em tabelas ao final do Capítulo 3, nas quais cada limite de detecção está relacionado a um valor do parâmetro α , usado no amortecimento exponencial simples.

No Capítulo 4 é apresentada a proposta de modelo de estoque para um sistema com duas camadas e revisão periódica. O sistema representa uma situação típica encontrada no varejo (um CD que abastece um conjunto de lojas). Esse capítulo se inicia com a descrição do modelo e da notação usada. Em seguida é apresentada a formulação para o cálculo dos estoques médios no CD e nas lojas, bem como o processo de alocação das faltas do CD nas lojas do varejo. Essas faltas são utilizadas na determinação do *lead time* efetivo, que permite tratar o sistema como um conjunto de instalações isoladas.

Ainda no capítulo 4, é apresentada a formulação para o modelo proposto em uma situação de demanda normal no varejo e no CD, seguida da descrição do algoritmo de otimização do modelo. Esse capítulo contém ainda uma ilustração numérica para um sistema com um CD que abastece três lojas. Para esse mesmo sistema, foi realizado um experimento de simulação a fim de observar a eficácia do modelo proposto. Como resultado, o nível de serviço no varejo foi satisfatório, sobretudo quando as variâncias das demandas nas lojas do varejo são similares. Porém, para variâncias muito diferentes, aquelas lojas com maior variância na demanda tendem a ficar abaixo da meta de nível de serviço, por outro lado, instalações com variância menor têm suas metas superadas.

As conclusões e propostas para trabalhos futuros estão reunidas no Capítulo 5. Ao final do texto são encontrados os apêndices, os quais contêm: o método de busca para os limites de detecção (APÊNDICE A), os limites iniciais de busca para o algoritmo proposto (APÊNDICE B) e as rotinas usadas na elaboração desta tese em linguagem MATLAB (APÊNDICE C).

1.4. Uso de software de Programação

O software escolhido foi o MATLAB em virtude de apresentar uma linguagem que permite a elaboração de funções personalizadas pelo usuário e por conter funções de geração de números aleatórios, que são nativas do programa. No Anexo, estão reunidas todas as rotinas utilizadas na elaboração desta tese juntamente com uma breve descrição do funcionamento de cada uma delas.

A metodologia proposta no Capítulo 3 envolve a obtenção de uma medida conhecida como número médio de períodos até um alarme (NMA), tanto para a obtenção dos limites de detecção das estatísticas *Tracking Signal*, quanto na avaliação de seus desempenhos. Para a obtenção dessa medida foram elaboradas quatro rotinas (uma para cada estatística estudada) capazes de calcular o NMA em função dos parâmetros do modelo constante, do parâmetro do amortecimento exponencial simples, e do parâmetro usado em cada estatística. Essas rotinas são denominadas: `funcTsT`; `funcTsT1`; `funcTsB`; `funcTsB1`.

Ainda segundo o Capítulo 3, a obtenção dos limites de detecção é realizada por um algoritmo de busca. Esse algoritmo foi programado em quatro rotinas chamadas: `findLTsT`; `findLTsT1`; `findLTsB`; `findLTsB1`. O algoritmo testa iterativamente as funções do tipo `funcTs*` até encontrar o limite de detecção que produz o NMA0 (número médio de períodos até um alarme falso) definido. A avaliação do desempenho na detecção do viés foi realizada a partir do NMA1 (número médio de períodos até um alarme verdadeiro), que foi calculado com as funções do tipo `funcTs*`, utilizando os limites de detecção encontrados.

A partir da formulação do modelo de estoque para um sistema com duas camadas (Capítulo 4) e do algoritmo de otimização proposto para esse modelo, foi elaborada uma rotina chamada `SdoisNiveis`. Essa rotina monta a função de custo do modelo e busca iterativamente pelos níveis de reposição no CD e nas lojas que minimizam o custo a partir dos dados de entrada: dados de custos, média e variância da demanda nas lojas, *Lead times*, intervalos de revisão nas lojas e no CD e metas de níveis de serviço nas lojas.

Para observar a eficácia da solução heurística proposta, foi elaborada uma rotina chamada `simulaEstoque`. Essa rotina simula a movimentação do estoque nas lojas e no CD, para o sistema descrito na ilustração numérica no Capítulo 4, com o objetivo de calcular o nível de serviço observado em cada loja ao final da

simulação. Nesse caso, os níveis de reposição do estoque são definidos pela rotina SdoisNiveis.

2. Previsão de Demanda e Sistemas de Controle de Estoques

Este capítulo é dedicado à descrição do sistema de previsão de demanda, ao monitoramento do viés dessa previsão, aos principais sistemas de controle de estoque e à modelagem de estoque em camadas. O capítulo se inicia com algumas definições e descrição das fases de um processo de previsão de demanda para o modelo constante. Neste capítulo, se encontra também breve discussão sobre a metodologia usada em pesquisas anteriores para calcular os limites de detecção das estatísticas *Tracking Signal*, bem como uma discussão sobre a medida de desempenho usada para avaliar a eficiência dessas estatísticas na detecção de um viés.

Neste capítulo são ainda descritos os principais sistemas de controle de estoque e discutidos importantes aspectos da modelagem de estoque em camadas. São apresentados os quatro principais sistemas de controle de estoque e discutidas as principais características da modelagem de estoque em camadas. Serão apresentados apenas os modelos voltados a sistemas de estoques com duas camadas e revisão periódica, destacando aqueles que se assemelham ao sistema estudado nesta tese. Ao final deste capítulo será apresentada uma técnica usada na solução de modelos de estoque em camadas, que consiste em desagregar o sistema em instalações isoladas.

2.1. Conceitos básicos de previsão de demanda

As previsões de demanda são estimativas do volume que será requisitado de bens ou serviços no futuro. Essas estimativas são o ponto de partida para o planejamento das funções marketing, finanças e operações, de maneira que todas essas são prejudicadas quando as previsões divergem muito da realidade. Do ponto de vista da logística de suprimento, previsões discrepantes da realidade resultam em clientes que não são atendidos no momento ou na quantidade desejados, ou em estoques em excesso.

Torna-se necessário então um sistema de previsão adequado para melhorar o nível de serviço aos usuários, para planejar o sistema de reposição de estoque e identificar a necessidade de uma maior capacidade de produção.

Na Figura 2.1 (adaptada de Silver *et al.*, 1998) encontra-se uma sugestão de uma estrutura de um sistema de previsão.

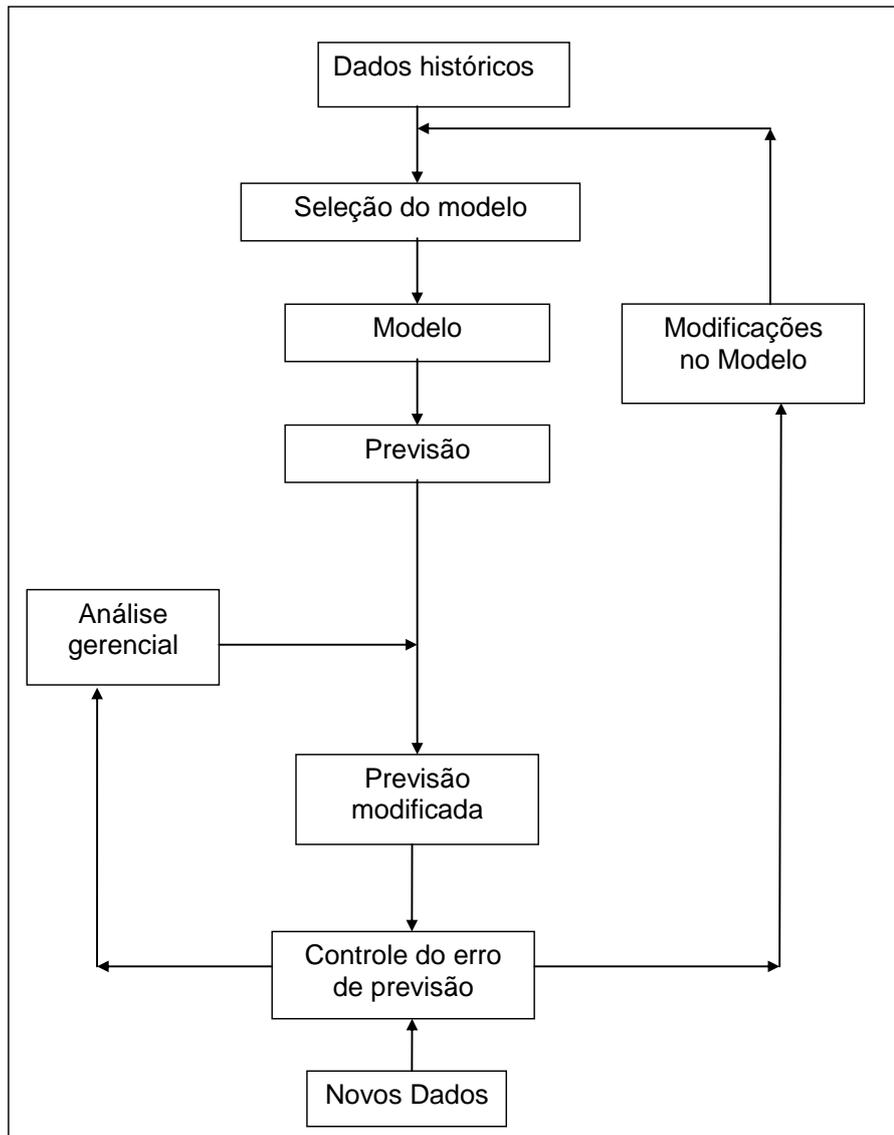


Figura 2.1- Estrutura de um sistema de previsão.

Nessa estrutura é possível uma intervenção no processo de previsão após análise gerencial do resultado fornecido pelo modelo. Essa análise é particularmente importante quando ocorrem fatos que podem interferir na demanda (novas firmas competidoras, firmas que saíram do mercado, promoções etc) que não existiam, e, portanto, não estavam implícitos nos dados passados.

Medir o erro de previsão é uma importante fase do sistema de previsão, pois serve para dimensionar os estoques de segurança, para avaliar se o modelo é adequado, fornecer previsões sem viés e para avaliar se as modificações adicionadas na análise gerencial são pertinentes.

Portanto, os objetivos de um sistema de previsão são: estimar o valor esperado da demanda; prever com antecedência suficiente para a tomada de decisão; atualizar as estimativas periodicamente; permitir a análise gerencial.

Segundo Ballou (2004), a vasta opção de técnicas para estimar o valor esperado da demanda pode ser agrupada em métodos qualitativos, métodos de séries temporais e modelos causais. Os métodos de previsão baseados em técnicas de séries temporais são de especial interesse para esta tese e, segundo Ballou (2004) e Bowersox et al. (2002), são indicados para demandas com comportamento relativamente estável, apresentando componentes de tendência e/ou sazonalidade bem definidos ao longo do tempo. Ainda de acordo com esses autores, a estabilidade do comportamento da demanda permite que esses métodos sejam eficientes, sobretudo para previsões de curto prazo. Nessa circunstância, mudanças no padrão de demanda são muito graduais.

Uma série temporal é um conjunto de valores gerados por um processo estocástico indexado no tempo. O processo estocástico é um mecanismo gerador de uma variável aleatória que não pode ser descrita por uma função determinística, ou seja, a cada passo de tempo a série temporal apresenta um único valor (ou estado), de maneira que antes de sua realização seu estado é uma variável aleatória.

Um processo estocástico pode ser classificado como contínuo ou discreto quanto ao seu espaço paramétrico (Morettin e Tolo, 2006). No caso de séries temporais, o espaço paramétrico diz respeito a maneira como o tempo é considerado, contínuo ou discreto. Um processo estocástico pode ainda ser contínuo ou discreto quanto ao espaço dos estados, que para uma série temporal representa o conjunto de valores apresentados a cada marca de tempo. Portanto, a demanda por bens e serviços ao longo do tempo pode ser interpretada como uma série temporal.

As técnicas de previsão de séries temporais são de interesse para esta tese por serem utilizadas nas previsões de curto prazo, que servem de base para as decisões relacionadas ao controle de estoques.

2.2. Previsão da demanda

Em geral, a expressão “modelo de previsão” está associada apenas ao modelo do processo gerador da demanda (ou modelo paramétrico). Métodos de séries temporais podem ser utilizados para estimar os parâmetros desse modelo partindo da premissa que o comportamento futuro da demanda reflete o comportamento passado. Logo, é essencial a disponibilidade de um histórico de dados para sua utilização.

Eventuais mudanças no comportamento da demanda podem comprometer a eficácia do método de previsão adotado. Essa perda de adequação do método de previsão é percebida no erro das previsões. No caso de previsão em séries temporais existem técnicas de monitoramento do erro dessas previsões que auxiliam na identificação de uma mudança estrutural na demanda.

Os modelos chamados estruturais assumem que uma série temporal é gerada pela combinação de um componente sistemático somado a um componente aleatório. O componente sistemático pode ser estimado por diferentes técnicas com base estatística. Em geral, para previsão da demanda de curto prazo são considerados três modelos: Modelo Constante - é o modelo em que não se observa tendência de crescimento ou decrescimento na demanda, a qual varia em torno de um nível; Modelo Linear - é o modelo em que se observa tendência de crescimento ou decrescimento na demanda; Modelo Sazonal - é o modelo em que, com ou sem tendência de crescimento ou decrescimento na demanda, se observam variações cíclicas na demanda.

2.2.1. Modelo constante de demanda e amortecimento exponencial simples

Conforme citado, o modelo constante é aquele em que não se observa tendência de crescimento ou decrescimento na demanda. Supondo que a demanda x_t é gerada pelo modelo constante, tem-se que

$$x_t = a + \varepsilon_t, \quad (2.1)$$

onde a é o componente sistemático independente do tempo, que representa a demanda média. Assume-se que o componente aleatório ε_t , também denominado ruído, tem distribuição normal com média zero e variância σ^2 .

Assim, a previsão pontual feita em um instante t para um instante futuro $(t + \tau)$ qualquer é a estimativa do valor esperado da demanda (parte sistemática) nesse instante futuro $t + \tau$. No caso do modelo de nível constante, a previsão pontual é, portanto, a estimativa do nível a no instante futuro $t + \tau$ em questão. Essa estimativa de a é a mais atual de que se dispõe no instante t em que a previsão é feita. Formalmente:

$$\hat{x}_{t+\tau} = \hat{a}_t. \quad (2.2)$$

Uma das formas de calcular essa estimativa é através do método de amortecimento exponencial simples. Segundo Brown (1962), a estimativa \hat{a}_t pode ser obtida ao fim de cada período t por

$$\hat{a}_t = \alpha x_t + (1 - \alpha)\hat{a}_{t-1}, \quad (2.3)$$

sendo, α o parâmetro de amortecimento, no intervalo $0 \leq \alpha \leq 1$. A estimativa inicial \hat{a}_0 é obtida a partir da média aritmética de uma amostra retirada dos dados da série temporal.

Observa-se que se $\alpha = 1$, então a estimativa do nível de a a cada instante é igual ao valor da última observação; se $\alpha = 0$, a estimativa do nível permanece constante e igual ao seu valor inicial \hat{a}_0 . Quanto maior α , maior é o peso que a última observação tem na revisão da estimativa. Mais adiante será discutido como escolher o melhor valor para α .

A Equação (2.3) pode ser reescrita considerando que são disponíveis t observações da demanda: x_1, x_2, \dots, x_t , e que o processo de previsão foi iniciado na ocasião da observação x_1 . Então, a previsão feita ao fim do período t , quando x_t foi observado, é:

$$\hat{a}_t = \alpha x_t + (1 - \alpha)\hat{a}_{t-1}$$

Mas, $\hat{a}_{t-1} = \alpha x_{t-1} + (1 - \alpha)\hat{a}_{t-2}$ e substituindo \hat{a}_{t-1} na equação acima resulta em:

$$\hat{a}_t = \alpha x_t + (1 - \alpha)[\alpha x_{t-1} + (1 - \alpha)\hat{a}_{t-2}] = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2 \hat{a}_{t-2}$$

Mas, $\hat{a}_{t-2} = \alpha x_{t-2} + (1 - \alpha)\hat{a}_{t-3}$ e substituindo \hat{a}_{t-2} na equação acima resulta em:

$$\begin{aligned}\hat{a}_t &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2[\alpha x_{t-2} + (1-\alpha)\hat{a}_{t-3}] \\ &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + (1-\alpha)^3 \hat{a}_{t-3}\end{aligned}$$

Continuando com as substituições sucessivas de $\hat{a}_{t-3}, \hat{a}_{t-4}, \dots, \hat{a}_1$, encontra-se a seguinte expressão:

$$\begin{aligned}\hat{a}_t &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \dots + \alpha(1-\alpha)^{t-1} \hat{a}_1 + (1-\alpha)^t \hat{a}_0 \\ \hat{a}_t &= \sum_{j=0}^{t-1} \alpha(1-\alpha)^j x_{t-j} + (1-\alpha)^t \hat{a}_0.\end{aligned}\tag{2.4}$$

Considerando que, em situações reais, quase sempre $0,01 \leq \alpha \leq 0,3$, temos que o termo $(1-\alpha)^t \hat{a}_0$ é aproximadamente igual a zero, para valores moderados de t . Logo, temos que:

$$\hat{a}_t = \sum_{j=0}^{t-1} \alpha(1-\alpha)^j x_{t-j}\tag{2.5}$$

Portanto, \hat{a}_t é uma média ponderada das demandas anteriores. O peso da observação x_{t-j} é $\alpha(1-\alpha)^j$.

Tomando o valor esperado de \hat{a}_t , pode-se demonstrar que $E(\hat{a}_t) = E(x_t)$, o que significa que a estimativa \hat{a}_t para a demanda é sem viés, uma vez que o valor esperado da previsão utilizada (\hat{a}_t) é igual ao valor esperado da variável que está estimando (x_t).

2.3. Monitoramento da previsão

Devido à aleatoriedade natural da demanda, toda previsão está sujeita a erro. Define-se o erro de previsão no período t como a diferença entre a demanda observada x_t e a estimativa de demanda para esse período, ou seja,

$$e_t = x_t - \hat{x}_{t-1,t},\tag{2.7}$$

sendo $\hat{x}_{t-1,t}$ a estimativa da demanda (previsão) realizada no período $t-1$ para o período t .

Quando se utiliza um modelo de previsão adequado, os valores desse erro oscilam em torno do zero. Mais rigorosamente, seu valor esperado ($E(e_t)$) é nulo. Contudo, quando o modelo de previsão deixa de ser aderente à realidade, as previsões começam a ficar enviesadas, tendendo a subestimar, ou a superestimar, a demanda. Assim, o valor esperado do erro de previsão torna-se diferente de zero, indicando a presença de viés. Nesse caso, a série que se acredita representar a demanda ou o método de previsão utilizado (ou os parâmetros usados nos mesmos) devem ser revistos. Existe vasta literatura sobre modelos de previsão de demanda, porém não se pode dizer o mesmo sobre técnicas formais para monitorar a existência de viés conforme as previsões vão sendo realizadas.

Como a inadequação do modelo resulta em $E(e_t) \neq 0$, as técnicas de monitoramento de previsão de demanda são baseadas no monitoramento das observações dos erros de previsão coletadas a cada instante de tempo; mais precisamente, de estatísticas dos mesmos destinadas a testar a hipótese nula $E(e_t) = 0$ contra a alternativa $E(e_t) \neq 0$.

Brown (1962) forneceu uma das primeiras propostas de monitoramento de modelos de previsão, que consiste em computar a cada período uma estatística conhecida como *Tracking Signal*, aqui denotada por TSB . Essa estatística é definida, no período t , como o módulo do quociente entre a soma cumulativa do erro EAC_t e o desvio absoluto médio DAM_t , ou seja,

$$TSB_t = \left| \frac{EAC_t}{DAM_t} \right|, \quad (2.8)$$

onde $EAC_t = e_t + EAC_{t-1}$ com valor inicial $EAC_0 = 0$; e DAM_t é a estimativa do desvio absoluto médio, atualizada por amortecimento exponencial simples pela fórmula recursiva $DAM_t = \alpha_1 |e_t| + (1 - \alpha_1) DAM_{t-1}$ onde $0 < \alpha_1 < 1$ é uma constante de amortecimento escolhida pelo usuário do método.

Como valor inicial DAM_0 para iniciar o cálculo recursivo, Brown (1962), bem como Gardner (1983, 1985), em seus trabalhos teóricos, usaram o valor

esperado de $|e_t|$, calculado em função do desvio-padrão do componente aleatório da série temporal, assumido conhecido num contexto teórico de análise de desempenho. No contexto prático, porém, em que não é conhecido o valor desse desvio-padrão, tornasse necessário estimá-lo a partir de dados históricos.

Assim, o monitoramento é realizado da seguinte forma: a cada período t a estatística TSB é processada; quando seu valor ultrapassa um determinado limite de controle, isso é considerado um sinal de existência de viés na previsão. Essa solução se mostrou de fácil aplicação, além de não exigir a manipulação de muitos dados, uma vez que na Equação (2.8) são consideradas apenas duas informações do período anterior: EAC_{t-1} e DAM_{t-1} , que são atualizados recursivamente.

Embasado nessa solução, Trigg (1964) propôs substituir, na Equação (2.8), a soma cumulativa do erro EAC_t pelo erro amortecido exponencialmente EAM_t , ou seja,

$$TST_t = \left| \frac{EAM_t}{DAM_t} \right|, \quad (2.9)$$

sendo o EAM atualizado pela fórmula recursiva $EAM_t = \alpha_1 e_t + (1 - \alpha_1)EAM_{t-1}$, onde, por sua vez, o valor inicial EAM_0 é o valor esperado do erro de previsão sob a hipótese nula (modelo adequado), ou seja, $EAM_0 = 0$.

Trigg (1964) afirma que a estatística TST tem variância menor que TSB e por esse motivo é mais eficiente na detecção de viés.

Computar a estatística TSB (ou TST) e compará-la com um determinado limite de detecção é equivalente a fazer uso de um gráfico de controle, que surgiu no Controle Estatístico de Processos (CEP). A aplicação de gráficos de controle na detecção de viés tem uma diferença essencial: em CEP as variáveis monitoradas têm distribuição conhecida, o que facilita a determinação dos limites de detecção, no caso chamados de limites de controle. Todavia as estatísticas TSB e TST têm distribuição conhecida apenas em algumas situações. Nos demais casos em que essa distribuição não é conhecida, os limites de detecção para TSB e TST são obtidos por simulação.

Segundo o corpo editorial da revista *International Journal of Forecasting* (Forecasting, Int. J. 2009), não houve publicações sobre o monitoramento das

previsões desde a década de 80, até o final dos anos 2000, quando alguns artigos foram publicados. Alguns desses estudos (Ord et. al, 2009; Neill, 2009) propuseram que, ao invés de se concentrar no erro das previsões, o monitoramento deve ter o foco sobre o comportamento da série temporal. Nesse caso, uma vez identificada uma mudança nesse comportamento o processo de previsão deve ser avaliado.

As estatísticas *Tracking Signal* foram o objeto de estudo de Cohen et. al (2009) que propuseram uma alternativa para determinar os limites de detecção (será discutida na próxima seção). Snyder e Koehler (2009) também voltaram sua atenção para essas estatísticas e propuseram um modelo de previsão que incorpora o erro amortecido com o intuito de corrigir automaticamente as previsões. Esses autores sugerem que, nesse caso, o monitoramento seria dispensado, pois a correção das previsões seria feita de forma automática.

Desde então, novas pesquisas vêm surgindo, como por exemplo, Krol (2013) que questionou o conceito de viés em modelos econométricos, afirmando que nesses modelos o erro da previsão pode ter média diferente de zero, mesmo quando não há viés. Ainda que as estatísticas *Tracking Signal* tenham sido amplamente estudadas até a década de 80, algumas questões ainda permaneceram, por exemplo, qual a melhor metodologia na determinação dos limites de detecção, ou mesmo qual o par (limite, α_1) que produz uma detecção de viés mais eficiente, além disso, qual dessas estatísticas tem o melhor desempenho na detecção.

A seguir são discutidas diferentes formas de obtenção dos limites de controle para as estatísticas *TST* e *TSB*, bem como os critérios para comparação de desempenho.

2.3.1. Limites de detecção

No centro da discussão sobre monitoramento da previsão com as estatísticas *Tracking Signal* está a determinação dos limites de detecção. Nas situações em que se consideram conhecidas as distribuições de *TST* e *TSB*, os limiares são obtidos como os limites de controle em CEP, ou seja, são os valores dessas estatísticas que correspondem a uma probabilidade pré-fixada de Erro do Tipo I. Entretanto, essas estatísticas têm distribuição conhecida apenas em algumas situações.

Segundo Brown (1962) a estatística TSB , quando obtida a partir dos erros do amortecimento exponencial simples, apresenta distribuição aproximadamente normal para $\alpha_1 < 0,1$. Gardner (1983) chegou à mesma conclusão. Para esses casos, Brown (1962) fornece uma expressão (em função de α_1) para calcular o desvio padrão de TSB e sugere que os limites de detecção sejam quatro vezes esse desvio.

A estatística TST , segundo Trigg (1964), também apresenta distribuição normal, desde que $\alpha_1 \leq 0,1$ e que os erros de previsão sejam independentes e normalmente distribuídos. Para essas condições, esse autor apresenta uma expressão para calcular os limites de TST em função de α_1 . Esses limites não são aplicados a erros de previsão por amortecimento exponencial pelo fato desses erros apresentarem autocorrelação.

Batty (1969) e McKenzie (1978) consideraram a autocorrelação e observaram que para $\alpha_1 < 0,1$ a hipótese de normalidade de TST se mantém. Baseados nessa hipótese, esses autores desenvolveram expressões para estimar a variância de TST em função de α_1 . A formulação de Batty (1969) considera o monitoramento de previsões por amortecimento exponencial de qualquer ordem. Além disso, para $\alpha_1 > 0,1$ Batty (1969) construiu tabelas com a distribuição cumulativa de TST , obtida através de simulação, para o caso de previsões por amortecimento exponencial simples e duplo. McKenzie (1978), entre outras contribuições, reviu a pesquisa de Batty (1969) e forneceu novo desenvolvimento, que resultou na expansão dos cálculos para previsões baseadas nos modelos sazonais de Holt-Winters.

Nas situações em que não se conhece a distribuição de TST e TSB resta recorrer à simulação para definir os limites de detecção para o monitoramento do viés, como fez Trigg (1964) para TST a partir de erros de previsão independentes e $\alpha_1 \in \{0,1, 0,2\}$; e Batty (1969) para o caso de erros de previsão por amortecimento exponencial simples, considerando os valores $\alpha_1 \in \{0,1, 0,2, 0,3, 0,4, 0,5\}$.

Gardner (1983, 1985) também utilizou a simulação para obter os limites de TST e TSB com o auxílio do NMA (número médio de períodos até o alarme). O método utilizado consiste em gerar aleatoriamente valores de demanda x_t segundo o modelo

$$x_t = y(t) + \varepsilon_t, \quad (2.10)$$

onde $y(t)$ é função conhecida no tempo e ε_t é uma variável aleatória normal com média $\mu = 0$ e variância σ^2 . Após a geração da demanda, a previsão \hat{x}_t foi feita por amortecimento exponencial simples (como na Equação (2.4)), o erro e_t foi calculado de acordo com a Equação (2.7) e usado nas estatísticas *TST* e *TSB*.

Uma vez que um limite L é adotado, pode ocorrer um alarme, mesmo que não haja mudança no comportamento da demanda, devido à variação natural de x_t . Nesse caso, tem-se um alarme falso. O período até um alarme é armazenado e o processo é repetido até 1000 replicações. Por fim, determina-se o número médio de períodos até um alarme falso (NMA0).

Assim, o procedimento se resume a fixar um valor para NMA0 e utilizar um método de busca para determinar o limite L correspondente a esse valor. Gardner não deu detalhes sobre o método de busca utilizado. O custo com alarmes falsos é tão menor quanto maior o NMA0, portanto, torna-se estratégico fixar esse valor em um patamar considerado aceitável pelo usuário para em seguida encontrar o limite correspondente. Essa metodologia usada por Gardner (1983, 1985), foi adaptada nesta tese, como será discutido nas seções 2.3.3. e 3.2.

Cohen et al. (2009) propuseram, alternativamente, estabelecer os limites de controle a partir de curvas características de operação do receptor (ROC), determinadas empiricamente. Essas curvas fornecem, para cada combinação de α_1 e limites, a taxa de alarmes verdadeiros (indicação de viés quando existe) e a taxa de alarmes falsos (indicação de viés quando não existe). A taxa de alarmes verdadeiros depende da alteração na série, que nesse caso, Cohen et al. consideraram mudanças no nível da série. Segundo esses autores, o usuário deve definir o melhor *tradeoff* entre as taxas de alarmes verdadeiros e falsos, para então localizar nas curvas os parâmetros α_1 e os limites de detecção correspondentes.

2.3.2. Medidas de desempenho

A existência de diferentes estatísticas para monitorar as previsões de demanda levanta a questão sobre qual delas seria mais eficiente na detecção do viés. Gardner (1983) foi um dos pioneiros em buscar uma resposta a essa questão fazendo a comparação de desempenho entre *TST* e *TSB*. A medida de desempenho usada na comparação foi o NMA1 (número médio de períodos até um alarme verdadeiro), foi obtido via simulação em um procedimento semelhante àquele usado para o NMA0.

Para a obtenção do NMA1 o primeiro passo é conhecer os limites de detecção, que são determinados conforme descrito na Seção 2.3.1 para um NMA0 pré-fixado. Na determinação do NMA0 sabe-se que todo o alarme é falso em virtude de não ter sido inserida nenhuma mudança de comportamento na geração da demanda. Por outro lado, para o cálculo do NMA1 é inserida uma mudança no comportamento da demanda para que ocorra um viés. Portanto, todo o alarme que ocorrer após essa mudança é considerado verdadeiro.

A partir de um dado período, em cada “corrida” da simulação é inserida, uma mudança de comportamento em $y(t)$ (como na Equação (2.10)). Assim que ocorrer o alarme, o número de períodos entre a mudança na demanda e a detecção é armazenado, o processo então é reiniciado. Esse procedimento é repetido um grande número de vezes para que por fim seja calculado o número médio de períodos até um alarme verdadeiro (NMA1). As “corridas” em que o alarme ocorre antes da mudança na demanda são descartadas.

Assim, o método mais eficiente na detecção do viés na previsão é aquele que apresentar o menor NMA1 para um mesmo NMA0. Gardner (1983) utilizou essa metodologia para comparar o desempenho das estatísticas *TST* e *TSB*. Em seu experimento, a demanda foi gerada pelo modelo constante $x_t = a + u_t$ (nesse caso $y(t) = a$ na Equação (2.10), onde a é constante). A estimativa \hat{x} foi calculada recursivamente por amortecimento exponencial simples, ou seja,

$$\hat{x}_t = \alpha x_t + (1 - \alpha) \hat{x}_{t-1}, \quad (2.11)$$

sendo $0 < \alpha < 1$. Ele considera $\alpha_1 = \alpha$, onde α_1 é a constante de amortecimento usada para calcular DAM_t , nas Equações (2.8) e (2.9).

Para que ocorresse um viés na previsão, a partir de um dado período foi inserida uma mudança de comportamento na demanda por meio de um incremento Δa no nível do modelo constante (chamado pelo autor de *step*). Esse incremento foi proporcional ao desvio padrão do ruído ε_t .

Para determinar o NMA0 e NMA1 para as estatísticas *TST* e *TSB* Gardner (1983) considerou os casos $\alpha \in \{0,1, 0,2\}$ e concluiu que para $\alpha = 0,1$ a estatística *TSB* tem desempenho equiparado ao de *TST*, porém para $\alpha = 0,2$ *TSB* é superior. O autor sugere então usar *TSB* na detecção de viés.

Uma vez que não há razão para forçar $\alpha_1 = \alpha$, McKenzie (1978) experimentou valores de $\alpha_1 \neq \alpha$, e concluindo que $\alpha_1 < \alpha$ proporciona um melhor desempenho na detecção do viés. Levando em conta esse resultado, Gardner (1985) fez comparações entre as estatísticas *TST* e *TSB* considerando uma demanda constante ($y(t) = a$) e uma previsão por amortecimento exponencial simples. Esse autor testou as combinações entre $\alpha_1 \in \{0,05, 0,1\}$ e $\alpha \in \{0,1, 0,2, 0,3\}$ para as duas estatísticas e concluiu que tanto *TST* quanto *TSB* melhoram o desempenho para $\alpha_1 < \alpha$ e que na maioria dos casos testados *TSB* apresenta melhor desempenho.

Portanto, as pesquisas voltadas para comparar o desempenho entre *TST* e *TSB* na detecção de viés se resumiram a um universo pequeno de combinações de α_1 e α . As previsões monitoradas nessas comparações de desempenho se restringiram àquelas por amortecimento exponencial simples. Além disso, apenas o *step* foi considerado como mudança na demanda que leva ao surgimento do viés.

2.3.3. Diferentes interpretações para NMA0

O número médio de períodos até um alarme falso (NMA0) é fixado pelo gestor com o intuito de calcular o limite de detecção correspondente. Em virtude do custo com alarmes falsos, deseja-se que NMA0 seja muito grande. Infelizmente, os objetivos de NMA0 o maior possível e NMA1 o menor possível são conflitantes, uma vez que quanto maior o NMA0 especificado, maior o limite de detecção, o que leva a um aumento no NMA1. Portanto, o valor adotado para

NMA0 é estratégico, e diferentes metodologias para estimá-lo podem levar a resultados divergentes.

A definição de NMA0 usada por Gardner (1983, 1985) sugere duas interpretações que podem levar a conclusões distintas a respeito do melhor desempenho (McClain, 1988). Essas interpretações são:

NMA0(1) – É o número médio de períodos entre alarmes falsos em um sistema de monitoramento da previsão em operação.

NMA0(2) – É o número médio de períodos até um alarme falso, contados a partir de um ponto arbitrário no tempo.

O procedimento padrão para calcular NMA0(1) é gerar por simulação uma série de demanda com um número de períodos muito grande, e a cada alarme falso armazenar o número de períodos entre esse e o último alarme ocorrido. O monitoramento é reiniciado e contam-se as amostras até que o próximo alarme ocorra. Repete-se o processo dessa forma até que um número determinado de alarmes ocorra, e ao final o NMA0(1) é estimado. Ainda segundo esse autor, existem duas variações de NMA0(1), que dependem da forma de inicialização de *TST* e *TSB* quando ocorre um alarme falso, de forma que:

NMA0(1a) – em caso de alarme falso, antes de reiniciar o monitoramento, as estatísticas do erro amortecido inicial EAM_0 e do erro acumulado inicial EAC_0 retornam para zero;

NMA0(1b) – em caso de alarme falso, antes de reiniciar o monitoramento, as estatísticas do erro amortecido inicial EAM_0 e do erro acumulado inicial EAC_0 retornam para zero, e tanto o desvio absoluto médio inicial DAM_0 quanto a estimativa inicial da previsão \hat{x}_0 são substituídos por seus valores esperados.

McClain (1988) batizou NMA0(1a) de “*no reset*” e NMA0(1b) de “*perfect reset*” e mostrou que os valores de NMA0 podem diferir em até 250%, dependendo do método utilizado em sua determinação. Esse autor concluiu que Gardner (1983, 1985) usou o NMA0(2), e nesse caso o DAM_0 foi calculado por uma expressão analítica que é função da variância do ruído da série temporal, e que a previsão inicial foi definida como o nível do modelo constante ($\hat{x}_0 = a$). Esse procedimento é criticado por McClain (1988) por ser inviável na prática conhecer perfeitamente a variância do ruído e a média da demanda. O autor concorda com o procedimento de manter fixo o NMA0. Entretanto, recomenda que o melhor procedimento para obter esse valor é o NMA0(1b) ou “*no reset*”.

Nesta tese, a obtenção do NMA0 é um recurso semelhante ao utilizado por Gardner (1983, 1985) diferindo-se apenas nas estimativas iniciais de DAM_0 e \hat{x}_0 , que nesse caso são obtidas a partir de uma amostra. Busca-se com esse procedimento contornar os pontos frágeis apontados por McClain (1988) com relação ao NMA0(2) usado por Gardner (1983, 1985).

2.3.4. Outras medidas de desempenho

Apesar de concordar com NMA0 como uma boa forma de determinar os limites de detecção, McClain (1988) afirma que o desempenho não deve ser medido usando NMA em virtude desta estatística apresentar uma variância grande. Como alternativa o autor sugere que o ideal é observar o número de detecções até o enésimo período. A ideia é que se deve definir quando o viés deve ser detectado. Portanto, o melhor método na detecção é aquele que apresenta maior número de detecções até o período desejado, ou de outra forma, aquele que tem maior probabilidade acumulada de detecção até aquele período.

Outra proposta para medir o desempenho foi sugerida por Cohen et al. (2009), que através das curvas características de operações afirmam que o melhor desempenho corresponde a maior taxa de alarmes verdadeiros juntamente com a menor taxa de alarmes falsos. As amostras usadas na obtenção das curvas características de operação foram extraídas de uma série temporal real. Assim como sugerido por McClain (1988) os alarmes falsos e verdadeiros contabilizados foram apenas aqueles que ocorreram dentro de um número máximo de períodos considerado.

2.3.5. Críticas ao uso de *TST* e *TSB* e oportunidades de contribuição da pesquisa

As duas técnicas descritas para detecção de viés se apresentam de fácil uso e necessitam da armazenagem de poucos dados. Em uma situação prática, esses aspectos são importantes. Apesar disso, cada uma dessas técnicas tem suas desvantagens: a principal crítica ao *TSB* se deve ao fato de seu numerador ser uma soma de variáveis aleatórias, conseqüentemente seu desvio-padrão cresce, fazendo com que a probabilidade de alarme falso também cresça.

A estatística TST não apresenta esse problema, pois seu desvio-padrão é limitado, uma vez que trabalha com o erro amortecido no numerador. Por outro lado, a estatística TST também tem sua desvantagem em relação à TSB : Como foi mostrado por Gardner (1983), o desempenho na detecção piora para $\alpha_1 > 0,2$ o que não ocorre com TSB , cujo desempenho mostrou-se estável nessa situação.

Os artigos citados não exploraram combinações entre α e α_1 suficientes para definir qual dessas técnicas é superior ou em que situação cada uma é preferível. Assim, nesta tese, resultados conclusivos foram obtidos investigando um número maior de combinações de α (18 valores) e α_1 (10 valores); tais resultados são apresentados na Seção 3.3. Além das estatísticas citadas, foram também testadas outras duas estatísticas, conforme será descrito na Seção 3.2.

2.4. Sistemas de Controle de Estoques

A decisão sobre o controle de estoque se inicia pela definição da frequência com que o nível do estoque (unidades em estoque + unidades pedidas – unidades em falta) será verificado, de acordo com Silver *et al.* (1998), existem duas opções: revisão contínua e revisão periódica. O autor fornece dois exemplos de revisão contínua observados na prática: sistemas que atualizam as informações do estoque a cada transação realizada (recebimento, venda, envio para o cliente, etc); ou sistemas que atualizam as informações a partir dos pontos de venda no varejo, ou seja, a cada venda o nível do estoque é atualizado. Nesses exemplos, as transações podem ocorrer a qualquer momento, por esse motivo, a revisão é chamada de contínua. No caso da revisão periódica, o estoque é verificado em intervalos regulares T . Portanto, entre duas verificações consecutivas não é conhecido o *status* do estoque.

A partir do momento que a frequência de verificação do nível de estoque é definida, segundo Silver *et al.* (1998), considera-se um entre quatro dos principais sistemas de controle de estoque descritos a seguir.

Ponto de pedido e lote fixo com revisão contínua – nesse sistema, o nível de estoque é verificado continuamente. Quando esse nível é igual ou menor a uma quantidade R , um pedido de Q unidades é enviado ao fornecedor, caso contrário, nada é feito. Tem-se então o sistema (R, Q) .

Ponto de pedido e estoque máximo com revisão contínua – nesse sistema, o nível de estoque é verificado continuamente. Quando esse nível for igual ou menor a uma quantidade R , o pedido então é feito de tal forma que o nível de estoque seja incrementado até um valor máximo S , como consequência o lote pedido não é fixo. Esse sistema é conhecido como (R, S) .

Estoque máximo com revisão periódica – nesse sistema, a cada intervalo T é realizado um pedido de tal forma que o nível de estoque seja incrementado até um valor máximo S . Esse sistema é conhecido como (T, S) .

O sistema (T, R, S) é a combinação de outros dois - nesse sistema, a cada intervalo T o nível de estoque é verificado. Quando esse nível é igual ou menor a uma quantidade R , é realizado um pedido de tal forma que o nível de estoque seja incrementado até um valor máximo S . Um caso particular desse sistema é conhecido como $(S-1, S)$. Nesse sistema, o ponto de pedido é definido por $R=S-1$, ou seja, quando o nível de estoque estiver ao menos uma unidade abaixo de S um pedido deve ser feito para que a meta S seja atingida novamente. Esse é o sistema de controle de estoque que está relacionado com o escopo desta tese, em virtude de ser usado no varejo.

Os valores s, S, R, T e Q , que definem os sistemas de controle de estoque, podem ser definidos por modelos de estoque embasados em formulações matemáticas que levam em consideração dados que caracterizam o sistema de controle e o sistema físico representado, por exemplo: demanda, custos com estoques, tipo de revisão, restrições de capital, etc.

Um desses dados é o *lead time*, que é definido como o intervalo de tempo entre a realização de um pedido e o seu recebimento. Outro dado de entrada é a demanda, que, dependendo da forma como é considerada, divide os modelos em dois grupos: modelos determinísticos, que consideram a demanda como determinística; e modelos probabilísticos, que consideram a demanda ou *lead time* como variáveis aleatórias.

O objetivo da maioria dos modelos de estoque é minimizar o custo total com estoques. De acordo com Silver *et al.* (1998), os custos considerados relevantes na modelagem de estoques são os de encomendar (ou fazer pedidos), custos de manutenção dos estoques e custos com faltas ou atraso no atendimento ao cliente. O custo de encomendar é proporcional ao número de pedidos em um dado período. Esse custo considera os gastos com o processamento do pedido e operações de recebimento. O custo de encomendar

é considerado constante e independente da quantidade de itens envolvida nos pedidos.

O custo de manter o estoque está relacionado com os gastos para manter os itens no estoque até o momento de seu envio para o cliente. O principal custo relacionado à manutenção do estoque é o custo de capital (ou de oportunidade) proveniente da depreciação do capital imobilizado em forma de itens estocados.

As penalidades provenientes da falta de produtos ou atrasos na entrega implicam nos custos de falta. Por definição, custos de falta devem englobar os custos diretos (venda perdida) mais os custos indiretos relacionados com a falta do item, como por exemplo, a perda da confiabilidade do cliente na loja. A tarefa de determinar o custo de falta pode ser complexa, sobretudo do ponto de vista da quantificação do custo indireto, em virtude da dificuldade de identificá-lo e rateá-lo entre as demais despesas.

Uma alternativa ao uso do custo de falta é impor ao modelo de estoque uma restrição de nível de serviço β conhecida como *fill rate*. Essa restrição representa a fração da demanda prontamente atendida a partir do estoque. Embora existam outras definições para nível de serviço a única considerada nesta tese será o *fill rate*.

Assim, modelos de estoque têm o objetivo de auxiliar nas decisões do sistema de controle do estoque, de forma que a demanda seja atendida ao menor custo, levando também em conta certas restrições, por exemplo, limite de quantidade pedida, nível de serviço mínimo, limite de custo, etc.

2.4.1. Modelos de estoque em camadas

Os modelos de estoque em camadas são elaborados a fim de representar a movimentação dos estoques em redes de suprimento. Essas redes apresentam ao longo de suas camadas processos de montagem, de produção e distribuição. A partir de uma determinada camada o fluxo de materiais passa a ter atividades somente de distribuição. Em geral, a camada que antecede as lojas do varejo apresenta somente atividades de distribuição.

Segundo Axsäter (2006) e Zipikin (2000), resultados teóricos e empíricos apontam que sistemas com múltiplos níveis devem ser modelados de forma conjunta, pois a prática de otimizar as instalações separadamente não resulta no menor custo total com estoques para o sistema.

A modelagem de estoques em um sistema com camadas implica no aumento da complexidade matemática, quando comparada com a modelagem isolada de cada instalação. Em muitos casos as soluções de modelos em camadas exigem grande esforço computacional, resultado da interação entre os níveis de reposição de estoques adotados em cada instalação. Essa complexidade aumenta com o número de instalações e camadas consideradas. Mesmo para modelos com duas camadas, por exemplo, um centro de distribuição (CD) que abastece lojas no varejo, poucas soluções ótimas (métodos exatos) foram desenvolvidos até hoje.

Alguns exemplos de soluções ótimas são encontrados em Graves (1985), Chen e Zheng (1997), Cheung e Hausman (2000) e Axsäter (2000), todos esses consideram revisão contínua e política de ponto de pedido com lote fixo. Soluções ótimas para sistemas com revisão periódica são ainda mais raras. Chu e Shen (2010) citam Cachon (2001) e Liljenberg (1996) como exemplos de soluções ótimas para sistemas com revisão periódica e lote fixo. Em função da complexidade em obter soluções ótimas, as pesquisas têm se concentrado em métodos heurísticos, que via de regra, são validados por experimentos de simulação.

Algumas abordagens de modelos de estoque em camadas têm como premissas o conhecimento do nível de estoque, dos níveis de reposição e da demanda em todas as instalações do sistema. Na prática, em uma rede de suprimentos ainda é um desafio o compartilhamento de informação ao longo de toda a rede, pois geralmente as empresas consideram essas informações estratégicas e preferem mantê-las em sigilo. Apesar disso, com a evolução do conceito de gestão logística integrada, cada vez mais existe o estímulo para que haja um maior compartilhamento dessas informações.

A ideia de que a informação está cada vez mais disponível tem se refletido na tendência observada recentemente nos modelos de estoque em duas camadas, como por exemplo, Cachon e Fisher (2000), Moizadeh (2002), Özer (2003), Axsäter e Marklund (2008), Bakal, Erkip e Güllü (2011) e Howard e Marklund (2011). Esses estudos propuseram novas políticas de estoque sob a premissa de obtenção de informações em tempo real. Informações como nível de estoque, custos com transporte e preços são utilizadas para definir como o centro de distribuição deve alocar o estoque disponível nas lojas do varejo. Segundo seus autores essas políticas resultaram em menor custo com estoques quando comparado com políticas tradicionais.

A seguir será descrita uma situação típica que permite assumir que exista o compartilhamento de informações em tempo real.

2.4.2. Hipóteses e a física do problema

Apesar da dificuldade em compartilhar informações em uma rede de suprimentos, algumas situações favorecem a aplicação desses modelos de estoque multi-camada. Um exemplo seria uma empresa que abrange mais de uma camada em uma rede, como no caso das lojas do varejo que são abastecidas pelo próprio centro de distribuição (CD). Nessa situação essas informações são enviadas por meio eletrônico das lojas para o CD, que tem plena condição de decidir sobre o envio de itens para as lojas, bem como de definir seus pedidos para os fornecedores externos. Assim, esse sistema com duas camadas pode ser representado como na Figura 2.2.

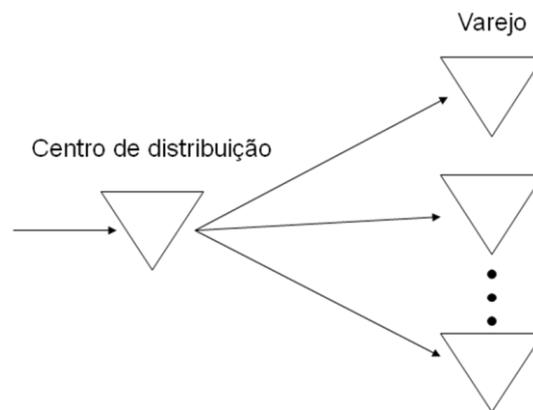


Figura 2.2 – Sistema centralizado com duas camadas

Esse sistema é centralizado, ou seja, todas as informações de estoque e demanda nas lojas são conhecidas no CD. Além disso, os parâmetros de reposição tanto nas lojas quanto no CD são definidos no CD.

Todas as instalações no sistema da Figura 2.2 adotam política de revisão periódica do estoque. O intervalo de revisão no varejo é igual em todas as lojas, enquanto que o CD adota um intervalo múltiplo inteiro daquele aplicado no varejo. A política de ressuprimento adotada nas instalações consiste em realizar um pedido a cada intervalo de revisão T visando atingir o nível máximo de reposição S . Essa política é conhecida como (T, S) . Os intervalos de revisão nas lojas, além de iguais, são sincronizados, de maneira que o CD verifica seus níveis de estoque simultaneamente.

Esse sistema é recorrente no varejo, sobretudo em supermercados e hipermercados no Brasil. As peculiaridades desse sistema, em relação aos outros apresentados na literatura, estão na diferença entre os intervalos de revisão no CD e no varejo, na verificação simultânea dos níveis de estoque nas lojas e na possibilidade de as necessidades de reposição nas lojas serem atendidas parcialmente. A revisão de literatura a seguir é direcionada para modelos aplicados a sistemas com as mesmas características apresentadas, pois esses são de interesse específico para esta tese.

Na revisão de literatura não foi encontrada uma proposta para um modelo de estoques em camadas que adotasse o sistema (T, S) em todas as instalações e que tivesse as seguintes características: permitir entregas parciais nas lojas, utilizar diferentes intervalos de revisão no CD e nas lojas; e que a demanda nas lojas não seja restrita a uma distribuição de Poisson ou Poisson composto.

2.4.3.

Sistemas com duas camadas e política (T, S)

Um sistema semelhante ao descrito na seção anterior foi estudado em Heijden (1992), nesse caso as instalações utilizam diferentes intervalos de revisão nas lojas do varejo, o que permite que os pedidos cheguem ao CD em momentos distintos e sejam atendidos com uma disciplina PEPS (primeiro a entrar, primeiro a ser servido).

Heijden (1992) forneceu formulações para estimar as variáveis: estoque médio, nível de serviço e atraso nos pedidos; todos em função dos níveis máximos de estoque S pré-definidos em cada instalação. Em seguida, as estimativas dessas variáveis foram usadas para criar um ajuste por uma distribuição de Cox (conforme Heijden (1988)), cujos resultados foram validados por simulação. Conforme citado, esse modelo depende de uma solução inicial S utilizada para estimar as variáveis de interesse do problema, sugerindo uma solução por um processo de tentativa e erro, o que não garante o resultado ótimo.

A abordagem de Heijden (1992) é geral, porém não é aplicável ao problema estudado: em primeiro lugar, os pedidos só podem ser atendidos em sua totalidade, ou seja, não é permitido entrega parcial, situação muito recorrente na prática; a segunda diferença está na premissa de que os pedidos

são atendidos pelo CD em uma disciplina PEPS, o que não é aplicável quando esses pedidos chegam simultaneamente.

Outro sistema semelhante encontrado na literatura foi estudado por Heijden (2000), nesse caso as instalações também utilizam revisão periódica, porém o sistema pode ter várias camadas caracterizando uma rede de suprimentos. O autor fornece equações aproximadas em função de uma solução inicial S , com o intuito de estimar os estoques médios e os níveis de serviço no varejo. O objetivo é minimizar o custo do sistema e garantir o nível de serviço no varejo, para tanto Heijden (2000) fornece um algoritmo de busca. Apesar da semelhança com o sistema estudado essa solução também não se aplica, pois para Heijden (2000) todas as instalações da rede possuem o mesmo intervalo de revisão e os pedidos são sincronizados em todas as camadas.

Recentemente, Wang (2013) forneceu uma solução para o sistema descrito na Seção 2.4.2., onde considera que a demanda no varejo seguem uma distribuição de Poisson com média igual em todas as lojas. Esse autor observou que seria necessário o CD aplicar uma regra de racionamento para alocar o estoque no varejo em virtude da verificação simultânea dos níveis de estoque nas lojas. Propôs então que a alocação dependa da informação do nível de estoque em cada loja, de forma que essa alocação seja feita com o objetivo de equilibrar os custos de estoque com os custos de falta. Além disso, o autor fornece ainda um algoritmo para otimizar de forma conjunta o intervalo de revisão no varejo e no CD juntamente com seus níveis de reposição. O algoritmo exige grande esforço computacional, exceto quando a média da demanda e o número de lojas são pequenos.

As pesquisas citadas são as soluções aplicadas a sistemas próximos àquele descrito na Seção 2.4.2., porém não foi encontrada uma solução específica para o sistema estudado.

2.4.4. Técnica de desagregação do sistema

Modelar um sistema de estoque com múltiplas camadas no qual a demanda no varejo não é determinística é um desafio. Uma técnica bastante usada para simplificar o problema consiste em encontrar uma forma de desagregar o sistema em N instalações, que possam ser tratadas isoladamente. Uma das formas de desagregar o sistema é determinar a

influência dos níveis de reposição do estoque de uma instalação no *lead time* da instalação que ela abastece (Sherbrook,1961).

Ainda que o tempo de transferência de produtos entre CD e varejo seja considerado constante, o intervalo entre a realização de um pedido para o CD e o recebimento desse pedido no varejo é uma variável aleatória. Isso é consequência de eventuais faltas no CD. Define-se então *lead time* efetivo no varejo i como

$$\bar{\ell}_i = L_i + \bar{w}_i, \quad (2.12)$$

sendo L_i o *lead time* no varejo i quando não há faltas e \bar{w}_i o atraso médio nos pedidos do varejo i .

Uma das formas de obtenção do \bar{w} é considerar que os pedidos do varejo pendentes no CD caracterizam uma fila. Dessa forma, o atraso médio em um pedido é o tempo médio de espera na fila. Segundo a Teoria de Filas, o tempo médio de espera na fila é obtido pela razão entre o comprimento médio da fila e a taxa de chegada dos clientes, resultado conhecido como fórmula de Little (Gross & Harris, 1997).

Portanto, o atraso médio nos pedidos é definido pela razão entre o valor esperado de faltas no CD (comprimento médio de fila) é o número médio de pedidos realizados em dado intervalo de tempo (taxa de chegada de clientes).

O método METRIC, apresentado por Sherbrook (1961), foi um dos primeiros trabalhos no contexto de estoque em múltiplas camadas que utilizaram a fórmula de Little, e que direcionaram muitos outros com a mesma filosofia. Em seu modelo, o atraso médio nos pedidos é acrescido ao *lead time* de todas as instalações no varejo, o que significa que o atraso médio nos pedidos é o mesmo para todas as lojas. A premissa de que o atraso nos pedidos independe de sua origem é aceitável quando a média da demanda no varejo é semelhante em todas as lojas.

A aplicação da fórmula de Little não é geral e uma de suas premissas define que a demanda observada no CD deve seguir um processo de Poisson, ou Poisson composto (Diks *et al.*,1996). Além disso, a demanda no varejo deve ser unitária, ou seja, cada pedido solicita apenas uma unidade, como ocorre no modelo METRIC e suas extensões, por exemplo, MOD-METRIC (Muckstad, 1973), VARI-METRIC (Sherbrook, 1992) e recentemente Lu e Yang (2012).

Apesar dessa restrição, Svoronos & Zipkin (1988) generalizam a aplicação da fórmula de Little em um sistema com demanda determinística D no CD, no qual o varejo encomenda uma quantidade fixa Q igual para todas as lojas. Em um sistema com essas mesmas características, Axsäter (2003) propôs a aplicação da fórmula de Little quando o lote pedido pelo varejo é restrito a nQ , sendo n um número inteiro e Q um lote fixo.

A determinação do atraso nos pedidos do varejo é importante na desagregação de sistemas com múltiplas camadas e, de acordo com Diks et al. (1996), a fórmula de Little não pode ser usada em sistemas com revisão periódica e sem a presença de lotes fixos, a exemplo do sistema descrito na Seção 2.4.2.

Assim, foi necessário um novo desenvolvimento para definir o *lead time* efetivo $\bar{\ell}$ (discutido na Seção 4.5.) fazendo uso do fator de racionamento (seção 4.4.). Além disso, foi elaborada uma função de custo e um algoritmo de otimização, conforme será apresentado na Seção 4.6.3.

3. Metodologia de Monitoramento da Previsão

Neste capítulo é apresentada a metodologia proposta para determinar os limites de controle das estatísticas *Tracking Signal*. Essa metodologia segue o princípio daquela usada em Gardner (1983, 1985) (descrita na Seção 2.3.1), diferenciando-se apenas na forma de obtenção das estimativas DAM_0 e x_0 , que são calculadas a partir de uma amostra. Entende-se que esse procedimento é uma representação mais próxima de uma situação prática.

Ainda neste capítulo é apresentada a metodologia para identificar qual das estatísticas é mais eficiente na detecção do viés, sendo que além das estatísticas *TSB* e *TST* nas Equações (2.8) e (2.9) respectivamente, são testadas duas de suas variantes sugeridas por Gardner (1983). Dessa forma, quatro estatísticas foram testadas, via simulação, o que apontou as mais eficientes sob certas condições. Ao final do capítulo são apresentados os limites de controle para cada α usado na previsão por amortecimento exponencial simples.

3.1. Variantes para as estatísticas *Tracking Signal*

O objetivo deste capítulo é investigar o desempenho na detecção de viés na previsão de demanda com modelo constante, de forma que

$$x_t = a + \varepsilon_t, \quad (3.1)$$

sendo a constante e ε_t definido como ruído, com média $\mu = 0$ e variância $\sigma^2 > 0$. A estimativa da demanda ($\hat{x}_{t,t+\tau}$) é para um passo a frente ($\tau = 1$), realizada no período t por amortecimento exponencial simples conforme Equação (2.3). O erro e_t é calculado de acordo com a Equação (2.7).

A mudança de comportamento na demanda é um incremento Δa inserido na média a a partir do período 20 após o início do monitoramento.

Além das estatísticas TSB e TST , são testadas duas de suas variantes sugeridas por Gardner (1983) com o intuito de verificar se existe melhoria no desempenho na detecção de viés. Essas variantes são obtidas pela substituição do DAM pela raiz do erro quadrático médio EQM nas equações (2.8) e (2.9). São elas:

$$TSB'_t = \frac{EAC_t}{\sqrt{EQM_t}} \quad (3.2)$$

e

$$TST'_t = \frac{EAM_t}{\sqrt{EQM_t}}, \quad (3.3)$$

Sendo EQM_t o valor estimado do erro quadrático médio, calculado recursivamente no período t pela fórmula $EQM_t = \alpha_1 e_t^2 + (1 - \alpha_1)EQM_{t-1}$. O valor inicial EQM_0 pode-se usar uma estimativa de $E(e_t^2)$ obtida a partir de dados históricos (por exemplo, simulando a previsão, ou, alternativamente, calculando os resíduos de um modelo ajustado aos dados históricos).

3.2.

Experimento de simulação e determinação dos limites de detecção

Para a determinação dos limites de detecção e para os testes de desempenho, sem que houvesse perda de generalidade para os resultados, a demanda foi gerada com nível $a=0$ e o ruído segundo uma distribuição normal com média $\mu=0$ e variância $\sigma^2=1$.

Os limites de detecção foram ajustados de maneira a fornecer $NMA0=50$, o que significa, por exemplo, esperar em média um alarme falso por ano em um sistema de previsões semanais. A forma de obtenção de $NMA0$ usada nesta tese, para um dado limite de detecção, é semelhante àquela encontrada em Gardner (1983) ($NMA0(2)$), diferenciando-se na obtenção das estimativas iniciais EQM_0 , \hat{x}_0 e DAM_0 , que são calculadas por

$$EQM_0 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}; \quad (3.4)$$

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}; \quad (3.5)$$

e

$$DAM_0 = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|. \quad (3.6)$$

Essas estimativas foram feitas a partir dos primeiros $N=36$ períodos de cada corrida de simulação;

A metodologia aqui proposta também se difere por utilizar um método de busca (ver Apêndice A) para encontrar os limites de detecção para cada par (α, α_1) . Para cada limite de detecção testado pelo método de busca foram utilizadas 10.000 corridas de simulação para o cálculo do NMA0 resultante. Os pares (α, α_1) nos casos simulados foram formados pela combinação dos valores: $\alpha \in \{0,045, 0,3\}$ a cada 0,015 e $\alpha_1 \in \{0,01, 0,1\}$ a cada 0,01.

No cálculo do NMA1 foram introduzidas, em cada corrida de simulação, mudanças nas demandas estudadas a partir do vigésimo período contado do início do monitoramento. O incremento no nível do modelo constante (*step*) foi expresso como múltiplos de σ . As mudanças de comportamento inseridas na demanda foram: $\Delta a \in \{0,5\sigma, 1\sigma, 1,5\sigma, 2\sigma\}$.

Em síntese, para cada par (α, α_1) foram determinados os limites $L \in \{LT, LT', LB, LB'\}$ das estatísticas TST, TST', TSB e TSB' , tendo como meta $NMA0=50$. Em seguida, a fim de determinar o desempenho de cada estatística na detecção do viés, o NMA1 foi calculado em cada caso via nova simulação para cada combinação $(\alpha, \alpha_1, L, \Delta a)$.

3.3. Apresentação dos resultados

Ao final do experimento foram obtidos 720 resultados de NMA1 para cada estatística testada em cada caso de mudança de comportamento na demanda. Para que os desempenhos fossem comparados, foi necessário adotar um critério de seleção do par (α_1, L) , para cada estatística em cada caso, que resultasse na

detecção mais eficiente. Observou-se então que, com poucas exceções, as combinações de (α_1, L) que tiveram melhor desempenho para pequenas alterações na demanda mantiveram essa superioridade para alterações maiores.

Portanto, esse critério foi utilizado para a seleção dos melhores desempenhos de cada estatística isoladamente. Sendo assim, os pares (α_1, L) que resultam nos menores NMA1's para $\Delta a = 0,5\sigma$, para cada estatística em cada caso, foram agrupados na Tabela 3.1 para fins de comparação entre as estatísticas.

Observa-se que todas as estatísticas testadas apresentaram desempenho semelhante. Nota-se ainda que para pequenos Δa as diferenças entre valores de NMA1 são maiores (para qualquer α em qualquer estatística). Contudo, com o aumento em Δa e α essas diferenças diminuem consideravelmente, ou seja, quanto maior o α e mais acentuada a mudança de comportamento na demanda, mais semelhantes se tornam os desempenhos.

Vale ressaltar que todos os resultados de desempenho independem do valor inicial de a e que os valores de NMA1 são constantes, desde que mantida a proporção $\Delta a/\sigma$, o que reflete a generalidade desses resultados.

Para que a estatística mais eficiente fosse identificada mesmo com a semelhança entre resultados, o erro-padrão da simulação foi utilizado para verificar se alguma estatística é significativamente melhor que outra. Dessa forma, considerando na Tabela 3.1 a faixa $[NMA1 - \text{erro}, NMA1 + \text{erro}]$, os resultados comparados onde ocorre intercessão nessa faixa são considerados equivalentes. Em caso contrário, um resultado será significativamente superior ao outro. Na Figura 3.1 encontra-se uma ilustração de como o erro-padrão, denominado de $e_{padrão}$, foi utilizado para comparação das estatísticas. Nessa ilustração, Técnica1 e Técnica2 representam duas estatísticas.

Dessa forma, concluiu-se que para valores de $\alpha \in \{0,165, 0,3\}$, as estatísticas TSB e TSB' apresentaram desempenho melhor que as estatísticas TST e TST' . Para os outros valores de α , de fato, não houve diferenças significativas entre desempenhos. Essas conclusões estão ilustradas na Figura 3.3 a partir da observação da Tabela 3.1.

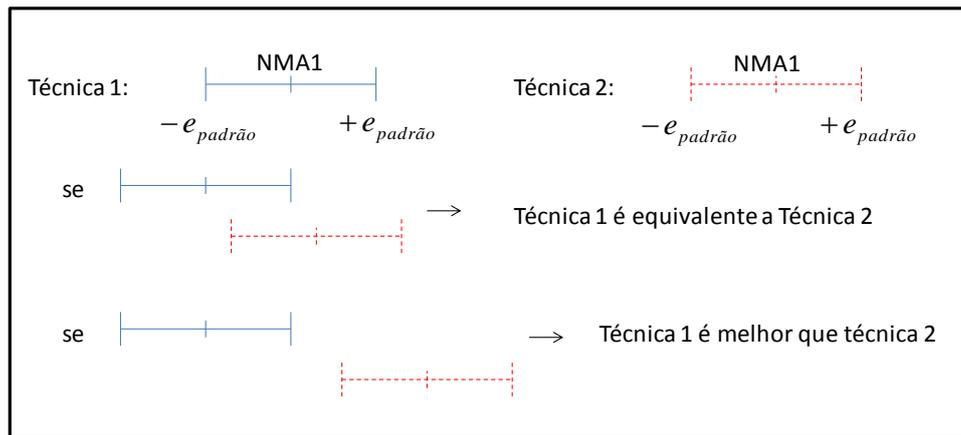


Figura 3.1 – comparação de desempenhos com o $e_{padrão}$

Tabela 3.1 – Valores de NMA1 em função de $(\alpha, \alpha_1, L, \Delta a)$

α	Δa	TST		TST'		TSB		TSB'					
		NMA1	erro	NMA1	erro	NMA1	erro	NMA1	erro				
0,045	0,5	α_1	15,8	0,18	α_1	15,6	0,19	α_1	16,3	0,18	α_1	16,5	0,29
	1	0,01	7	0,04	0,01	6,93	0,04	0,06	7,99	0,04	0,01	7,32	0,06
	1,5	LT	4,6	0,02	LT'	4,58	0,02	LB	5,51	0,02	LB'	4,87	0,03
	2	0,06	3,48	0,01	0,05	3,48	0,01	6,89	4,32	0,01	5,49	3,64	0,02
0,06	0,5	α_1	16,1	0,20	α_1	15,9	0,21	α_1	16,2	0,20	α_1	16,3	0,33
	1	0,01	6,84	0,05	0,01	6,79	0,05	0,04	7,43	0,06	0,03	7,43	0,09
	1,5	LT	4,41	0,03	LT'	4,38	0,03	LB	4,95	0,03	LB'	4,88	0,05
	2	0,06	3,32	0,02	0,05	3,3	0,02	6,41	3,81	0,02	5,12	3,75	0,03
0,075	0,5	α_1	16,5	0,25	α_1	16	0,24	α_1	16,3	0,22	α_1	16,1	0,36
	1	0,02	6,6	0,05	0,01	6,74	0,05	0,04	7,15	0,06	0,03	7,2	0,09
	1,5	LT	4,25	0,03	LT'	4,27	0,03	LB	4,71	0,03	LB'	4,6	0,05
	2	0,11	3,21	0,02	0,05	3,22	0,02	6,03	3,64	0,02	4,8	3,59	0,03
0,09	0,5	α_1	16,3	0,24	α_1	15,7	0,23	α_1	16,4	0,23	α_1	16,1	0,38
	1	0,01	6,5	0,05	0,01	6,4	0,05	0,04	7,07	0,06	0,03	6,86	0,09
	1,5	LT	4,14	0,03	LT'	4,1	0,03	LB	4,57	0,03	LB'	4,51	0,05
	2	0,06	3,1	0,02	0,04	3,08	0,02	5,73	3,47	0,02	4,56	3,43	0,03
0,105	0,5	α_1	17	0,23	α_1	16,6	0,23	α_1	16,3	0,25	α_1	16,3	0,39
	1	0,01	6,39	0,05	0,01	6,28	0,05	0,03	6,83	0,06	0,04	6,84	0,10
	1,5	LT	4,01	0,03	LT'	3,97	0,03	LB	4,32	0,03	LB'	4,45	0,05
	2	0,05	2,98	0,02	0,04	2,98	0,02	5,46	3,3	0,02	4,35	3,4	0,03
0,12	0,5	α_1	16,8	0,23	α_1	16,4	0,23	α_1	16,4	0,26	α_1	16,2	0,43
	1	0,01	6,3	0,05	0,01	6,21	0,05	0,04	6,81	0,06	0,07	7,02	0,10
	1,5	LT	3,89	0,03	LT'	3,86	0,03	LB	4,31	0,03	LB'	4,59	0,05
	2	0,05	2,92	0,02	0,04	2,9	0,02	5,23	3,23	0,02	4,18	3,61	0,03
0,135	0,5	α_1	17,7	0,26	α_1	17,2	0,25	α_1	16,7	0,28	α_1	16,6	0,44
	1	0,01	6,34	0,06	0,01	6,24	0,05	0,08	6,99	0,06	0,03	6,55	0,10
	1,5	LT	3,86	0,03	LT'	3,81	0,03	LB	4,57	0,03	LB'	4,17	0,05
	2	0,05	2,86	0,02	0,04	2,84	0,02	5,02	3,54	0,02	4,02	3,15	0,03
0,15	0,5	α_1	17,5	0,27	α_1	16,8	0,28	α_1	16,5	0,29	α_1	16,6	0,47
	1	0,01	6,3	0,06	0,01	6,2	0,06	0,01	6,41	0,06	0,03	6,57	0,10
	1,5	LT	3,79	0,03	LT'	3,74	0,03	LB	3,92	0,03	LB'	4,08	0,05
	2	0,05	2,79	0,02	0,04	2,77	0,02	4,87	2,86	0,02	3,88	3,04	0,03
0,165	0,5	α_1	19	0,30	α_1	18,4	0,29	α_1	17,1	0,30	α_1	16,8	0,51
	1	0,01	6,3	0,06	0,01	6,21	0,06	0,07	6,87	0,07	0,07	6,73	0,12
	1,5	LT	3,74	0,03	LT'	3,71	0,03	LB	4,28	0,03	LB'	4,39	0,05
	2	0,05	2,77	0,02	0,04	2,75	0,02	4,67	3,26	0,02	3,75	3,31	0,03

Tabela 3.1 (continuação) – Valores de NMA1 em função de $(\alpha, \alpha_1, L, \Delta a)$

α	TST				TST'				TSB				TSB'				
	Δa		NMA1	erro	α_1		NMA1	erro	α_1		NMA1	erro	α_1		NMA1	erro	
0,18	0,5	α_1	19	0,31	α_1	18,2	0,30	α_1	17,2	0,36	α_1	17	0,59				
	1	0,01	6,31	0,06	0,01	6,18	0,06	0,02	6,4	0,07	0,03	6,51	0,12				
	1,5	LT	3,74	0,03	LT'	3,69	0,03	LB	3,87	0,03	LB'	3,94	0,05				
	2	0,04	2,72	0,02	0,04	2,7	0,02	4,58	2,82	0,02	3,63	2,93	0,03				
0,195	0,5	α_1	19,4	0,31	α_1	18,7	0,30	α_1	17,3	0,37	α_1	17,2	0,60				
	1	0,01	6,35	0,06	0,01	6,21	0,06	0,05	6,55	0,08	0,03	6,44	0,12				
	1,5	LT	3,68	0,03	LT'	3,63	0,03	LB	4	0,03	LB'	3,88	0,05				
	2	0,04	2,67	0,02	0,03	2,65	0,02	4,42	3,02	0,02	3,53	2,87	0,03				
0,21	0,5	α_1	19,3	0,32	α_1	18,5	0,33	α_1	17,6	0,38	α_1	17	0,62				
	1	0,01	6,24	0,06	0,01	6,16	0,06	0,02	6,31	0,08	0,03	6,42	0,13				
	1,5	LT	3,64	0,03	LT'	3,61	0,03	LB	3,75	0,03	LB'	3,85	0,05				
	2	0,04	2,62	0,02	0,03	2,61	0,02	4,34	2,75	0,02	3,44	2,84	0,03				
0,225	0,5	α_1	19,4	0,34	α_1	18,7	0,34	α_1	17,8	0,40	α_1	17,8	0,65				
	1	0,01	6,22	0,07	0,01	6,08	0,07	0,04	6,46	0,09	0,07	6,71	0,15				
	1,5	LT	3,64	0,03	LT'	3,61	0,03	LB	3,88	0,04	LB'	4,14	0,06				
	2	0,04	2,59	0,02	0,03	2,58	0,02	4,39	2,82	0,02	3,34	3,13	0,03				
0,24	0,5	α_1	20,2	0,37	α_1	19,5	0,36	α_1	18,2	0,47	α_1	17,8	0,72				
	1	0,01	6,26	0,07	0,01	6,15	0,07	0,05	6,51	0,10	0,03	6,39	0,16				
	1,5	LT	3,6	0,03	LT'	3,56	0,03	LB	3,92	0,03	LB'	3,8	0,06				
	2	0,04	2,6	0,02	0,03	2,58	0,02	4,09	2,86	0,02	3,28	2,77	0,03				
0,255	0,5	α_1	20,7	0,35	α_1	19,6	0,35	α_1	18,1	0,46	α_1	18,1	0,79				
	1	0,01	6,36	0,06	0,01	6,21	0,06	0,03	6,44	0,11	0,03	6,47	0,16				
	1,5	LT	3,55	0,03	LT'	3,49	0,03	LB	3,78	0,04	LB'	3,73	0,06				
	2	0,04	2,54	0,02	0,03	2,51	0,02	4,02	2,69	0,02	3,21	2,69	0,03				
0,27	0,5	α_1	21,2	0,36	α_1	20,1	0,36	α_1	18,8	0,50	α_1	18,3	0,80				
	1	0,01	6,39	0,07	0,01	6,24	0,07	0,07	6,92	0,11	0,08	6,86	0,17				
	1,5	LT	3,54	0,03	LT'	3,51	0,03	LB	4,01	0,04	LB'	4,15	0,06				
	2	0,04	2,53	0,02	0,03	2,52	0,02	3,9	2,92	0,02	3,11	3,07	0,04				
0,285	0,5	α_1	21,6	0,39	α_1	20,3	0,38	α_1	18,8	0,51	α_1	18,4	0,87				
	1	0,01	6,4	0,07	0,01	6,21	0,07	0,05	6,55	0,12	0,05	6,61	0,18				
	1,5	LT	3,58	0,03	LT'	3,53	0,03	LB	3,88	0,04	LB'	3,84	0,07				
	2	0,04	2,51	0,02	0,03	2,48	0,02	3,82	2,75	0,02	3,05	2,76	0,03				
0,3	0,5	α_1	21,1	0,40	α_1	20,3	0,39	α_1	19	0,53	α_1	19	0,86				
	1	0,01	6,59	0,09	0,01	6,4	0,09	0,02	6,52	0,13	0,04	6,56	0,21				
	1,5	LT	3,57	0,03	LT'	3,51	0,03	LB	3,64	0,04	LB'	3,79	0,07				
	2	0,04	2,49	0,02	0,03	2,47	0,02	3,8	2,57	0,02	3,01	2,7	0,04				

A partir de quatro valores de $\alpha \in \{0,045, 0,135, 0,21, 0,3\}$ selecionados da Tabela 3.1, os gráficos na Figura 3.2 foram construídos com o intuito de ilustrar o comportamento de NMA1, para cada estatística. De acordo com essa Figura, conforme Δa diminui, é possível verificar diferenças de NMA1 para qualquer α . Contudo, quando Δa e α aumentam, essas diferenças diminuem consideravelmente.

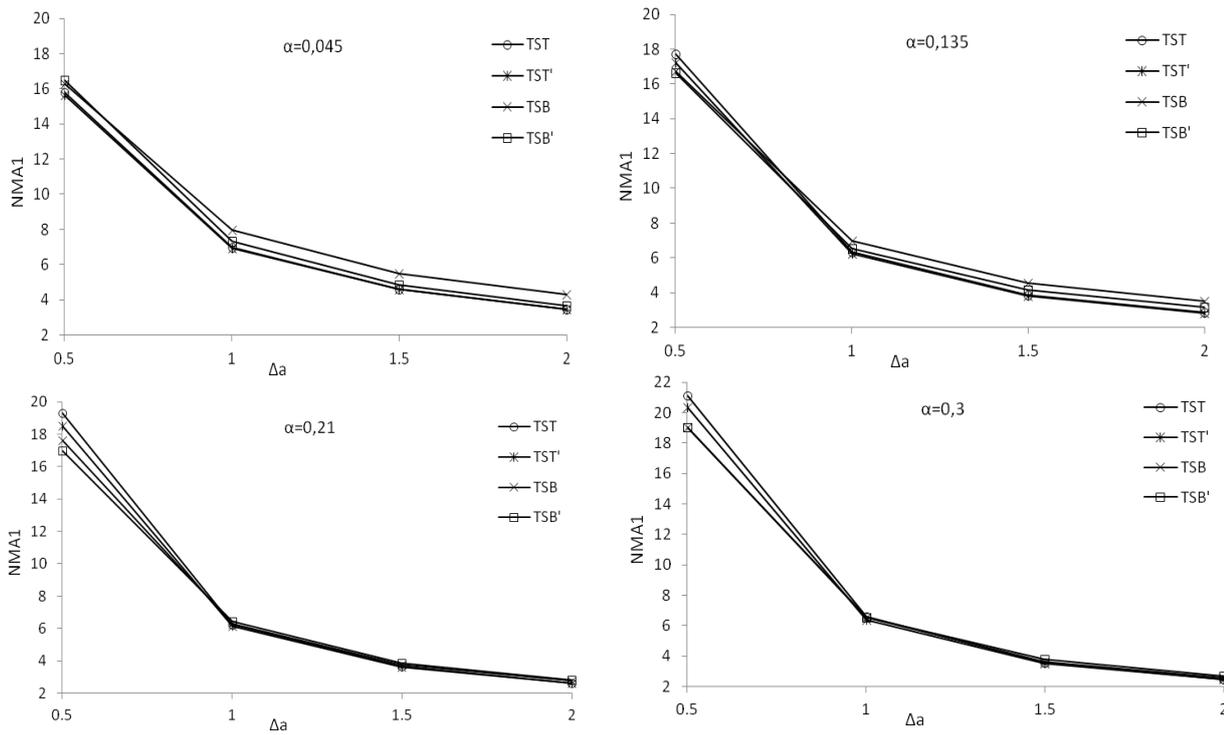


Figura 3.2 – NMA1 para variações Δa e $\alpha \in \{0,045, 0,135, 0,21, 0,3\}$

α	TST	TST'	TSB	TSB'
0,045				
⋮				
0,15				
0,165				
⋮				
0,3				

Melhor desempenho

Figura 3.3 – Estatísticas com melhor desempenho para cada α

Portanto, pode-se utilizar a Figura 3.3, juntamente com a Tabela 3.2, como ferramenta para identificar os parâmetros (α, L) a serem utilizados no monitoramento de suas previsões a partir de um valor de α .

Tabela 3.2 – Parâmetros ($\alpha_1 L$) por estatística em função de α

α	<i>TST</i>		<i>TST'</i>		<i>TSB</i>		<i>TSB'</i>	
	α_1	LT	α_1	LT'	α_1	LB	α_1	LB'
0.05	0.01	0.065	0.01	0.051	0.06	6.887	0.01	5.491
0.06	0.01	0.061	0.01	0.048	0.04	6.411	0.03	5.123
0.08	0.02	0.111	0.01	0.046	0.04	6.032	0.03	4.796
0.09	0.01	0.055	0.01	0.044	0.04	5.726	0.03	4.563
0.11	0.01	0.052	0.01	0.042	0.03	5.456	0.04	4.355
0.12	0.01	0.050	0.01	0.040	0.04	5.234	0.07	4.175
0.14	0.01	0.049	0.01	0.039	0.08	5.023	0.03	4.017
0.15	0.01	0.047	0.01	0.037	0.01	4.875	0.03	3.880
0.17					0.07	4.670	0.07	3.748
0.18					0.02	4.584	0.03	3.633
0.2					0.05	4.420	0.03	3.535
0.21					0.02	4.339	0.03	3.444
0.23					0.04	4.395	0.07	3.337
0.24					0.05	4.086	0.03	3.280
0.26					0.03	4.016	0.03	3.207
0.27					0.07	3.901	0.08	3.110
0.29					0.05	3.822	0.05	3.047
0.3					0.02	3.803	0.04	3.005

A metodologia proposta nesta tese foi utilizada por Jardim (2013) para as mesmas estatísticas *Tracking Signal* aqui estudadas, com o intuito de ampliar esta pesquisa. Jardim considerou outros tipos de vieses, levando em conta também o modelo de demanda linear com previsão por amortecimento exponencial duplo.

4. Modelo de Estoque Proposto

Este capítulo contém a proposta de modelo de estoque para um sistema com duas camadas e revisão periódica. O sistema representa uma situação encontrada no varejo que utiliza um centro de distribuição. Além da formulação proposta para o modelo foi desenvolvido um algoritmo de otimização, testado via simulação para observar sua eficácia. Ao fim deste capítulo encontra-se ainda uma ilustração numérica.

4.1. Descrição do modelo

No sistema de estoque em camadas estudado nesta tese os *lead times* no CD e no varejo i , L_0 e L_i respectivamente, são considerados constantes. O fornecedor externo sempre tem estoque disponível para atender ao CD, condição conhecida como capacidade infinita. Em todas as instalações é permitido estocar itens e não há restrições de capacidade de estocagem e transporte. As demandas nas lojas são variáveis aleatórias identicamente distribuídas e independentes. O varejo i deve atender ao cliente final com um nível de serviço β_i , que nesse caso representa a fração de demanda prontamente atendida pelo estoque na loja, sendo β_i uma meta gerencial.

Ao final de cada intervalo entre revisões T_0 , o CD faz um pedido ao fornecedor externo, em um instante t_0 , visando atingir seu nível máximo de reposição S_0 . No instante em que ocorre a reposição no CD é realizada a primeira revisão no varejo. O intervalo T entre revisões no varejo é igual para todas as lojas e as revisões são sincronizadas. A cada instante t , o CD verifica o nível de estoque em cada varejo i e calcula a quantidade de itens de reposição para que este atinja seu nível máximo S_i . Eventualmente, o CD não terá estoque disponível para atender a essas quantidades, nesse caso, a parcela não atendida é enviada assim que houver reposição no CD.

O custo de manutenção de estoque no sistema é cobrado ao final de cada intervalo T . A Figura 4.1 é uma ilustração da movimentação do nível de estoque de um item no CD e em um determinado varejo i , nesse caso $T_0 = 2T$.

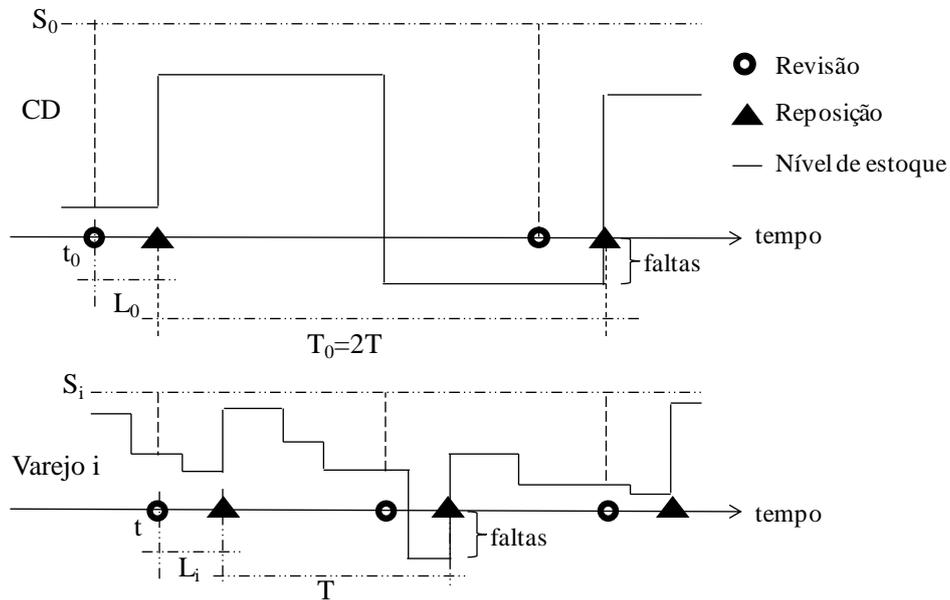


Figura 4.1 – Nível de estoque no CD e no varejo i para $T_0 = 2T$

4.2. Notação

N - número de lojas no varejo;

T - intervalo entre revisões no varejo. Igual em todas as lojas;

m - total de revisões no varejo dentro de um ciclo de reposição do CD;

T_0 - intervalo entre revisões no CD, restrito a $T_0 = mT$, sendo m um inteiro positivo;

μ_i - demanda média no varejo i por unidade de tempo;

σ_i^2 - variância da demanda média no varejo i por unidade de tempo;

h_i, h_0 - custo de manter uma unidade em estoque por unidade de tempo, no varejo e no CD, no intervalo T , respectivamente;

L_i, L_0 - *lead times* no varejo i e no CD, respectivamente;

p_i - fração de faltas no CD alocadas no varejo i ;

S_i, S_0 - níveis máximos de reposição no varejo i e no CD, respectivamente;

$D_i(a)$ - demanda no varejo i durante um intervalo $[0, a)$;

$D_0(a)$ - demanda no CD durante um intervalo $[0, a)$;

B_{ij} - faltas alocadas no varejo i em sua j -ésima revisão;

$$(x)^+ = \max\{x, 0\};$$

$$(x)^- = \max\{-x, 0\}.$$

4.3.

Determinação do estoque médio no CD e no varejo

O estoque médio no CD, assim como em cada varejo i , é usado para calcular o valor esperado do custo de manter o estoque em cada uma dessas instalações. O estoque médio a ser considerado é aquele presente em um ciclo de reposição, que é definido como o intervalo de tempo entre o instante de chegada de uma encomenda e o instante imediatamente anterior à chegada da próxima encomenda. Os níveis de estoque no varejo i nesses dois instantes são respectivamente: $(S_i - D_i(L_i))$ e $(S_i - D_i(L_i + T))$. Em virtude de o nível de estoque poder assumir valores positivos (estoque disponível) ou valores negativos (faltas), apenas a parte positiva desses níveis deve ser considerada para o cálculo do estoque médio no varejo i , que será calculado por

$$\bar{I}_i = \left(E(S_i - D_i(L_i))^+ + E(S_i - D_i(L_i + T))^+ \right) / 2, \quad (4.1)$$

sendo, $E(S_i - D_i(L_i))^+$ e $E(S_i - D_i(L_i + T))^+$ os valores esperados do estoque disponível no início e no final do ciclo de reposição do varejo.

O cálculo do estoque médio no CD segue o mesmo princípio do estoque médio no varejo. Porém, é necessário observar que um ciclo de reposição T_0 contém m ciclos T e, como consequência da política adotada, o nível de estoque no CD sofre alterações somente nas revisões do varejo, ou seja, a cada intervalo T (ver Figura 4.1). Por esse motivo, o cálculo do valor esperado de estoque antes da próxima reposição no CD é obtido por $E(S_0 - D_0(L_0 + (m-1)T))^+$, que se mantém constante até a próxima reposição. O valor esperado de estoque imediatamente após a reposição é calculado por $E(S_0 - D_0(L_0))^+$. Portanto, o estoque médio no CD é obtido por

$$\bar{I}_0 = \left(E(S_0 - D_0(L_0))^+ + E(S_0 - D_0(L_0 + (m-1)T))^+ \right) / 2, \quad (4.2)$$

4.4. Alocação das faltas no varejo

Eventuais faltas no CD provocam atrasos nas entregas do varejo, que nesse caso aguarda um tempo maior que L_i para a reposição de seu estoque. As faltas no CD são distribuídas (ou alocadas) em cada varejo i , ou seja, o CD determina as quantidades que serão entregues com atraso em cada loja do varejo. Portanto, as faltas no CD, alocadas no varejo i , têm relação direta com o *lead time* efetivo $\bar{\ell}_i$ (detalhado na próxima seção). Sendo assim, é necessário conhecer o valor esperado de faltas alocadas em cada varejo i a cada revisão.

Conforme citado, alterações no nível de estoque do CD ocorrem apenas nos instantes de revisão no varejo, logo as faltas no CD são registradas pela parte negativa do nível de estoque nessas ocasiões. Portanto, o valor esperado de faltas no CD até a j -ésima revisão do varejo é calculado por

$$E(S_0 - D_0(L_0 + jT))^- \quad j \in \{0, 1, \dots, m-1\}, \quad (4.3)$$

de forma que para $j=0$ ocorre a primeira revisão do varejo dentro de um ciclo de reposição do CD, e a última revisão ocorre para $j=m-1$.

Considerando um intervalo entre duas revisões consecutivas no varejo, o valor esperado de faltas no CD no início e no final desse intervalo são, respectivamente, $E(S_0 - D_0(L_0 + (j-1)T))^-$ e $E(S_0 - D_0(L_0 + jT))^-$. O valor esperado de faltas na j -ésima revisão do varejo é a diferença entre esses valores esperados. A única exceção é para a primeira revisão no varejo ($j=0$) que coincide com a reposição no CD; nesse caso o valor esperado de faltas nessa revisão é dado por $E(S_0 - D_0(L_0))^-$. Dessa forma, as faltas a cada revisão do varejo são descritas por

$$\begin{aligned} B_{0,j} &= E(S_0 - D_0(L_0 + jT))^- - E(S_0 - D_0(L_0 + (j-1)T))^- \quad j \in \{1, \dots, m-1\} \\ B_{0,0} &= E(S_0 - D_0(L_0))^- \quad \text{para } j=0 \end{aligned},$$

para qualquer valor não-negativo de a $E(S_0 - D_0(a))^- = E(D_0(a) - S_0)^+$, logo,

$$\begin{aligned}
 B_{0j} &= E(D_0(L_0 + jT) - S_0)^+ - E(D_0(L_0 + (j-1)T) - S_0)^+ \quad j \in \{1, \dots, m-1\} \\
 B_{00} &= E(D_0(L_0) - S_0)^+ \quad \text{para } j = 0
 \end{aligned} \quad (4.4)$$

Assim, o valor esperado de faltas alocadas no varejo i em sua j -ésima revisão é obtido por

$$B_{ij} = p_i B_{0j} \quad \text{para } i > 0, \quad (4.5)$$

sendo p_i a fração de faltas alocadas no varejo i quando não há estoque disponível no CD. Nessas situações, é necessário definir a quantidade de cada pedido que será atendida. Essa decisão é chamada racionamento. O racionamento também pode ser interpretado como o ato de alocar as faltas do CD nas lojas do varejo, ou seja, é a fração p_i ($0 \leq p_i \leq 1$) de cada pedido que não será atendida no varejo i , de maneira que $\sum_{i=1}^N p_i = 1$.

A determinação da fração p_i é importante para o desempenho do sistema, pois ela influencia a probabilidade de faltas no varejo i , bem como seu nível de serviço. Além disso, eventualmente o racionamento pode ser feito de tal maneira que o volume alocado de faltas no varejo i seja maior que o pedido realizado ao CD, o que recebe o nome de *imbancing* ou alocação negativa de estoque.

Algumas das propostas mais conhecidas para determinar a fração p_i são: *Consistent Appropriate Share* (CAS) (De Kok, 1990), *Fair Share* (FS) (Eppen & Schrage, 1981) e *Balanced Rationing* (BR) (Heijden, 1997). Essas abordagens se diferenciam no objetivo: em CAS o princípio é manter a mesma probabilidade de faltas em todos os pontos do varejo; o procedimento em FS prioriza atender a meta do nível de serviço, e o fenômeno da alocação negativa é minimizado em BR. Muitas outras propostas surgiram a partir dessas. Heijden *et al.* (1997) propôs uma solução para p_i independente de S_i , diferente das outras propostas que determinam p_i e S_i simultaneamente. Nesse caso, somente a variância da demanda no varejo influencia o racionamento, de forma que:

$$p_i = \frac{1}{2N} + \frac{\sigma_i^2}{2\sum_{j=1}^N \sigma_j^2}. \quad (4.6)$$

Heijden *et al.* (2000) testaram a heurística na Equação (4.6) por meio de simulação juntamente com CAS, FS e BR; a heurística da Equação (4.6) teve o melhor desempenho no atendimento ao nível de serviço no varejo. Por esse motivo, as frações obtidas na Equação (4.6) são usadas nesta tese como regra de racionamento, ou seja, serão usadas para alocar as faltas do CD no varejo.

4.5.

Lead time efetivo no varejo e desagregação do sistema

Em um sistema de estoque com duas camadas, a movimentação do estoque no varejo é influenciada pelos níveis de reposição de estoque adotados no CD. O princípio básico na modelagem de um sistema de estoque com duas camadas é encontrar uma forma de desagregar o sistema em N instalações que possam ser tratadas isoladamente. Assim, uma das formas que permite a decomposição do sistema é determinar o lead time efetivo $\bar{\ell}_i$. O lead time efetivo é influenciado pelo nível de reposição no CD através do atraso nas entregas provocado pelas faltas no CD. Portanto, uma vez determinado $\bar{\ell}_i$ o varejo i pode ser tratado de forma isolada, pois a influência do nível de reposição do CD está sendo levada em conta de forma implícita.

As faltas no CD são uma variável aleatória, como consequência o *lead time* efetivo no varejo i é uma variável aleatória e seu valor esperado é definido pela soma do lead time L_i ao atraso médio nos pedidos \bar{w}_i , ou seja,

$$\bar{\ell}_i = L_i + \bar{w}_i, \quad (4.7)$$

sendo \bar{w}_i o valor esperado do atraso no pedido do varejo i .

Em virtude da política de suprimento adotada, as reposições acontecem em pontos definidos no tempo. Logo, o atraso no pedido é uma variável aleatória discreta.

Dessa forma, tomando com referência a Figura 4.1, caso ocorram faltas na primeira revisão do varejo i em t , as unidades em falta levariam um tempo total

$L_i + mT$, nesse caso $L_i + 2T$, até serem entregues, ou seja, com um atraso de mT . Em caso de faltas na segunda revisão ($t + T$) esse tempo é reduzido para $L_i + (m-1)T$, nesse caso, tem-se um atraso em $(m-1)T$. Conclui-se que os atrasos possíveis de reposição no varejo i ocorrem em $(m-j)T$ sendo $j \in \{0, 1, \dots, m-1\}$.

Os atrasos ocorrem somente quando ocorrem faltas no CD. Nesse caso, de todas as unidades pedidas ao CD em um ciclo T_0 apenas B_{ij} estarão em falta no varejo i na revisão j . Portanto, a fração da demanda em falta na j -ésima revisão é definida pela fração $B_{ij}/\mu_i mT$, sendo $B_{ij} = p_i B_{0j}$ como na Equação (4.5) e μ_i o valor esperado da demanda no varejo por unidade de tempo.

Portanto, o valor esperado do atraso \bar{w}_i é definido pelo somatório

$$\bar{w}_i = \sum_{j=0}^{m-1} [(m-j)T] B_{ij} / \mu_i mT. \quad (4.8)$$

Uma vez conhecido $\bar{\ell}_i$ (Equação (4.7)) é possível tratar o sistema como $N+1$ instalações independentes, para isso basta utilizar $\bar{\ell}_i$ no lugar de L_i . Logo, o estoque médio no varejo i , calculado na Equação (4.1), pode ser reescrito como

$$\bar{I}_i = \left(E(S_i - D_i(\bar{\ell}_i))^+ + E(S_i - D_i(\bar{\ell}_i + T))^+ \right) / 2.$$

Sabendo que, para qualquer valor não-negativo a , $E(S_i - D_i(a))^+ = E(D_i(a) - S_i)^-$, logo

$$\bar{I}_i = \left(E(D_i(\bar{\ell}_i) - S_i)^- + E(D_i(\bar{\ell}_i + T) - S_i)^- \right) / 2.$$

Além disso, $E(D_i(a) - S_i)^- = E(D_i(a) - S_i)^+ - E(D_i(a) - S_i)$, que substituindo na expressão anterior vem a ser

$$\bar{I}_i = \left(E(D_i(\bar{\ell}_i) - S_i)^+ - E(D_i(\bar{\ell}_i) - S_i) + E(D_i(\bar{\ell}_i + T) - S_i)^+ - E(D_i(\bar{\ell}_i + T) - S_i) \right) / 2.$$

Sabendo que $E(D_i(a)) = \mu_i a$ segue que $E(D_i(a) - S_i) = \mu_i a - S_i$. Portanto, a expressão anterior pode ser reescrita como

$$\bar{I}_i = \left(E(D_i(\bar{\ell}_i) - S_i)^+ + E(D_i(\bar{\ell}_i + T) - S_i)^+ + 2S_i - \mu_i(2\bar{\ell}_i + T) \right) / 2. \quad (4.9)$$

Analogamente, o mesmo desenvolvimento pode ser feito para reescrever o estoque médio no CD (Equação (4.2)), que nesse caso resulta em

$$\bar{I}_0 = \left(E(D_0(L_0) - S_0)^+ + E(D_0(L_0 + (m-1)T) - S_0)^+ + 2S_0 - \mu_0(2L_0 + (m-1)T) \right) / 2. \quad (4.10)$$

O nível de serviço no varejo, aqui denotado por β_i , é a fração de demanda prontamente atendida (*fill rate*). β_i também deve considerar o lead time efetivo $\bar{\ell}_i$. Para o cálculo do *fill rate* é preciso conhecer as faltas em um ciclo de reposição no varejo, que nesse caso são calculadas pela diferença entre o valor esperado das faltas no início e no final desse ciclo, que são $E(S_i - D_i(\bar{\ell}_i))^-$ e $E(S_i - D_i(\bar{\ell}_i + T))^-$ respectivamente. A demanda esperada em um ciclo é definida por $\mu_i T$. Logo a fração de demanda não atendida é calculada por $\left(E(S_i - D_i(\bar{\ell}_i + T))^- - E(S_i - D_i(\bar{\ell}_i))^- \right) / \mu_i T$. Portanto, a fração de demanda prontamente atendida vem a ser

$$\beta_i = 1 - \left(E(S_i - D_i(\bar{\ell}_i + T))^- - E(S_i - D_i(\bar{\ell}_i))^- \right) / \mu_i T,$$

que levando em conta que $E(S_i - D_i(a))^- = E(D_i(a) - S_i)^+$, se torna

$$\beta_i = 1 - \left(E(D_i(\bar{\ell}_i + T) - S_i)^+ - E(D_i(\bar{\ell}_i) - S_i)^+ \right) / \mu_i T. \quad (4.11)$$

4.6. Formulação para demanda normalmente distribuída

A distribuição normal é uma escolha comum para aproximar a demanda de itens, sobretudo quando essa demanda é elevada. A demanda de um item é expressa por intervalo de tempo, por exemplo, semana, que por sua vez é composta pela soma de demandas em uma unidade de tempo menor, por exemplo, dias. Segundo o Teorema do Limite Central, a soma de várias variáveis aleatórias independentes tem distribuição aproximadamente normal, o que justifica o uso dessa distribuição em muitas situações.

Segundo Axsäter (2006), para a demanda ser aproximada por uma normal o coeficiente de variação $cv = \sigma/\mu$ deve ser razoavelmente menor que um, o que diminui a probabilidade de a distribuição conter valores negativos de demanda. Silver *et al* (1998) recomendam como regra o uso da normal quando $cv \leq 0,5$.

4.6.1. Determinação do nível de estoque para demanda normal

No desenvolvimento desta seção será utilizado um resultado conhecido na literatura como *Loss Function* (para maiores detalhes, consultar Zipkin, 2000). Para um determinado valor S de uma variável aleatória X , com média μ_x , com desvio-padrão σ_x e função densidade $f(x)$, a perda de primeira ordem (*Loss Function*) é calculada por $E(X - S)^+ = \int_S^{\infty} (y - S)f(y)dy = F^1(S)$. Sendo que a perda $F^1(S)$ pode ser expressa por uma distribuição normal padronizada como

$$F^1(S) = \sigma_x \Phi^1(z), \quad (4.12)$$

pode-se mostrar que

$$\Phi^1(z) = \phi(z) - z[1 - \Phi(z)], \quad (4.13)$$

sendo $\Phi^1(z)$ a função de perda de primeira ordem, $\phi(z)$ a função densidade e $\Phi(z)$ a função de distribuição cumulativa da normal padronizada, todas definidas em $z = (S - \mu_x)/\sigma_x$.

Para o intervalo a , tem-se que a demanda tem média μa e desvio-padrão $\sigma\sqrt{a}$. Dessa forma, podemos reescrever as equações (4.4), (4.9), (4.10) e (4.11), levando em consideração a função de perda dada por $E(D(a) - S)^+ = \sigma\sqrt{a}\Phi^1(z)$, para qualquer valor não-negativo de a , sendo $z = (S - \mu a)/\sigma\sqrt{a}$.

Assim, a Equação (4.4) passa a ser

$$B_{0j} = \Phi^1(z)\sigma_0\sqrt{L_0 + jT} - \Phi^1(z)\sigma_0\sqrt{L_0 + (j-1)T} \quad j \in \{1, \dots, m-1\}$$

$$B_{00} = \Phi^1(z)\sigma_0\sqrt{L_0} \quad \text{para } j = 0$$

ou ainda

$$B_{0j} = \Phi^1(z)\sigma_0\left(\sqrt{L_0 + jT} - \sqrt{L_0 + (j-1)T}\right) \quad j \in \{1, \dots, m-1\}$$

$$B_{00} = \Phi^1(z)\sigma_0\sqrt{L_0} \quad \text{para } j = 0 \quad . \quad (4.14)$$

A Equação (4.9) passa a ser

$$\bar{I}_i = \left(\sigma_i\sqrt{\bar{\ell}_i}\Phi^1(z) + \sigma_i\sqrt{\bar{\ell}_i + T}\Phi^1(z) + 2S_i - \mu_i(2\bar{\ell}_i + T)\right)/2 ,$$

ou

$$\bar{I}_i = \left(\Phi^1(z)\sigma_i\left(\sqrt{\bar{\ell}_i} + \sqrt{\bar{\ell}_i + T}\right) + 2S_i - \mu_i(2\bar{\ell}_i + T)\right)/2 . \quad (4.15)$$

A Equação (4.10) passa a ser

$$\bar{I}_0 = \left(\sigma_0\sqrt{L_0}\Phi^1(z) + \sigma_0\sqrt{L_0 + (m-1)T}\Phi^1(z) + 2S_0 - \mu_0(2L_0 + (m-1)T)\right)/2$$

ou

$$\bar{I}_0 = \left(\Phi^1(z)\sigma_0\left(\sqrt{L_0} + \sqrt{L_0 + (m-1)T}\right) + 2S_0 - \mu_0(2L_0 + (m-1)T)\right)/2 . \quad (4.16)$$

A Equação (4.11) passa a ser

$$\beta_i = 1 - \left(\Phi^1(z) \sigma_i \sqrt{\bar{\ell}_i + T} - \Phi^1(z) \sigma_i \sqrt{\bar{\ell}_i} \right) / \mu_i T$$

ou

$$\beta_i = 1 - \left(\Phi^1(z) \sigma_i \left(\sqrt{\bar{\ell}_i + T} - \sqrt{\bar{\ell}_i} \right) \right) / \mu_i T. \quad (4.17)$$

Nas expressões anteriores μ_i e σ_i representam a média e o desvio padrão da demanda no varejo i em uma unidade de tempo. Analogamente, μ_0 e σ_0 são a média e o desvio padrão da demanda por unidade de tempo no CD,

de forma que $\mu_0 = \sum_{i=1}^N \mu_i$ e $\sigma_0^2 = \sum_{i=1}^N \sigma_i^2$.

4.6.2. Função de custo e o problema de otimização

A reposição dos estoques no CD e nas lojas será modelado como um problema de otimização. Na função de custo, o único custo considerado é o custo de manutenção de estoque, que é cobrado ao final de cada ciclo de reposição no varejo. Quanto ao nível de serviço, são considerados apenas os níveis de serviço nas lojas, que representam a fração atendida da demanda do cliente final. Assim, a única restrição do problema diz respeito ao nível de serviço no varejo, que deve atingir um parâmetro β_i definido. Sejam então h_i e h_0 os custos de manter uma unidade em estoque, no varejo e no CD respectivamente, durante o período T , o custo total de manter o estoque no sistema durante esse período é dado por $c = h_0 \bar{I}_0 + \sum_{i=1}^N h_i \bar{I}_i$, sendo \bar{I}_i e \bar{I}_0 calculados nas equações (4.15) e (4.16).

Deseja-se determinar os níveis de reposição S_i e S_0 que atendam a meta de nível de serviço no varejo ao menor custo no sistema, o que resulta no problema de minimização

$$\begin{aligned} \min \quad & c = h_0 \bar{I}_0 + \sum_{i=1}^N h_i \bar{I}_i \\ \text{sujeito a} \quad & , \\ & 1 - \left(\Phi^1(z) \sigma_i \left(\sqrt{\bar{\ell}_i + T} - \sqrt{\bar{\ell}_i} \right) \right) / \mu_i T = \beta_i \quad i \in \{1, 2, \dots, N\} \end{aligned} \quad (4.18)$$

O custo c é uma combinação linear entre os estoques médios I_0 e I_i . Os estoques médios I_0 e I_i são obtidos a partir dos valores esperados de estoque, que nesse caso são obtidos a partir de funções de perda de primeira ordem para a normal (ver Seção 4.6.1.). Segundo Axsäter (2006), as funções de perda são contínuas, decrescentes e convexas. Portanto, a função objetivo do Problema (4.18) resume-se a uma combinação de funções convexas, o que resulta, segundo Bazaraa *et al.* (2006), em uma função convexa.

A expressão utilizada para calcular o nível de serviço na Equação (4.17) também é convexa, pois utiliza o valor esperado de faltas obtido a partir da função de perda. Portanto, uma vez calculado $\bar{\ell}_i$, existe somente um z que satisfaz a restrição do Problema (4.18). Esse valor de z , que é função de S_i , pode ser encontrado por um método de busca aplicado na Equação (4.17), como por exemplo, o método da secante.

Além disso, o lead time efetivo $\bar{\ell}_i$ está relacionado com S_0 através das faltas B_{0j} . Conclui-se então, que para cada S_0 existe apenas um S_i que atende a restrição do problema. Portanto, o custo c é função apenas de S_0 , pois os valores de S_i são consequência da restrição do problema. Portanto, o custo c é convexo e possui apenas um argumento.

Nesta tese, propõe-se uma solução heurística para o Problema (4.18). Devido às características da função de custo, é possível utilizar um método de busca para encontrar o custo mínimo, optou-se então por um algoritmo (apresentado a seguir) baseado no método da seção áurea. Conforme esse algoritmo, cada solução viável do custo c é calculada em duas etapas. Na primeira etapa, a partir de um dado nível de reposição S_0 , são calculados: o estoque médio I_0 , as faltas B_{0j} e os *lead times* efetivos $\bar{\ell}_i$. A segunda etapa se inicia encontrando, através de uma busca na Equação (4.17), os níveis de reposição S_i que atendem as restrições do problema; em seguida, são calculados os estoques médios I_i e por fim o custo c . A determinação dos níveis S_0 em cada iteração do algoritmo segue a lógica do método da seção áurea.

4.6.3. Algoritmo de otimização

O algoritmo desenvolvido é baseado no algoritmo da seção áurea (Bazaraa *et al.*, 2006). Esse método necessita de um intervalo de busca, e a convergência para o ótimo ocorre quando a diferença entre dois pontos testados é menor que um critério de parada δ adotado. Na prática os valores de S_0 são discretos, logo, é razoável que o menor intervalo testado tenha um comprimento mínimo equivalente a uma unidade, ou seja, $\delta = S_{0(1)}^k - S_{0(2)}^k \geq 1$ sendo k o número da iteração e $S_{0(1)}^k$ e $S_{0(2)}^k$ os pontos testados.

Além disso, o método da seção áurea é iniciado a partir de limites extremos de busca para S_0 , esses limites são desenvolvidos no Apêndice B, são eles:

$$\mu_0(L_0 - T) < S_0 < 5\sqrt{L_0 + (m-1)T}\sigma_0 + \mu_0[L_0 + (m-1)T]. \quad (4.19)$$

Segue o algoritmo baseado no método da seção áurea:

Dados: $\mu_i, \sigma_i, L_i, T, h_i, \beta_i, \mu_0, \sigma_0, L_0, T_0, h_0, m, p_i$.

Faça

$$A^1 = \mu_0(L_0 - T);$$

$$B^1 = 5\sqrt{L_0 + (m-1)T}\sigma_0 + \mu_0[L_0 + (m-1)T];$$

$$\delta = A^1 - B^1; k = 1 \text{ (número da iteração);}$$

Calcule as estimativas iniciais de $S_{0(1)}^k$ e $S_{0(2)}^k$ para $k = 1$:

$$S_{0(1)}^1 = A^1 + (1 - 0,618)(B^1 - A^1) \text{ e } S_{0(2)}^1 = A^1 + 0,618(B^1 - A^1)$$

Enquanto $\delta > 1$ faça:

Para cada opção $\{S_{0(1)}^k, S_{0(2)}^k\}$ faça:

Calcule o estoque médio no CD $I_{0(1)}^k$ e $I_{0(2)}^k$ como na Equação (4.16)

Para cada revisão no varejo $j \in \{0, 1, \dots, m-1\}$ faça:

Calcule as faltas no CD $B_{0j(1)}^k$ e $B_{0j(2)}^k$ como nas Equações (4.14)

fim

para cada varejo $i \in \{1, 2, \dots, N\}$ faça:

Obtenha os lead times $\bar{\ell}_{i(1)}^k$ e $\bar{\ell}_{i(2)}^k$ como na Equação (4.7)

Com um método de busca, partindo de $S_{0(1)}^k$ e $S_{0(2)}^k$ defina $z_{i(1)}^k$ e $z_{i(2)}^k$

para β_i na Equação (4.17)

Calcule os estoques médios $I_{i(1)}^k$ e $I_{i(2)}^k$ como na Equação (4.15)

fim

Calcule os custos $c_{(1)}^k$ e $c_{(2)}^k$ como no Problema (4.18)

Se ($c_{(1)}^k > c_{(2)}^k$) então

$$A^{k+1} = S_{0(1)}^k; B^{k+1} = B^k;$$

$$S_{0(1)}^{k+1} = A^{k+1} + (1 - 0,618)(B^{k+1} - A^{k+1}); S_{0(2)}^{k+1} = A^{k+1} + 0,618(B^{k+1} - A^{k+1});$$

$$c \min = c_{(2)}^k, S_0^* = S_{0(2)}^k \text{ e } S_i^* = S_{i(2)}^k.$$

Ou se ($c_{(1)}^k \leq c_{(2)}^k$) então

$$A^{k+1} = A^k; B^{k+1} = S_{0(2)}^k;$$

$$S_{0(2)}^{k+1} = A^{k+1} + 0,618(B^{k+1} - A^{k+1}); S_{0(1)}^{k+1} = A^{k+1} + (1 - 0,618)(B^{k+1} - A^{k+1});$$

$$c \min = c_{(1)}^k, S_0^* = S_{0(1)}^k \text{ e } S_i^* = S_{i(1)}^k.$$

fim

$$\delta = A^k - B^k; k = k + 1.$$

fim

4.6.4. Ilustração numérica e simulação

Para testar o método heurístico proposto, foi elaborado um exemplo de uma rede de suprimentos de duas camadas, considerando a reposição de um determinado produto em um sistema com três lojas no varejo ($N = 3$), abastecidas por um CD. A revisão dos níveis de estoques nas lojas é feita pelo CD a cada dia ($T = 1$). O nível de estoque no CD é verificado a cada três dias ($T_0 = 3$), para fazer pedidos ao fornecedor externo. No varejo os valores de média e variância da demanda diária são: $\mu_1 = 27; \sigma_1^2 = 23$; $\mu_2 = 81; \sigma_2^2 = 39$; $\mu_3 = 54; \sigma_3^2 = 31$. Outros dados do problema: $L_0 = 1; h_0 = 1; L_i = 1; h_i = 4; \beta_i = 0,9$.

No algoritmo da Seção 4.6.3 a cada iteração é gerado um conjunto de valores obtidos a partir de referências de S_0^1 e S_0^2 . Os valores obtidos em algumas iterações estão reunidos na Tabela 4.1 onde se observa a diminuição do custo c^* a cada iteração. O nível de reposição S_0^* também diminui, tornando maiores as faltas no CD, o que produz *lead times* efetivos $\bar{\ell}_i$ maiores. Como resultado, S_1^* , S_2^* e S_3^* aumentam para atingir a meta de nível de serviço $\beta_i = 0,9$. Ao término das iterações o algoritmo encontrou como solução o custo mínimo $c^* = R\$ 329,79$ para os níveis de reposição: $S_0^* = 153$, $S_1^* = 106$, $S_2^* = 220$ e $S_3^* = 162$.

Tabela 4.1- Resultado das iterações utilizando o algoritmo da Seção 4.6.3.

	A	B	S_0^*	S_1^*	S_2^*	S_3^*	c^*
It. 1	1.335,84	2.161,12	1.335,84	54,01	154,96	104,20	1.324,96
It. 4	315,29	510,08	315,29	72,43	177,97	124,75	395,64
It. 7	74,41	120,39	120,39	115,84	232,38	173,33	332,25

Os gráficos na Figura 4.3 foram elaborados com o intuito a concavidade da função de custo c e dos níveis de reposição nas lojas S_1, S_2, S_3 em função dos níveis de reposição S_0 adotados no CD. Como esperado, o aumento de S_0 provoca redução nos níveis de reposição S_i , consequência da redução no *lead time* efetivo. Porém, a redução de $\bar{\ell}_i$ é limitada ao próprio lead time L_i e a partir deste ponto S_i se torna fixo mesmo que S_0 continue aumentando. Além disso, é possível observar o comportamento convexo da função de custo.

Uma simulação foi realizada com o intuito de observar como os valores de S_i^* e S_0^* , fornecidos pelo algoritmo proposto (Seção 4.6.3.), influenciam no comportamento do *fill rate* em relação à meta estabelecida β_i . Foram simulados três casos diferentes de demanda no sistema do exemplo citado. A simulação foi realizada através da rotina `simulaEstoque` (APÊNDICE C) programada em MATLAB. Em cada caso simulado, foram geradas 10.000 demandas diárias, obtidas a partir do gerador de números aleatório padrão do MATLAB.

A partir dos dados do problema e das demanda geradas, para cada dia, em cada loja, foi calculada a parcela da demanda não atendida (caso a demanda fosse maior que o estoque disponível). Ao fim de cada dia, foi calculada a

necessidade de reposição em cada loja. Nos casos em que o CD não dispunha de estoque suficiente, foi utilizado o racionamento para definir as unidades em atraso e as quantidades de reposição em cada loja. As unidades em atraso tiveram suas reposições programadas assim que o CD recebia itens de seu fornecedor. No fim da simulação, foi calculado o nível de serviço em cada loja = $1 - (\text{total de demanda não atendida} / \text{demanda total})$. Os resultados estão reunidos na Tabela 4.2.

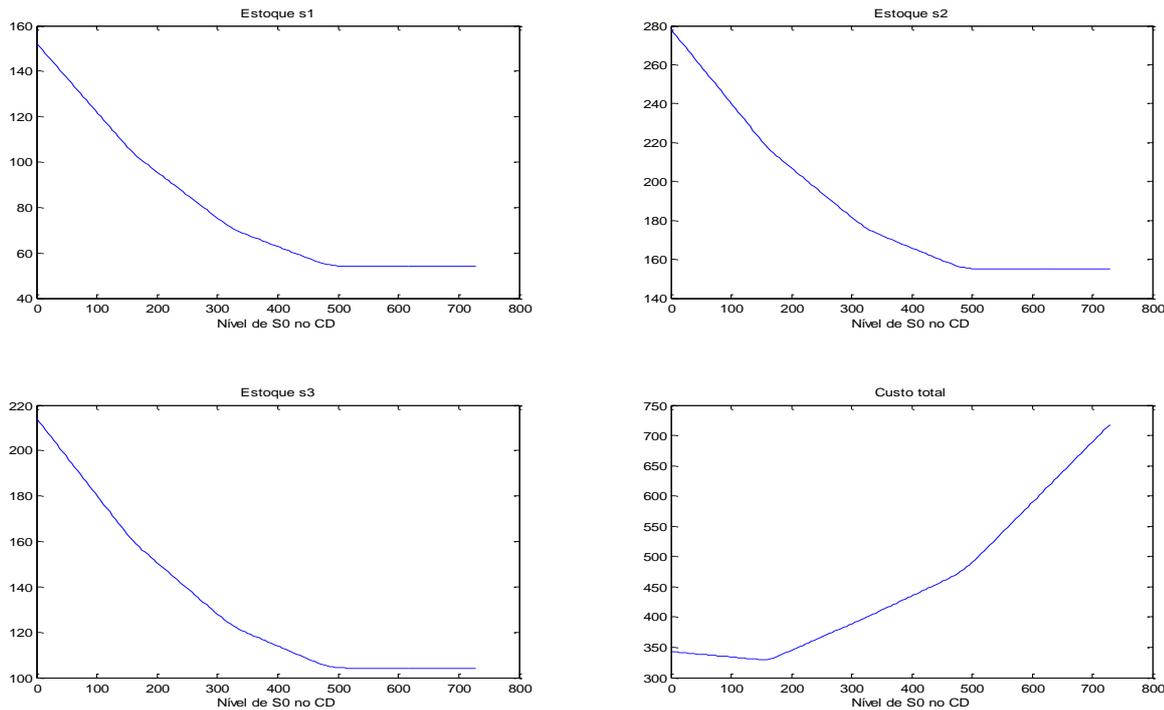


Figura 4.3 – Valores de S_1 , S_2 , S_3 e c para $S_0 \in \{1, 2, \dots, 729\}$.

Conforme observado na Tabela 4.2, a solução proposta pelo algoritmo produz um *fill rate* próximo a meta definida, sobretudo quando as variâncias das demandas no varejo são similares. Porém, à medida que existem disparidades entre variâncias a meta de *fill rate* não é atendida (valores em negrito na Tabela 4.2). Em virtude da regra de racionamento na Equação (4.6) as instalações do varejo com maior variância na demanda têm mais faltas alocadas pelo CD, como consequência o *fill rate* sofre uma defasagem, e para aquelas com variância menor ocorre o contrário a meta de nível de serviço é superada.

Tabela 4.2 – Resultados da simulação

Demanda no varejo		Níveis de reposição				meta	β observado (%)		
caso 1		S_0^*	S_1^*	S_2^*	S_3^*	β	β_1	β_2	β_3
$\mu_1 = 27$	$\sigma_1^2 = 22$	75	79	80	80	80%	85,13	85,38	86,14
$\mu_1 = 27$	$\sigma_1^2 = 23$	78	83	84	84	90%	91,24	91,56	91,97
$\mu_1 = 27$	$\sigma_1^2 = 24$	80	86	87	88	95%	94,97	95,30	95,94
		84	94	94	94	99%	98,83	98,94	99,01
caso 2		S_0^*	S_1^*	S_2^*	S_3^*	β	β_1	β_2	β_3
$\mu_1 = 81$	$\sigma_1^2 = 39$	185	223	167	194	80%	80,25	88,59	82,86
$\mu_2 = 54$	$\sigma_2^2 = 31$	190	231	173	201	90%	85,33	96,52	89,68
$\mu_3 = 67$	$\sigma_3^2 = 35$	193	237	177	206	95%	89,37	98,44	94,17
		196	247	186	215	99%	94,18	99,83	98,36
caso 3		S_0^*	S_1^*	S_2^*	S_3^*	β	β_1	β_2	β_3
$\mu_1 = 27$	$\sigma_1^2 = 23$	149	102	212	156	80%	99,32	78,21	83,68
$\mu_2 = 81$	$\sigma_2^2 = 39$	153	106	220	162	90%	99,84	81,37	88,28
$\mu_3 = 54$	$\sigma_3^2 = 31$	156	110	226	167	95%	99,97	84,58	92,13
		160	116	235	175	99%	99,99	88,74	97,40

5. Conclusões

No que tange o monitoramento do erro das previsões esta tese atingiu seu objetivo de identificar, em uma amostra de uma série temporal conhecida, quais estatísticas *Tracking Signal* (TST, TST', TSB e TSB') são preferíveis em determinadas faixas de valores de α (usado no amortecimento exponencial simples) para um dado valor de $NMA0$. Ainda que as diferenças de desempenhos sejam pequenas, as estatísticas TSB e TSB' apresentaram melhor desempenho na maioria dos casos estudados. Nas demais situações todas as estatísticas foram equivalentes. O desempenho na detecção do viés das estatísticas testadas tende a ser equivalente à medida que se aumenta a magnitude da mudança do comportamento da demanda.

As substituições do DAM pela raiz do EQM nas estatísticas TSB e TST , que resultaram nas estatísticas TSB' e TST' , não apresentaram diferença significativa de desempenho.

Apesar de se utilizar o mesmo recurso na determinação de $NMA0$ que em Gardner (1983, 1985), uma importante diferença foi introduzida nesta tese: as estimativas iniciais de EQM_0, DAM_0, \hat{x}_0 e \hat{a}_0 são obtidas a partir de uma amostra conhecida, o que torna o procedimento mais próximo de uma situação prática.

Sugere-se que estudos futuros abordem a detecção do viés em modelos com sazonalidade, bem como outros métodos de previsão de séries temporais, como, por exemplo, os modelos ARIMA, que também devem ser analisados sob a ótica do monitoramento do viés. O parâmetro de amortecimento α_1 pode ser mais estudado, pois para TST , McClain (1988) sugere que diferentes parâmetros de amortecimento no numerador e no denominador produzem melhores resultados.

Quanto ao modelo de estoque proposto, a partir da simulação realizada conclui-se que o modelo é eficaz. Diferente de outras soluções encontradas na literatura, o modelo proposto considera revisões em intervalos iguais e sincronizados nas lojas, que podem ter seus pedidos atendidos parcialmente

pelo CD. Ademais, a formulação proposta para o modelo supõe que as distribuições da demanda nas lojas do varejo e no CD sejam independentes e identicamente distribuídas com distribuições normais. Porém, esse modelo pode ser desenvolvido para outras distribuições, como por exemplo, gama e Poisson. Para esse fim, é necessário que o valor esperado $E(D(a) - S)^+$ seja escrito para essas distribuições a exemplo do que foi feito para a normal.

A função de custo formulada para o sistema estudado é convexa, e para cada nível de reposição adotado no CD existe apenas um nível de reposição correspondente em cada instalação do varejo que atende as metas de nível de serviço. Portanto, a função de custo além de convexa possui um único argumento. Essas duas características permitiram o desenvolvimento de uma heurística de solução em duas etapas.

O *lead time* efetivo obtido a partir da metodologia proposta permite desagregar o sistema estudado em N problemas de única instalação, bastando para isso substituir o *lead time* pelo *lead time* efetivo tornando a obtenção de S_i um problema isolado em cada instalação do varejo.

Para os casos simulados, conclui-se que a solução obtida pelo algoritmo proposto, de forma geral, resulta em um nível de serviço próximo do estipulado, chegando até a superar a meta em alguns casos. Porém, à medida que as variâncias das demandas no varejo se tornam muito diferentes, instalações com maior variância na demanda tendem a ficar abaixo da meta de nível de serviço. Por outro lado, nas instalações com variâncias menores as metas são superadas.

A solução heurística proposta, representa uma alternativa de solução para o modelo de estoque proposto. A partir dessa solução, é possível determinar os níveis de reposição de um item em uma rede de suprimentos com duas camadas (CD e lojas), em que o CD centraliza as decisões relacionadas à reposição. Por se tratar de um sistema comumente observado no varejo, a solução proposta representa uma contribuição relevante na teoria dos estoques.

Sugere-se para estudos futuros explorar, a partir das mesmas hipóteses consideradas nesta tese, abordar sistemas com maior complexidade, seja na dimensão da quantidade de itens (reposição de múltiplos itens), seja na dimensão do número de camadas (problemas em rede). Essas considerações tornarão o modelo ainda mais próximo de situações observadas no cotidiano das redes de suprimento do varejo.

6. Referências Bibliográficas

Axsäter, S. **Inventory Control**. 2th ed. New York: Springer, 2006.

Axsäter, S; Marklund, J. **Optimal position-based warehouse ordering in divergent two-echelon inventory systems**. *Operations Research*, vol. 56, 4, pp. 976–991, 2008.

Azevedo, C. R.; Araújo, O. **Ruptura: causas e impactos na cadeia de abastecimento e no comportamento do consumidor**. Congresso ECR América Latina. 2004. Disponível em: <
<http://www.ecrbrasil.com.br/ecrbrasil/page/buscabiblioteca.asp?palavra=rupturas&midia=0&pagina=1>> Acesso em: 02/08/2011.

Bakal, I.S.; Erkip, N.; Güllü, R. **Value of supplier's capacity information in a two-echelon supply chain**. *Ann Oper Res*, vol. 191, pp.115–135, 2011.

Ballou, R. H. **Gerenciamento da cadeia de suprimentos: planejamento, organização e logística empresarial**. 4ª ed. Porto Alegre: Bookman, 2004.

Batty, M. **Monitoring an Exponential Smoothing Forecasting System**. *Operational Research Quarterly*. Vol. 20, 3, 319-325, 1969.

Batty, M. **Monitoring an Exponential Smoothing Forecasting System**. *Operational Research Quarterly*, 20, 319-325, 1969.

Bazaraa, M.S. Sherali, H. D.; Shetty, C. M. **Nonlinear Programming: Theory and Algorithms**. 3th ed. New Jersey: Wiley-Interscience, 2006.

BEA. Bureau of Economic Analysis. **Gross Domestic Product: First Quarter (Third Estimate)**. 2011. Disponível em:<

http://www.bea.gov/newsreleases/national/gdp/2011/gdp1q11_adv.htm
Acesso: em 02/08/2011.

BNDES. Banco Nacional do Desenvolvimento. **Comércio Varejista: Supermercados. 1996.** Disponível em:<
http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/relato/supmerca.pdf> Acesso: 02/08/2011.

Bowersox, D. J.; Closs, D.; Cooper, M. **Gestão da Cadeia de Suprimentos e Logística. 2ª ed.** Rio de Janeiro: Elsevier, 2004.

Brown, R. G. **Smoothing, Forecasting, and Prediction of Discrete Time Series.** Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.

Cachon, G.P.; Fisher, M. **Supply chain inventory management and the value of shared Information.** Management Science, vol. 46, 8, pp. 1032–1048, 2000.

Chu, L.Y.; Shen, Z.M. **A Power-of-Two ordering policy for one-warehouse multiretailer systems with stochastic demand.** Operations Research, vol. 58, 2, pp. 492–502, 2010.

Cohen, J.; Garman, S.; Gorr, W. **Empirical calibration of time series monitoring methods using receiver operating characteristic curves.** International Journal of Forecasting, 25, 484–497, 2009.

De Kok, A.G. **Hierarchical production planning for consumer goods.** European Journal of Operational Research, vol.45, pp. 55-69, 1990.

ECRBRASIL. **O jogo perde-perde.** Disponível em:<
<http://www.ecrbrasil.com.br/ecrbrasil/page/buscabiblioteca.asp?palavra=rupturas&midia=0&pagina=1>, 2008> Acesso em: 02/08/2011.

Editorial. **Introduction to time series monitoring.** International Journal of Forecasting, 25, 463–466, 2009.

Eppen, G.; Schrage, L. **Centralized ordering policies in a multi-warehouse system with lead times and random demand.** TIMS Studies in the Management Sciences, vol. 16, pp. 51-67, 1981.

Gardner JR, E. S. **Automatic Monitorign of Forecasting Errors.** International Journal of Forecasting, 2, 1-21, 1983).

Gardner JR, E. S. **CUSUM vs Smoothed-Error Forecast Monitoring Schemes: Some Simulation Results.** Operational Research Society, 36, 43-47, 1985.

Heijden, M.C. **Analysing divergent logistic networks with local (R, S) inventory control.** International Journal of Production Economics, vol. 27, pp. 187-219, 1992.

Heijden, M.C. **Near cost-optimal inventory control policies for divergent networks under fill rate constraints.** Int. J. Production Economics. vol. 63, pp. 161-179, 2000.

Heijden, M.C. **Supply rationing in multi-echelon divergent systems.** European Journal of Operational Research. vol. 101, pp. 532-549, 1997.

Heijden, M.C.; De Kok, A.G. **Customer waiting times in an (R, S) inventory system with compound Poisson demand.** Methods and Models of Operations Research, vol. 36, pp. 315-332, 1992.

Heijden, M.C.; Diks, E.B.; De Kok, A.G. **Stock allocation in general multi-echelon distribution systems with (R, S) order-up-to-policies.** International Journal of Production Economics. vol. 49, pp.157-174, 1997.

Howard, C.; Marklund, J. **Evaluation of stock allocation policies in a divergent inventory system with shipment consolidation.** European Journal of Operational Research. Vol. 211, pp. 298–309, 2011.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Tabela 1 - Dados comparativos das empresas comerciais, segundo as divisões, grupos e classes de atividades - Brasil - 2007-2008. 2008b.** Disponível

em:<

http://www.ibge.gov.br/home/estatistica/economia/comercioeservico/pac/2008/tabelas_pdf/parte1_tab1.pdf> Acesso: em 02/08/2011.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Tabela 9 - Unidades locais, pessoal ocupado total e assalariado em 31.12, salários e outras remunerações e salário médio mensal, segundo faixas de pessoal ocupado total e seção da classificação de atividades - Brasil**

- **2007-2008.** 2008a. Disponível em:<
http://www.ibge.gov.br/home/estatistica/economia/cadastroempresa/2008/tabelas_pdf/tabela9.pdf> Acesso: em 02/08/2011.

Jardim, F.S. **Técnicas para Detecção de Viés em Previsão de Demanda – Comparação de Desempenhos.** Dissertação de Mestrado da Engenharia Industrial da PUC-RJ , 2013.

Juracy, P. **Varejo no Brasil: Gestão e Estratégia.** 1a edição. São Paulo: Atlas, 2000.

Kotler,P.; Armstrong, G. **Princípios de Marketing.** Rio de Janeiro: Prentice-Hall do Brasil LTDA, 1993.

Krol, R. **Evaluating state revenue forecasting under a flexible loss function.** International Journal of Forecasting, 29, 282–289, 2013.

Levy, M.; Weitz, B. A. **Administração de Varejo.** São Paulo: Atlas, 2000.

Lotze, T. H.; Shmueli, G. **How does improved forecasting benefit detection? An application to biosurveillance.** International Journal of Forecasting, 25, 467–483, 2009.

McClain, J. O. Dominant Tracking Signals. **International Journal of Forecasting.** 4, 563-572, 1988.

McKenzie, ED. **The Monitoring of Exponentially Weighed Forecasts.** Journal of operational Research Society, 29, 449-458, 1978.

MOINZADEH, K. **A multi-echelon Inventory System with Information Exchange**. *Management science*, 48, n. 3, 414-426, 2002.

Morettin, P.A.; Toloi, L.M.C. **Análise Séries Temporais**. 2ª Ed. ABE Projeto Fisher, Ed. Edgard Blucher, 2006.

Neill, D. B. **Expectation-based scan statistics for monitoring spatial time series data**. *International Journal of Forecasting*, 25, 498–517, 2009.

Ord, J. K.; Koehler, A. B.; Snyder, R.D.; Hyndman, R. J. **Monitoring processes with changing variances**. *International Journal of Forecasting*, 25, 518–525, 2009.

Özer, Ö. **Replenishment strategies for distribution systems under advance demand information**. *Management Science*, vol. 49, 3, pp. 255–272, 2003.

Parente, J. **Varejo no Brasil: Gestão e Estratégia**. Editora Atlas, São Paulo, 2000.

Rosa, J. P.; Coral, M. L. S.; Flamengo, L. C. B. **Gestão de varejo: estratégia de negócio e gestão de estoques**. FCV Empresarial, v.1, p. 73-98, 2007.

Silver, E. A.; Pyke, D. F.; Peterson, R. **Decision Systems for Inventory Management and Production Planning and Scheduling**. 3th ed. New Jersey: John Wiley, 1998.

Snyder, R.D.; Koehler, A. B. **Incorporating a tracking signal into a state space model**. *International Journal of Forecasting*, 25, 526–530, 2009.

Trigg, D. W. **Monitoring a Forecasting System**. *Operational Research Quarterly*, 15, 271-274, 1964.

Wang Q. **A periodic-review inventory control policy for a two-level supply chain with multiple retailers and stochastic demand**. *European Journal of Operational Research*. vol. 230, pp. 53–62, 2013.

Zipkin, P. H. **Foundations of Inventory management**. New York: McGraw-Hill, 2000.

APÊNDICE A – Método de busca dos limites de detecção correspondentes a um determinado NMA0

Para quaisquer das estatísticas *Tracking Signal* testadas nesta tese, o método de busca adotado para determinar o limite L para um dado NMA0 foi o método da secante, para encontrar o zero de uma função univariada no intervalo $[A B]$. A partir da definição de um valor para L é possível definir o NMA0 por meio de simulação. Seja $Sim(\ell)$ o número médio de períodos até o primeiro alarme falso, encontrado após a simulação para um determinado valor de L , pode-se definir a seguinte expressão

$$g(\ell) = Sim(\ell) - NMA_0, \quad (A.1)$$

sendo NMA_0 a meta estipulada que limita o número esperado de alarmes falsos por ano e $(\alpha \alpha_1)$ são parâmetros fixos e conhecidos. Dessa forma, deseja-se encontrar um determinado ℓ que resulte em $g(\ell) \cong 0$. Nesse caso, o método da secante pode ser aplicado pelo algoritmo:

1 - Com os valores $(\alpha \alpha_1 NMA_0)$ encontrar um intervalo $[\ell_1, \ell_2]$ de forma que $\ell_1 < \ell_2$ e $g(\ell_1)g(\ell_2) < 0$;

2- Definir um valor para precisão ξ e iniciar o erro de forma que $e \gg \xi$;

3- Enquanto $e > \xi$ faça:

 Calcule :

$$g(\ell_1) \text{ e } g(\ell_2);$$

$$\ell_3 = \ell_2 - g(\ell_2) \left[(\ell_2 - \ell_1) / (g(\ell_2) - g(\ell_1)) \right];$$

$$e = |\ell_3 - \ell_2| = (g(\ell_2)\ell_1 - g(\ell_1)\ell_2) / (g(\ell_2) - g(\ell_1));$$

$$\text{faça } \ell_1 = \ell_2 \quad \text{e} \quad \ell_2 = \ell_3;$$

 fim

Ao final, ℓ_3 será o ponto que atenderá $g(\ell_3) \cong 0$.

APÊNDICE B – Limites iniciais de busca para S_0 no algoritmo da Seção 4.6.3.

Como foi visto no algoritmo de otimização da Seção 4.6.3. é definido um intervalo inicial para a variável S_0 . Dessa forma é necessário encontrar os limites superior e inferior para S_0 . A métrica escolhida para esse objetivo foi a fração de demanda não atendida no CD no intervalo $T_0 = mT$.

O valor esperado do total de faltas no CD (demanda não atendida) em um intervalo mT é dado por $E(D_0(L_0 + (m-1)T) - S_0)^+$, sendo que o valor esperado de demanda no CD nesse mesmo intervalo é de $\mu_0 mT$. Portanto, a fração de demanda não atendida prontamente no CD é dada por:

$$f = \frac{E(D_0(L_0 + (m-1)T) - S_0)^+}{\mu_0 mT}.$$

Evidentemente, $0 \leq f \leq 1$, e levando em conta que $E(D_0(a) - S_0)^+ = \sigma_0 \sqrt{a} \Phi^1(z)$ obtém-se:

$$0 \leq \frac{\sqrt{L_0 + (m-1)T} \sigma_0 \Phi^1\left\{\frac{[S_0 - \mu_0(L_0 + (m-1)T)]}{\sqrt{L_0 + (m-1)T} \sigma_0}\right\}}{\mu_0 mT} \leq 1 \quad (\text{B.1})$$

Considerando apenas o limite superior em (B.1) obtém-se:

$$\frac{\sqrt{L_0 + (m-1)T} \sigma_0 \Phi^1\left\{\frac{[S_0 - \mu_0(L_0 + (m-1)T)]}{\sqrt{L_0 + (m-1)T} \sigma_0}\right\}}{\mu_0 mT} \leq 1,$$

ou ainda:

$$\Phi^1\left\{\frac{[S_0 - \mu_0(L_0 + (m-1)T)]}{\sqrt{L_0 + (m-1)T} \sigma_0}\right\} \leq \frac{\mu_0 mT}{\sqrt{L_0 + (m-1)T} \sigma_0}. \quad (\text{B.2})$$

Na Equação (B.2) $\mu_0 mT$ e $\sqrt{L_0 + (m-1)T} \sigma_0$ são a média e o desvio padrão da demanda no CD em um intervalo mT . Essa demanda tem distribuição normal, e segundo Axsäter (2006), para a demanda ser aproximada por uma normal o quociente σ/μ deve ser menor que 1. Silver, et. al (1998) recomenda que $\sigma/\mu < 0,5$ ou seja, $\mu/\sigma > 2$. Logo, o lado direito da Inequação (B.2) é sempre maior ou igual a 2. Para $z \geq 2$, pode-se considerar a aproximação $\Phi^1(-z) = z$. Dessa forma, para que Inequação (B.2) seja satisfeita tem-se que:

$$[S_0 - \mu_0(L_0 + (m-1)T)] / \sqrt{L_0 + (m-1)T} \sigma_0 \geq - \frac{\mu_0 mT}{\sqrt{L_0 + (m-1)T} \sigma_0}$$

Após alguma manipulação algébrica, tem-se que:

$$S_0 \geq \mu_0(L_0 - T) \quad (\text{B.3})$$

Considerando agora o limite inferior na Expressão (A.1) tem-se que,

$$0 \leq \frac{\sqrt{L_0 + (m-1)T} \sigma_0 \Phi^1 \left\{ [S_0 - \mu_0(L_0 + (m-1)T)] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\}}{\mu_0 mT}$$

ou ainda:

$$\Phi^1 \left\{ [S_0 - \mu_0(L_0 + (m-1)T)] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\} \geq 0. \quad (\text{B.4})$$

Para $z > 5$, pode-se considerar a aproximação $\Phi^1(z) = 0$, então:

$$[S_0 - \mu_0(L_0 + (m-1)T)] / \sqrt{L_0 + (m-1)T} \sigma_0 \leq 5$$

ou ainda:

$$S_0 \leq 5\sqrt{L_0 + (m-1)T} \sigma_0 + \mu_0[L_0 + (m-1)T] \quad (\text{B.5})$$

APÊNDICE C – Rotinas e funções programadas em MATLAB

Nesta tese foram utilizadas rotinas e funções programadas em linguagem MATLAB, as quais serão apresentadas a seguir juntamente com uma breve descrição de seu funcionamento.

Conforme descrito na Seção 1.4., a determinação do NMA é parte integrante da obtenção dos limites de detecção das estatísticas *Tracking Signal*. As funções que realizam essa tarefa para cada estatística estudada são: funcTsT (para *TST*); funcTsT1 (para *TST'*); funcTsB (para *TSB*); funcTsB1 (para *TSB'*). Essas funções retornam o NMA mediante um grupo de parâmetros de entrada conforme apresentado a seguir:

parâmetros da geração da demanda:

- a - nível constante da demanda;
- M_e - média do ruído da demanda;
- dp_e - desvio padrão do ruído da demanda;
- N - tamanho da série temporal de demanda;

parâmetros da previsão de demanda:

- n - tamanho da amostra usada para estimativa inicial de a;
- alfa - parâmetro do amortecimento exponencial;

parâmetros da estatística usada na detecção do viés:

- Lt ou Lt1 ou LB ou LB1 - limite de detecção;
- alfa1 - usado no amortecimento do erro da previsão e na atualização do DAM;

parâmetros da mudança de comportamento na demanda:

- t_Delta - período de inserção da mudança de comportamento (>n);
- b - taxa de crescimento para o caso de inclusão de tendência;
- delta_a - incremento no nível do modelo constante;

parâmetro da simulação:

- T - número de replicações;

Código das funções:

```

*Para TST
function NMA =funcTsT(Lt,T,M_e,dp_e,a,delta_a,b,N,n,t_Deltab,alfa,alfa1)
%vetor que guarda NA (número de períodos até o alarme) de cada replicação
R0=zeros(1,1);%esse vetor cresce dinamicamente
%correção na contagem dos períodos no caso de inclusão de tendência
corr=0;
if(b~=0 && t_Deltab>0)
    corr=1;
end
%início do processo de simulação
for s=1:T
%geração
et=random('norm',M_e,dp_e,[N,1]);%geração do erro aleatório
Xt=a+et;%série temporal(modelo constante)
%inclusão da tendência e nível
for i=t_Deltab+n+1:N
Xt(i)=Xt(i)+b*(i-t_Deltab-n)+delta_a;
end
%método de previsão
M_Xt=sum(Xt(1:n,1))/n;%estimação da média
%estimativa do DAM
DAM=0;%estimado o DAM
for i=1:n
    DAM=DAM+abs(M_Xt-Xt(i))/n; %estimação do DAM(desvio absoluto médio)
end
%previsão, estimativa do erro de previsão
prev=zeros(N-n+1,1); %vetor das previsões
prev(1)=M_Xt; % previsão inicial
Tst=zeros(N-n,1);% vetor das estatísticas TST
e=zeros(N-n+1,1); % vetor dos erros de previsão
e(1)=Xt(n)-M_Xt; % erro de previsão inicial
et=zeros(N-n+1,1); % vetor dos erros amortecidos
dam=zeros(N-n+1,1); % vetor dos valores de DAM
dam(1)=DAM; % valor inicial do DAM
%cálculo das previsões, dos erros e das estatísticas para uma replicação
for t=n+1:N
    prev(t-n+1)=alfa*Xt(t)+(1-alfa)*prev(t-n);
    e(t-n+1)=(Xt(t)-prev(t-n));

```

```

%estatísticas
    et(t-n+1)=alfa1*e(t-n+1)+(1-alfa1)*et(t-n);
    dam(t-n+1)=alfa1*abs(e(t-n+1))+(1-alfa1)*dam(t-n);
    Tst(t-n)=et(t-n+1)/dam(t-n+1);
%comparando a estatística com o limite de detecção
    if(abs(Tst(t-n))>Lt)%se o valor absoluto "furar" o limite
        if((t-n)>(t_Deltab+corr))%se o ponto que rompeu o limite for > t_Delta
            R0=[R0;t_(t_Deltab+n+corr)];
        end %se o alarme ocorrer antes de t_Delta o resultado é descartado
        break
    end
end%fim do loop das estatísticas
end
%cálculo do NMA
[linha,coluna]=size(R0);
if(mean(R0(1:linha))==0)
    NMA=T;
else
    NMA=mean(R0(2:linha));
end
end

*Para TST'

function NMA =funcTsT1 (LT1,T,M_e,dp_e,a,b,delta_a,N,n,t_Deltab,alfa,alfa1)
%início do processo de simulação
R01=zeros(1,1);%vetor para guardar TS_1
for s=1:T
    corr=0;
    if(b~=0 && t_Deltab>0)
        corr=1;
    end
%geração
    et=random('norm',M_e,dp_e,[N,1]);%geração do erro aleatório
    Xt=a+et;%série temporal
%inclusão da tendência e nível
    for i=t_Deltab+n+1:N
        Xt(i)=Xt(i)+b*(i-t_Deltab-n);
    end
end

```

```

%método de previsão
M_Xt=sum(Xt(1:n,1))/n;%estimação da média
%estimativa do EQM
EQM=0;% EQM
    for i=1:n
        EQM=EQM+(M_Xt-Xt(i))^2/(n-1);
    end
%previsão, estimativa do erro de previsão
prev=zeros(N-n+1,1); prev(1)=M_Xt;
Tst1=zeros(N-n,1);
e=zeros(N-n+1,1);
et1=zeros(N-n+1,1);
eqm1=zeros(N-n+1,1);eqm1(1)=EQM;
for t=n+1:N %nesse loop todas estatísticas são processadas
    prev(t-n+1)=alfa*Xt(t)+(1-alfa)*prev(t-n);%previsão
    e(t-n+1)=(Xt(t)-prev(t-n));%erro da previsão
%estatísticas
    et1(t-n+1)=alfa1*e(t-n+1)+(1-alfa1)*et1(t-n);%erro amortecido
    eqm1(t-n+1)= alfa1*(e(t-n+1))^2+(1- alfa1)*eqm1(t-n);% EQM
    Tst1(t-n)=et1(t-n+1)/sqrt(eqm1(t-n+1));%estatística do TS_1
%avaliando as estatísticas
    if(abs(Tst1(t-n))>LT1)%se o valor absoluto do sinal "furar" o limite
        if((t-n)>(t_Deltab+corr))%se o ponto que rompeu o limite for > t_Delta
            R01=[R01;t-(t_Deltab+n+corr)];
        end
        break
    end
end%fim do loop da estatística
end%fim do loop das replicações

[linha1,coluna1]=size(R01);
if(mean(R01(1:linha1))==0)
    NMA=T;
else
    NMA=mean(R01(2:linha1));
end
end

```

```

*Para TSB
function NMA =funcTSB (Lb,T,M_e,dp_e,a,delta_a,b,N,n,t_Deltab,alfa,alfa1)
%início do processo de simulação
R03=zeros(1,1);%vetor para guardarTS_b
corr=0;
if(b~=0 && t_Deltab>0)
    corr=1;
end
for s=1:T
%geração
et=random('norm',M_e,dp_e,[N,1]);%geração do erro aleatório
Xt=a+et;%série temporal
%inclusão da tendência e nível
for i=t_Deltab+n+1:N
Xt(i)=Xt(i)+b*(i-t_Deltab-n-1)+delta_a;
end
%método de previsão
M_Xt=sum(Xt(1:n,1))/n;%estimação da média
%estimativa do DAM
DAM=0;%EQM=0;%estimado o DAM e o EQM
for i=1:n
    DAM=DAM+abs(M_Xt-Xt(i))/n; %estimação do DAM(desvio absoluto médio)
end
%previsão, estimativa do erro de previsão
prev=zeros(N-n+1,1); prev(1)=M_Xt;
TsB=zeros(N-n,1);
e=zeros(N-n+1,1);
dam3=zeros(N-n+1,1);dam3(1)=DAM;
cusum=zeros(N-n+1,1);
for t=n+1:N %nesse loop todas estatísticas são processadas
    prev(t-n+1)=alfa*Xt(t)+(1-alfa)*prev(t-n);%previsão
    e(t-n+1)=(Xt(t)-prev(t-n));%erro da previsão
    cusum(t-n+1)=cusum(t-n)+e(t-n+1);%erro cumulativo
end
%estatística Brown

```

```

    dam3(t-n+1)=alfa1*abs(e(t-n+1))+(1-alfa1)*dam3(t-n);%atualização da
estimativa DAM
    TsB(t-n)=cusum(t-n+1)/dam3(t-n+1);%estatística do TS_3
%comparando as estatísticas com os limites
    if(abs(Tst3(t-n))>Lb)%se o valor absoluto do sinal "furar" o limite
        if((t-n)>(t_Deltab+corr))%se o ponto que rompeu o limite for > t_Delta
            R03=[R03;t-(t_Deltab+n+corr)];
        end
        break
    end
end%fim do loop da estatística
end%fim do loop das replicações

[linha3,coluna3]=size(R03);
if(mean(R03(1:linha3))==0)
    NMA=T;
else
    NMA=mean(R03(2:linha3));
end
end

*Para TSB'
function NMA =funcTSB1(Lb1,T,M_e,dp_e,a,delta_a,b,N,n,t_Deltab,alfa,alfa1)
%início do processo de simulação
R02=zeros(1,1);
corr=0;
if(b~=0 && t_Deltab>0)
    corr=1;
end
for s=1:T
%geração
et=random('norm',M_e,dp_e,[N,1]);%geração do erro aleatório
Xt=a+et;%série temporal
%inclusão da tendência e nível
for i=t_Deltab+n+1:N
Xt(i)=Xt(i)+b*(i-t_Deltab-n-1)+delta_a;
end

```

```

%método de previsão
M_Xt=sum(Xt(1:n,1))/n;%estimação da média
%estimativa do DAM
EQM=0;%estimado o EQM
    for i=1:n
        EQM=EQM+(M_Xt-Xt(i))^2/(n-1);
    end
%previsão, estimativa do erro de previsão
prev=zeros(N-n+1,1); prev(1)=M_Xt;
Tsb1=zeros(N-n,1);e=zeros(N-n+1,1);
eqm2=zeros(N-n+1,1);eqm2(1)=EQM;
cusum=zeros(N-n+1,1);
for t=n+1:N %nesse loop as estatísticas são calculadas
    prev(t-n+1)=alfa*Xt(t)+(1-alfa)*prev(t-n);%previsão
    e(t-n+1)=(Xt(t)-prev(t-n));%erro da previsão
    cusum(t-n+1)=cusum(t-n)+e(t-n+1);%erro cumulativo
%estatísticas
    eqm2(t-n+1)=alfa1*(e(t-n+1))^2+(1-alfa1)*eqm2(t-n);%atualização do EQM
    Tsb1(t-n)=cusum(t-n+1)/sqrt(eqm2(t-n+1));%estatística do TSB'
%avaliando as estatísticas
    if(abs(Tsb1(t-n))>Lb1)%se o valor absoluto do sinal "furar" o limite
        if((t-n)>(t_Delta+corr))%se o ponto que rompeu o limite for > t_Delta
            R02=[R02;t-(t_Delta+n+corr)];
        end
    end
    break
end
end%fim do loop da estatística
end%fim do loop das replicações
[linha2,coluna2]=size(R02);
%número médio de períodos até a detecção
if(mean(R02(1:linha2))==0)
    NMA=N;
else
    NMA=mean(R02(2:linha2));
end
end
end

```

Para o cálculo dos limites de detecção o algoritmo nos programas findL* testa iterativamente as funções do tipo funcTs* até encontrar o limite de detecção que produz o NMA0 definido.

Além dos parâmetros das funções funcTs*, nos programas findL* são necessários os parâmetros:

I0 - valor inferior para o limite de detecção;

I1 - valor máximo para o limite de detecção;

NMA0 - meta de NMA0;

*Para TST:

```
function L =findKTsT (NMA0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1,I0,I1)
```

```
x0=I0; x1=I1;
```

```
e=0.0001;erro=1000;
```

```
while(erro>e)
```

```
    fx0=funcTsT (x0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
```

```
    fx1=funcTsT (x1,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
```

```
    x2=(fx1*x0-x1*fx0)/(fx1-fx0);
```

```
    erro=sqrt((x1-x2)^2);
```

```
    x0=x1;
```

```
    x1=x2;
```

```
end
```

```
L=x2;
```

```
end
```

*Para TST':

```
function L =findLTsT1 (NMA0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1,I0,I1)
```

```
x0=I0; x1=I1;
```

```
e=0.0001;erro=1000;
```

```
while(erro>e)
```

```
    fx0=funcTsT1 (x0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
```

```
    fx1=funcTsT1 (x1,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
```

```
    x2=(fx1*x0-x1*fx0)/(fx1-fx0);
```

```
    erro=sqrt((x1-x2)^2);
```

```
    x0=x1;
```

```
    x1=x2;
```

```
end
```

```
L=x2;
```

```
end
```

*Para TSB:

```
function L =findLTsB (NMA0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1,l0,l1)
x0=l0; x1=l1;
e=0.0001;erro=1000;
while(erro>e)
    fx0=funcTsB (x0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
    fx1=funcTsB (x1,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
    x2=(fx1*x0-x1*fx0)/(fx1-fx0);
    erro=sqrt((x1-x2)^2);
    x0=x1;
    x1=x2;
end
L=x2;
end
```

*Para TSB':

```
function L =findLTsB1 (NMA0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1,l0,l1)
x0=l0; x1=l1;
e=0.0001;erro=1000;
while(erro>e)
    fx0=funcTsB1 (x0,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
    fx1=funcTsB1 (x1,T,M_e,dp_e,a,b,N,n,t_Deltab,alfa,alfa1)-NMA0;
    x2=(fx1*x0-x1*fx0)/(fx1-fx0);
    erro=sqrt((x1-x2)^2);
    x0=x1;
    x1=x2;
end
L=x2;
end
```

Para que o desempenho na detecção de viés fosse avaliado foi necessário calcular o NMA1 com o auxílio das funções funcTs*. Para isso, basta adotar para o parâmetro $\delta_a > 0$ e $t_{\text{deltab}} > n$, induzindo o surgimento do viés. Dessa forma o programa retorna o NMA1.

O programa chamado SdoisNiveis monta a função de custo do modelo e busca iterativamente pelos níveis de reposição no CD e nas lojas que minimizam o custo. Os dados de entrada são:

$m_i = []$ - vetor de médias das demandas nas lojas;
 $v_i = []$ - vetor de variâncias das demandas nas lojas;
 $l_i = []$ - vetor de *lead times* nas lojas;
 L_0 - *lead time* no CD;
 T - intervalo de revisão no CD;
 t - %Intervalo de revisão nas lojas;
 β - meta de nível de serviço nas lojas;
 h_0 - custo de manter o estoque no CD;
 h_i - custo de manter o estoque nas lojas;

A rotina apresentada abaixo foi escrita para um sistema com um CD e três lojas no varejo.

```

%pré-processamento
k=T/t;V=sum(vi);M=sum(mi);
pi=zeros(1,N);%fator de desagregação
Auxpi=2*sum(vi);
for i=1:1:N
    pi(i)=1/(2*N)+vi(i)/Auxpi;
end
%encontrando o mínimo pelo método da seção áurea
%cálculo dos limites superior e inferior para os níveis
Nit=0;%número de iterações até o ótimo
A=((L0-t)/T)/(M*k*t);%limite inferior
B=((40.69*sqrt((L0+(k-1)*t)*V)+M*(L0+(k-1)*t))/(M*k*t));%limite superior
criterio=10000;
si=zeros(2,N);CT=zeros(1,2);S0otimo=1000;siOtimo=zeros(2,N);
while (criterio>=(1/(M*k*t)))
    %processando no CD
    S0=[(B-0.618*(B-A))*M*k*t,(A+0.618*(B-A))*M*k*t];
    for m=1:2
        %faltas a cada momento T no CD
        B0=zeros(1,k);
        B0(1)=sqrt(L0*V)*G((S0(m)-M*L0)/sqrt(L0*V));
        for j=1:1:(k-1)
            B0(j+1)=sqrt(L0+j*t)*sqrt(V)*G((S0(m)-M*(L0+j*t))/(sqrt(L0+j*t)*sqrt(V)))-
                sqrt(L0+(j-1)*t)*sqrt(V)*G((S0(m)-M*(L0+(j-1)*t))/(sqrt(L0+(j-1)*t)*sqrt(V)));
        end
    end
  
```

```

%estoque médio no CD
I0I=S0(m)-M*L0+sqrt((L0)*V)*G((S0(m)-M*(L0))/sqrt((L0)*V));
I0F=S0(m)-M*(L0+(k-1)*t)+sqrt((L0+(k-1)*t)*V)*G((S0(m)-M*(L0+(k-
1)*t))/sqrt((L0+(k-1)*t)*V));
I0=(I0I+I0F)/2;
%lead time médio no varejo i
Li=zeros(1,N);
%Li para cada varejo
for i=1:N
aux=0;
for j=0:k-1
aux=aux+(B0(j+1)/(mi(i)*k*t))^(k-j)*t;
end
Li(i)=li(i)+pi(i)*aux;
end
%cálculo de S para cada varejo dada uma meta de nível de serviço
for i=1:1:N
si(m,i)=fzero(@(s)(sqrt(vi(i)*(Li(i)+t))*G((s-
mi(i)*(Li(i)+t))/sqrt(vi(i)*(Li(i)+t)))-sqrt(vi(i)*(Li(i)))*G((s-
mi(i)*(Li(i)))/sqrt(vi(i)*(Li(i))))))/(mi(i)*t)-1+beta,0);
end
%estoque médio no varejo
li=zeros(1,N);
for i=1:1:N
li(i)=(2*si(m,i)-mi(i)*(2*Li(i)+t)+sqrt(vi(i)*(Li(i)+t))*G((si(m,i)-
mi(i)*(Li(i)+t))/sqrt(vi(i)*(Li(i)+t)))+sqrt(vi(i)*(Li(i)))*G((si(m,i)-
mi(i)*(Li(i)))/sqrt(vi(i)*(Li(i))))))/2;
end
%processando custo total e nível de serviço
CT(m)=(sum(li)*hi+I0*h0);
end
if (CT(1)<CT(2))
B=A+0.618*(B-A);
Ctmin=CT(1);
S0otimo=S0(1);
siOtimo=si(1,:);
elseif (CT(2)<=CT(1))

```

```

A=B-0.618*(B-A);
Ctmin=CT(2);
S0otimo=S0(2);
siOtimo=si(2,:);
end
Nit=Nit+1;
criterio=B-A;
end

```

Observa-se que no programa SdoisNiveis é recorrente a função G(). Essa função foi programada com o objetivo de calcular a perda de primeira ordem conforme rotina a seguir:

```

function g =G (x)
g=normpdf(x)-x*(1-normcdf(x));

```

Para observar a eficácia do modelo de estoque proposto, foi elaborada uma rotina chamada simulaEstoque. Esse programa simula a movimentação do estoque nas lojas e no CD, para o sistema descrito na ilustração numérica no Capítulo 4, com o objetivo de calcular o nível de serviço observado em cada loja ao final da simulação. Nesse caso, os níveis de reposição do estoque são definidos pela rotina SdoisNiveis.

```

dados de entrada
mi=[]; vetor de médias das demandas no varejo;
vi=[]; vetor de variâncias das demandas no varejo;
li=[]; vetor de lead times nas lojas;
L0 - lead time no CD;
T0 - intervalo de revisão no CD;
h0 - custo de manutenção do estoque no CD;
hi - custo de manutenção de estoque nas lojas;
S=[] vetor de níveis de reposição no CD e nas lojas;
T - horizonte de simulação.


```

%pre-processamento
[x,N]=size(mi);%total de lojas no varejo
%racionamento
pi=zeros(1,N);
demTotal=zeros(1,N);demNaoSatisfeita=zeros(1,N);estoqueMedio_em_t=zeros(
N,T);
for i=1:1:N

```


```

```

    pi(i)=1/(2*N)+vi(i)/(2*sum(vi));
end
%simulação
%criando matrizes para guardar valores
data=zeros(N*5+2,T+max(li)+L0+2);
%a primeira coluna de data é reservada para o estoque inicial máximo
data(2,1)=S(1);
for i=1:N
    data(5*i+1,1)=S(i+1);%o estoque em mãos inicia com Si
end
for t=2:1:T
    dem=zeros(1,N);%demanda
    for i=1:N
        dem(i)=normrnd(mi(i),sqrt(vi(i)));
        demTotal(i)=demTotal(i)+dem(i);
    end
    %movimentação no varejo
    O=zeros(1,N);%pedidos para o CD
    for i=1:N
        %registrando o estoque à mão para cálculo do estoque médio e fração de
        %demanda não atendida
        if((data(5*i+1,t-1)+data(5*i-2,t))<dem(i))
            data(5*i+2,t)=data(5*i+2,t)+dem(i)-(data(5*i+1,t-1)+data(5*i-2,t));
            demNaoSatisfeita(i)=demNaoSatisfeita(i)+data(5*i+2,t);
        elseif((data(5*i+1,t-1)+data(5*i-2,t))>dem(i))
        %registrando estoque a mão
            data(5*i+1,t)=data(5*i+1,t)+(data(5*i+1,t-1)+data(5*i-2,t))-dem(i);
        end
        %pedidos para o CD
        if((data(5*i+1,t-1)+data(5*i-2,t))>=S(i+1))
            if(data(5*i+1,t)>=S(i+1))
                O(i)=0;
            else
                O(i)=S(i+1)-data(5*i+1,t);%apenas a diferença
            end
        else
            if(data(5*i+1,t)>=S(i+1))%se a demanda for zero

```

```

        O(i)=0;
    else
        O(i)=(data(5*i+1,t-1)+data(5*i-2,t))-data(5*i+1,t);
    end
end
data(5*i,t)=O(i); %registrando o pedido
estoqueMedio_em_t(i,t-1)=(data(5*i+1,t-1)+data(5*i-2,t)+data(5*i+1,t))/2;
end
%movimentação no CD
Ot=sum(O);%pedido total do varejo
%definindo reposição e atraso no varejo
for i=1:N
    if((data(2,t-1)+data(1,t))==0)%nesse caso o pedido está em atraso
        data(5*i-1,t)=O(i);
    elseif((data(2,t-1)+data(1,t))>0 && (data(2,t-1)+data(1,t))<Ot)%racionar as
faltas e reposição
        data(5*i-1,t)=pi(i)*(Ot-(data(2,t-1)+data(1,t)));%faltas alocadas
        if(O(i)-data(5*i-1,t)>0)
            data(5*i-2,t+li(i)+1)=data(5*i-2,t+li(i)+1)+O(i)-data(5*i-1,t);
        end
    elseif((data(2,t-1)+data(1,t))>0 && (data(2,t-1)+data(1,t))>=Ot)%nesse
caso não há atrasos
        data(5*i-2,t+li(i)+1)=data(5*i-2,t+li(i)+1)+O(i);
        data(2,t)=(data(2,t-1)+data(1,t))-Ot;
    end
end
%definindo pedidos e reposição no CD
teste=(t-1)/T0;
if(teste-round(teste)==0)%se for positivo é dia de revisão no CD
    %definindo as reposições do varejo e do CD a partir do CD
    atrasos=zeros(1,N);
    for i=1:N
        for j=0:T0-1
            atrasos(i)=atrasos(i)+ data(5*i-1,t-j);%computando os atrasos
        end
    end
end
data(1,t+L0)=data(1,t+L0)+S(1)-data(2,t);%reposição no CD

```

```
    for i=1:N
        data(5*i-2,t+L0+li(i)+1)=data(5*i-2,t+L0+li(i)+1)+atrasos(i);%reposição no
varejo
    end
end
end
end
%computando o nível de serviço efetivo
servico=zeros(1,N);estoqueMedio=zeros(1,N);
```