U.PORTO

FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

PREVISÃO DE VENDAS NA CADEIA DE ABASTECIMENTO NO SETOR DO RETALHO INTEGRANDO ATIVIDADE PROMOCIONAL

Mariana Cardoso Teixeira

Dissertação
Mestrado em Economia

Orientado por
José Manuel Soares Oliveira
Patrícia Alexandra Gregório Ramos

2022

Resumo: Os retalhistas apoiam-se na previsão de vendas para planear a sua atividade e garantir uma gestão adequada da cadeia de abastecimento. No caso do setor do retalho, a previsão de vendas é necessária em diversos níveis de agregação e, nesse sentido, a abordagem da previsão hierárquica tem sido bastante explorada. Além disso, em resultado da crescente concorrência no setor do retalho, a realização de atividades promocionais pelos retalhistas tem-se tornado cada vez mais frequente. Dado o forte impacto dessas atividades na decisão de compra dos clientes, esta dissertação pretende, assim, integrar informação adicional, tal como atividade promocional, nos métodos de previsão hierárquica, de forma a aumentar a precisão das previsões de vendas no setor do retalho. Para isso, é realizado um estudo com base em dados de vendas disponibilizados pelo Grupo Jerónimo Martins. Para gerar as previsões base para os diferentes níveis da hierarquia, são utilizados os modelos ARIMA e uma DNN. Como estratégias de reconciliação das previsões base, são considerados os métodos *bottom-up* e *top-down* e as estratégias de reconciliação ótima com estimadores OLS e WLS (*struct*). Os resultados permitem concluir que a integração da atividade promocional nos modelos ARIMA permite aumentar a precisão das previsões, ainda que o mesmo não aconteça para o caso da DNN, realçando a necessidade de se utilizarem métodos de *Deep Learning* mais complexos que apresentam, no entanto, a desvantagem de terem custos de processamento e tempos de execução mais elevados.

Códigos JEL: C53, L81, M37

Palavras-chave: Previsão hierárquica, Gestão da cadeia de abastecimento, Atividade promocional, Modelos de séries temporais, *Deep Learning*

Abstract: Retailers need sales forecasts to plan their activity and ensure a proper supply chain management. In the retail sector, sales forecasts are needed at several aggregation levels. Therefore, hierarchical forecasting methods have been widely used. Because of the increasing competition in the retail sector, promotions have become a common practice. Given the strong impact of promotions on consumers' purchasing decisions, this dissertation intends to integrate additional information, such as promotional data, into hierarchical forecasting, in order to increase the sales forecasts accuracy. A study is performed based on a sales dataset from Pingo Doce stores of the Jerónimo Martins Group. Base forecasts for different hierarchical levels are generated using ARIMA models and a Dense Neural Network, while bottom-up and top-down methods and optimal reconciliation strategies with OLS and WLS (struct) estimators are considered as reconciliation methods. The results show that the integration of promotional data into ARIMA models increase forecasts accuracy, although the same does not apply to the DNN case, which emphasizes the need for more complex Deep Learning methods that present, however, higher processing costs.

JEL codes: C53, L81, M37

Keywords: Hierarchical forecasting, Supply chain management, Promotions, Time series models, Deep Learning

Índice

Resumo ii
Abstract iii
Índice de figuras v
Índice de tabelas vi
Lista de abreviaturas vii
1. Introdução 1
2. Revisão de literatura 4
2.1 Previsão de vendas no setor do retalho 4
2.2 Previsão hierárquica 8
2.3 Fatores determinantes na venda de produtos de retalho 10
3. Metodologia 14
3.1 Previsão hierárquica 14
3.1.1 Séries temporais hierárquicas 14
3.1.2 Estratégias de reconciliação de previsões 15
3.2 Modelos de previsão de séries temporais 20
3.2.1 Modelos ARIMA 21
3.2.2 *Dense Neural Networks* 25
4. Caso de estudo 29
4.1 Conjunto de dados 29
4.2 Procedimento metodológico 31
4.3 Resultados 33
5. Conclusões 40
6. Referências 42

Índice de figuras

Figura 3.1 – Representação de uma estrutura hierárquica com três níveis.....14

Figura 3.2 – Representação de uma NN simples.....25

Figura 4.1 – Estrutura hierárquica de cada SKU.....29

Figura 4.2 – Vendas por nível hierárquico de um SKU representativo.....30

Figura 4.3 – Abordagem *rolling*.....32

Figura 4.4 – Valor da função objetivo (a azul) e do MAE (a vermelho) nos conjuntos de treino e validação para a DNN (à esquerda) e a DNN com regressores (à direita) de um SKU típico de uma das lojas.....33

Figura 4.5 – Série das vendas (a vermelho) e das vendas previstas (a azul) pelos modelos ARIMA, ARIMAX, DNN e DNN com regressores para um SKU típico de umas das lojas.....34

Índice de tabelas

Tabela 4.1 – Número de séries por nível de agregação.....30

Tabela 4.2 – MASE e RMSSE das previsões geradas pelos modelos ARIMA 36

Tabela 4.3 – MASE e RMSSE das previsões geradas pelos modelos ARIMAX 37

Tabela 4.4 – MASE e RMSSE das previsões geradas pela DNN 38

Tabela 4.5 – MASE e RMSSE das previsões geradas pela DNN com regressores 39

Lista de abreviaturas

ACF – *AutoCorrelation Function*

ADL – *Autoregressive Distributed Lag*

AIC – *Akaike's Information Criterion*

ANN – *Artificial Neural Network*

ARIMA – *Autoregressive Integrated Moving Average*

ARIMAX – *Autoregressive Integrated Moving Average with eXogenous variables*

BIC – *Bayesian Information Criterion*

CNN – *Convolutional Neural Network*

DNN – *Dense Neural Network*

FSS – *Forecasting Support System*

GLS – *Generalized Least Squares*

KPSS – *Kwiatkowski-Phillips-Schmidt-Shin*

LASSO – *Least Absolute Shrinkage and Selection Operator*

LSTM – *Long Short-Term Memory neural network*

MAE – *Mean Absolute Error*

MASE – *Mean Absolute Scaled Error*

MinT – *Minimum Trace*

MLPNN – *Multi-Layer Perceptron Neural Networks*

NN – *Neural Network*

OLS – *Ordinary Least Squares*

PACF – *Partial Autocorrelation Function*

PCA – *Principal Component Analysis*

ReLU – *Rectified Linear Unit*

RMSSE – *Root Mean Squared Scaled Error*

RNN – *Recurrent Neural Network*

SES – S*imple Exponential Smoothing*

SGD – *Stochastic Gradient Descent*

SKU – *Stock Keeping Unit*

WLS – *Weighted Least Squares*

1. Introdução

A crescente complexidade das cadeias de abastecimento reforça a necessidade de os retalhistas se apoiarem na previsão de vendas para planear a sua atividade e tomar decisões. Quanto maior a precisão dessa previsão, maior a eficiência da gestão da cadeia de abastecimento (Villegas & Pedregal, 2018). Neste sentido, a escolha do método de previsão de vendas adotado pelos retalhistas deve ter em conta fatores como a sua facilidade de utilização e os seus requisitos computacionais, mas fundamentalmente deve ter como objetivo principal obter previsões de elevada precisão que permitam reduzir as perdas, para evitar a gestão ineficiente de *stocks* e a insatisfação de clientes. Efetivamente, previsões de elevada precisão permitem aos retalhistas reduzir custos derivados do excesso de *stocks*, mitigando a deterioração e o desperdício de produtos, assim como evitar ruturas de *stocks* e o potencial risco de os clientes recorrerem a empresas concorrentes (Fildes et al., 2019).

Os retalhistas precisam de obter informação sobre as vendas futuras nos vários níveis da sua organização interna (SKU – *Stock Keeping Unit*, loja, centro de distribuição, etc.), de modo a poderem tomar decisões adequadas em todos esses níveis (Syntetos et al., 2016). Este conceito de previsão de vendas nos diversos níveis de agregação consiste na previsão hierárquica. Podem ser geradas previsões individualmente para cada nível da hierarquia, porém esta solução não reflete a estrutura hierárquica do retalhista, não garantindo que a soma das previsões do nível desagregado seja igual à previsão do nível agregado correspondente. Contudo, esta coerência entre as previsões dos vários níveis é necessária para que os retalhistas possam tomar decisões apropriadas ao longo de toda a cadeia de abastecimento. Em alternativa, podem ser utilizadas estratégias de reconciliação que asseguram a coerência entre as previsões dos diferentes níveis da hierarquia e permitem, além disso, melhorar a precisão dessas previsões se a estratégia de reconciliação escolhida for apropriada, tendo em consideração aspetos como as particularidades da série temporal, a estrutura de agregação hierárquica e o horizonte temporal de previsão.

Em resultado da crescente concorrência no setor do retalho, a adoção de atividades promocionais tem-se tornado cada vez mais frequente, o que implica a utilização de modelos de previsão que tenham em consideração o efeito que essas atividades têm nas vendas. De facto, a integração de informação promocional na previsão de vendas torna-se relevante, no sentido de aumentar a precisão da previsão (Abolghasemi, Hurley, et al., 2020). No entanto, esta questão ainda não foi estudada para o caso da previsão hierárquica

de vendas e, portanto, existe uma grande motivação para explorar a forma como a precisão dessa previsão pode ser afetada, com a integração das atividades promocionais.

Assim, a presente dissertação tem como objetivo a integração de informação adicional em métodos de previsão hierárquica, de forma a aumentar a precisão das previsões de vendas no setor do retalho. Primeiramente, é necessário definir, atendendo ao melhor interesse do retalhista, a estrutura hierárquica mais adequada a adotar para a agregação dos dados. De seguida, para gerar as previsões base para os diferentes níveis da hierarquia, são considerados os modelos tradicionais *Autoregressive Integrated Moving Average* (ARIMA) e uma *Dense Neural Network* (DNN). Além disso, são incorporadas no processo de previsão informações relativas a preços, dias da semana e eventos de calendário, com o objetivo de aumentar a precisão das previsões. A utilização da DNN para a previsão de vendas é motivada por resultados de estudos recentes que indicam que métodos de *Deep Learning* conseguem superar os modelos de previsão tradicionais. O papel destes métodos na previsão de vendas não foi ainda amplamente explorado pela literatura, pelo que há espaço para progressão e para entender efetivamente quais as vantagens que apresentam face aos restantes. O estudo é suportado pela utilização de dados reais de vendas de produtos de lojas Pingo Doce, disponibilizados pelo Grupo Jerónimo Martins.

Desta forma, esta dissertação permite avançar sobre a determinação do impacto que a integração de regressores no processo de previsão hierárquica tem na precisão das previsões. Além disso, investiga-se se de facto uma DNN permite obter erros de previsão inferiores, em comparação com o modelo ARIMA. Garantir uma elevada precisão em previsões de vendas é sobretudo relevante para grandes retalhistas, visto que, por apresentarem um volume de vendas considerável, pequenos erros de previsão podem ter um impacto muito significativo nas suas receitas. Assim, este estudo ganha relevância para grandes retalhistas, como o Grupo Jerónimo Martins, aumentarem a eficiência das suas decisões, o que se reflete numa gestão mais adequada da oferta e níveis superiores de satisfação dos clientes.

Esta dissertação divide-se em seis secções. Após a presente secção relativa à introdução, segue-se a Secção 2 onde se apresenta a revisão de literatura que descreve os principais conceitos-chave e expõe o enquadramento teórico da questão de investigação com recurso aos diversos contributos da literatura. Na Secção 3 apresenta-se a metodologia a ser utilizada, de forma a solucionar a questão de investigação. A Secção 4 diz respeito ao caso

de estudo aplicado ao Grupo Jerónimo Martins e respetivos resultados obtidos. Finalmente, a Secção 5 sintetiza as principais conclusões, possíveis limitações do estudo e algumas questões que poderão ser objeto de investigação futura, enquanto a Secção 6 apresenta as referências utilizadas para a elaboração da dissertação.

2. Revisão de literatura

2.1 Previsão de vendas no setor do retalho

A previsão de vendas consiste na estimação de valores futuros das vendas, a partir de dados históricos das mesmas e de variáveis que as possam influenciar. Dada a incerteza associada à procura, a previsão de vendas é fundamental para que os retalhistas consigam planear de forma eficiente a distribuição e a reposição dos seus produtos em loja. A obtenção de previsões de elevada precisão permite assim, nomeadamente, evitar diminuições nos lucros dos retalhistas causadas pela gestão ineficiente de *stocks*.

O processo de previsão de vendas geralmente é baseado na utilização de um *Forecasting Support System* (FSS), ou seja, de um sistema que gera previsões que são subsequentemente conjugadas com ajustamentos realizados pelos responsáveis da empresa pela gestão da procura, de modo a se considerarem fatores que não foram incluídos no cálculo das previsões iniciais (Fildes et al., 2009). Contudo, é importante confirmar se esses ajustamentos enviesam as previsões, em vez de contribuírem para o aumento da precisão das mesmas (Davydenko & Fildes, 2013; Franses & Legerstee, 2010). Além disso, a grande escala do problema de previsão das vendas ao nível do produto, no setor do retalho, evidencia a necessidade de maior eficiência e rapidez dos sistemas de previsão computacionais (Seaman, 2018), assim como a perda de viabilidade dos ajustamentos manuais das previsões geradas por esses sistemas. Nesse sentido, cada vez mais os processos de previsão devem integrar informação adicional como as promoções, de forma a diminuir os erros de previsão e, consequentemente, a necessidade de ajustamento das previsões (Abolghasemi, Hurley, et al., 2020; Trapero et al., 2013).

Embora habitualmente cada um dos intervenientes da cadeia de abastecimento utilize o seu FSS de forma independente, existem esquemas colaborativos que permitem a partilha de informação sobre as vendas entre os fornecedores e os retalhistas. As parcerias entre membros da cadeia de abastecimento são essencialmente criadas com o objetivo de reduzir custos e controlar *stocks* (Lee et al., 2000; Yu et al., 2001). Alguns autores concluem que a partilha de informação melhora a precisão das previsões (Hosoda et al., 2008; Trapero et al., 2012), o que pode refletir uma diminuição do efeito *bullwhip*, que é definido por Lee et al. (1997) como uma distorção das informações sobre a procura ao longo da cadeia de abastecimento, quando a variação das vendas do retalhista é menor do que a variação dos pedidos do retalhista aos fornecedores. Em última análise, o efeito *bullwhip* pode resultar

em excesso de *stock* de matérias-primas, custos adicionais de produção, armazenamento e distribuição, derivados do excesso de capacidade e perda de qualidade do serviço prestado aos clientes (Lee & Whang, 2000).

Para realizar a previsão de vendas, os retalhistas geralmente baseiam-se em dados de vendas observadas. No entanto, quando não é garantida oferta suficiente de certos produtos e ocorrem ruturas de *stock*, a procura real não é totalmente observada a partir das vendas, o que pode enviesar negativamente a previsão da procura futura desses produtos (Jain et al., 2015). Além disso, perante a falta de disponibilidade de algum produto, os clientes podem recorrer a produtos substitutos, o que pode resultar num enviesamento positivo da previsão da procura desses produtos (Vulcano et al., 2012). De facto, Kim et al. (2019) demonstram que a ocorrência de rutura de *stock* de um produto dificulta o processo de estimação da procura dos produtos substitutos.

De modo a assegurar níveis adequados de oferta, os retalhistas recorrem frequentemente a *stocks* de segurança (Boone, Boylan, et al., 2019). A dimensão destes está dependente do nível de incerteza da procura e dos respetivos erros de previsão (Beutel & Minner, 2012). Além disso, encontrar um equilíbrio entre o custo de manutenção do *stock* de segurança e as perdas de receita originadas por ruturas de *stock* constitui um dos desafios dos retalhistas.

Surge ainda a questão da previsão da procura de novos produtos, para os quais não foram observadas vendas, e do potencial efeito de canibalização, ou seja, do aumento das vendas desses produtos em detrimento de produtos substitutos (Fisher & Vaidyanathan, 2014). A estratégia de previsão mais comum aplicada a novos produtos baseia-se em dados históricos de vendas de produtos com características semelhantes. Porém, Kahn (2014) realça a importância de se distinguir a previsão de produtos já existentes e a previsão de novos produtos, devendo a última basear-se em análises qualitativas que envolvam julgamentos e suposições.

Para além dos dados históricos das vendas, cada vez mais os retalhistas acedem a informação sobre os seus clientes e sobre os produtos e respetivos preços dos seus concorrentes, de forma a tomar decisões mais eficientes e inovar os seus modelos de negócio (Fisher & Raman, 2018). Efetivamente, um maior conhecimento sobre os comportamentos de compra e preferências dos clientes pode permitir aos retalhistas melhorar a previsão da procura (Boone, Ganeshan, et al., 2019). Estudos recentes demonstram que novas fontes de informação como as avaliações de produtos, as pesquisas

online e as redes sociais conseguem melhorar o desempenho de previsão das vendas (Boone et al., 2018; Chern et al., 2015; Cui et al., 2018; Lau et al., 2018).

Relativamente às abordagens de previsão, verifica-se uma ampla utilização dos modelos ARIMA desenvolvidos por Box e Jenkins (1970). Estes são modelos univariados, pois apenas utilizam a série temporal para efeitos de previsão. Além de gerarem previsões, estes modelos acrescentam a possibilidade de gerar intervalos de previsão que, no caso do setor do retalho, são especialmente importantes para a definição de níveis adequados de *stock* de segurança. Os métodos de *Deep Learning* constituem uma abordagem de previsão promissora, dada a sua capacidade de analisar de forma rápida grandes volumes de dados e de fornecer representações complexas dos mesmos (Najafabadi et al., 2015).

Alon et al. (2001) utilizam as *Artificial Neural Networks* (ANNs) para a previsão de vendas no setor do retalho e mostram que estas são capazes de captar os padrões de tendência e sazonalidade não lineares dos dados e de gerar melhores resultados do que métodos estatísticos tradicionais, como o ARIMA, quando as condições económicas são voláteis. Por outro lado, alguns autores demonstram que, se não for realizado um pré-processamento adequado dos padrões de tendência e sazonalidade dos dados de vendas, o modelo ARIMA permite obter melhores resultados do que as *Neural Networks* (NNs) (Chu & Zhang, 2003; Zhang & Qi, 2005). Estes estudos confirmam os resultados de Nelson et al. (1999) que concluem que, quando é efetuado um ajustamento prévio da sazonalidade dos dados, as NNs geram previsões com maior precisão. Contrariamente ao expectável, o estudo de Aras et al. (2017) não encontra diferenças significativas de desempenho na previsão de vendas entre as ANNs e os modelos ARIMA. Aburto e Weber (2007) desenvolvem um modelo de previsão da procura híbrido que combina os modelos ARIMA com as NNs e os resultados sugerem melhorias em termos da precisão das previsões, em comparação com os modelos individuais.

Dados os resultados mistos sobre o desempenho das ANNs tradicionais, recentemente métodos de previsão mais avançados têm sido objeto de estudo para a previsão de vendas. Vallés-Pérez et al. (2022) demonstram que as *Recurrent Neural Networks* (RNNs) podem ser utlizadas para prever de forma eficaz as vendas ao nível da loja ou do produto no setor do retalho. As *Long Short-Term Memory neural networks* (LSTMs), introduzidas por Hochreiter e Schmidhuber (1997), constituem uma classe das RNNs capaz de aprender as dependências de longo prazo das séries temporais. Nesse sentido, há um maior interesse neste método

para a previsão de vendas, em comparação com as RNNs tradicionais que têm dificuldade em aprender conexões entre informações separadas por um longo período de tempo. Abbasimehr et al. (2020) e Ensafi et al. (2022) demonstram a superioridade das LSTMs com múltiplas camadas sobre métodos de *Machine Learning* e modelos de previsão tradicionais. Falatouri et al. (2022) comparam os desempenhos do modelo ARIMA sazonal e das LSTMs para a previsão de vendas no setor do retalho e concluem que o primeiro permite obter melhores resultados para produtos com comportamento sazonal, enquanto que as LSTMs são mais adequadas para a previsão de produtos com procura estável. Encontramos ainda exemplos na literatura de modelos híbridos que combinam as LSTMs com métodos de *Machine Learning* (Punia et al., 2020; Weng et al., 2020). Wang et al. (2019) mostram que, embora as RNNs e as LSTMs permitam alcançar melhores resultados em termos de capacidade de generalização e precisão das previsões, o modelo ARIMA apresenta um tempo de execução mais baixo, assim como um custo de processamento menor.

As *Convolutional Neural Networks* (CNNs) consistem numa abordagem de *Deep Learning* proposta por LeCun et al. (1990) para a classificação de imagens. Contudo, devido à capacidade das CNNs de reconhecimento de padrões, a sua aplicação estende-se a mais áreas como a previsão de séries temporais. Ma e Fildes (2021) propõem uma estratégia de previsão de vendas baseada em CNNs capaz de melhorar a precisão das previsões das vendas ao nível do produto, em relação a um vasto conjunto de métodos de previsão alternativos. Alguns autores desenvolvem abordagens híbridas que combinam as CNNs e as LSTMs, no sentido de aproveitar as vantagens de ambos os métodos (Kaunchi et al., 2021; Liu et al., 2021; Nithin et al., 2022).

No contexto de retalho *online*, Bandara et al. (2019) desenvolvem uma estratégia de previsão de vendas baseada numa LSTM que integra informação sobre as correlações existentes entre os padrões de procura dos produtos. A aplicação da estratégia a um conjunto de dados da Walmart permite concluir que esta supera os modelos univariados tradicionais. Pan e Zhou (2020) utilizam as CNNs para gerar previsões de vendas *online* com base em dados fornecidos pelo grupo Alibaba relativos ao histórico de vendas e a outras variáveis como os números de pesquisas e de visualizações dos produtos. Os autores mostram que o método proposto permite obter melhores resultados do que o modelo ARIMA. Chen (2020) e He et al. (2022) aplicam as LSTMs ao mesmo conjunto de dados de vendas de um

retalhista *online* e observam que estas permitem reduzir o erro de previsão face a métodos de previsão alternativos.

2.2 Previsão hierárquica

A gestão da cadeia de abastecimento consiste na coordenação das ações dos seus diversos intervenientes, desde os produtores aos consumidores finais, com o principal objetivo de satisfazer a procura dos consumidores (Syntetos et al., 2016). No caso do setor do retalho, a previsão de vendas é necessária em diferentes níveis de agregação para que possam ser tomadas decisões ao longo da cadeia de abastecimento. Neste sentido, a abordagem da previsão hierárquica tem sido bastante explorada, por assumir que as séries temporais podem ser desagregadas mediante várias dimensões relevantes para os retalhistas, tais como os intervalos de tempo (hierarquia temporal) e as categorias dos produtos (hierarquia *cross-sectional*).

No caso da agregação temporal, nos níveis mais agregados predominam componentes como a tendência e a ciclicidade dos dados, enquanto nos níveis inferiores, as séries temporais são caracterizadas por padrões de sazonalidade. Kourentzes et al. (2014) propõem um método de previsão, denominado *Multiple Aggregation Prediction Algorithm*, que se baseia no uso da agregação temporal para modelar os vários componentes das séries temporais. Athanasopoulos et al. (2017) mostram que a previsão com hierarquias temporais permite obter melhores resultados do que a previsão tradicional.

No caso das hierarquias *cross-sectional*, para obter as previsões das vendas, os retalhistas devem especificar o nível de agregação considerando três eixos principais: produto (SKU, categoria, área), localização (loja, centro de distribuição, região, país) e tempo (dia, semana, mês, ano) (Fildes et al., 2019). A escolha do nível de agregação deve ainda ter em consideração que o mesmo tem influência na precisão das previsões (Zotteri et al., 2005). O objetivo é gerar um conjunto de previsões coerentes que reflitam a estrutura hierárquica adotada, isto é, assegurar que a soma das previsões do nível desagregado seja igual à previsão do nível agregado correspondente. Para isso, são utilizadas estratégias que reconciliam as previsões base geradas diretamente para os níveis da hierarquia.

As três estratégias básicas de reconciliação frequentemente utilizadas são: o método *bottom-up*, que consiste em gerar previsões para o nível mais desagregado e agregá-las sucessivamente para obter previsões para os níveis superiores; o método *top-down*, que consiste em gerar previsões para o nível mais agregado e desagregá-las com base em

proporções apropriadas; e o método *middle-out* que corresponde a uma abordagem intermédia entre os dois métodos anteriores (Athanasopoulos et al., 2020). Estes métodos apenas utilizam parte da informação disponível, visto que geram previsões para um único nível de agregação. O método *bottom-up* não tem em conta as correlações entre as séries temporais, apresentando piores resultados nos níveis superiores de agregação. Por sua vez, o método *top-down* perde informação devido à agregação dos dados, gerando previsões com menor precisão nos níveis inferiores da hierarquia. Por forma a resolver esta questão, Athanasopoulos et al. (2009) apresentam um método *top-down* que utiliza proporções com base em previsões, em vez de proporções históricas (Gross & Sohl, 1990), para desagregar as previsões geradas para o nível mais agregado.

Não obstante, Hyndman et al. (2011) constatam que nenhuma das três estratégias básicas de reconciliação tem em consideração as possíveis correlações existentes entre os níveis da hierarquia. Neste sentido, os autores propõem uma abordagem distinta que consiste em gerar previsões de forma independente para cada nível hierárquico e, posteriormente, usar um método para reconciliar essas previsões, para que reflitam a estrutura de agregação. Nesse estudo, o método de reconciliação usado é o método de mínimos quadrados ordinários (OLS – *Ordinary Least Squares*) e os resultados demonstram que esta nova abordagem supera os métodos de reconciliação tradicionais. Hyndman et al. (2016) propõem o uso do método de mínimos quadrados ponderados (WLS – *Weighted Least Squares*) para reconciliar as previsões base e sugerem o uso de algoritmos, no caso de um elevado número de séries temporais, para que a reconciliação seja viável. Wickramasuriya et al. (2019) demonstram que a solução de reconciliação proposta por Hyndman et al. (2011) não é exequível na prática e apresentam a abordagem de reconciliação *Minimum Trace* (MinT). Spiliotis et al. (2021) desenvolvem uma estratégia de reconciliação baseada em métodos de *Machine Learning*, que admite combinações não lineares das previsões base. Esta nova abordagem apresenta, em média, melhores resultados, em termos de precisão e de enviesamento das previsões, comparativamente às estratégias de reconciliação básicas e às lineares desenvolvidas por Hyndman et al. (2011), Hyndman et al. (2016) e Wickramasuriya et al. (2019).

Pennings e van Dalen (2017) desenvolvem uma estratégia de previsão hierárquica integrada, baseada na estimação de um modelo de espaço de estados multivariado com um filtro de Kalman, de modo a considerar na previsão de vendas as relações de complementaridade e

substituição entre os produtos. Deste modo, esta estratégia incorpora toda a informação disponível e gera, tendo em consideração a estrutura hierárquica, previsões para todos os níveis de agregação, permitindo, assim, superar os métodos *bottom-up* e *top-down*. Villegas e Pedregal (2018) apresentam uma abordagem de previsão hierárquica também baseada em modelos de espaço de estados, evidenciando o facto de a mesma garantir a consistência das previsões ao longo do tempo.

Apesar dos diversos estudos na área da previsão hierárquica, ainda não é claro se as previsões devem ser obtidas diretamente para cada nível da hierarquia ou se devem ser geradas através de métodos de previsão hierárquica (Oliveira & Ramos, 2019). Esta questão também se coloca quando a previsão integra regressores como a atividade promocional, havendo a necessidade de explorar sobre qual das abordagens a integração de regressores conduz a uma maior precisão da previsão. Além disso, acresce ainda a dúvida sobre quais os modelos de previsão que devem ser utilizados para cada nível da hierarquia, assim como quais as estratégias de reconciliação que conduzem a melhores resultados.

2.3 Fatores determinantes na venda de produtos de retalho

As decisões de compra dos clientes são influenciadas por vários fatores como o preço, promoções, eventos de calendário (feriados e épocas festivas), condições meteorológicas e sazonalidade. Neste sentido, por forma a gerir de forma eficiente os *stocks* e a promover níveis elevados de satisfação dos clientes, os processos de previsão de vendas dos retalhistas devem considerar, para além dos dados históricos das vendas, fatores que sejam determinantes na venda dos seus produtos.

Devido à incerteza associada às previsões das condições meteorológicas, esta é uma variável ainda pouco usada na previsão de vendas, apesar dos vários estudos apresentados. Divakar et al. (2005) desenvolvem um modelo de previsão de vendas de bebidas que integra o efeito de variáveis como preço, promoções, temperatura média, eventos de calendário e lançamentos de novos produtos. A integração das previsões da temperatura na previsão das vendas é justificada pelo aumento que se verifica na procura de bebidas em consequência do aumento da temperatura. Ramanathan e Muyldermans (2010) analisam um conjunto de fatores determinantes na procura de refrigerantes, nomeadamente, promoções, eventos de calendário e condições meteorológicas. No entanto, os autores observam que o único fator que influencia de forma significativa a procura de todos os produtos são as promoções. Num contexto de retalho *online*, Steinker et al. (2017) mostram

que a integração de informações meteorológicas no processo de previsão reduz de forma significativa o erro de previsão das vendas, sobretudo dos fins de semana e dos dias em que se registam melhores condições meteorológicas. Liu e Ichise (2017) apresentam um método de *Deep Learning*, que usa informações meteorológicas para prever de forma eficaz as vendas de bebidas de uma cadeia de supermercados, que supera as previsões geradas por métodos tradicionais de *Machine Learning*. Hirche et al. (2021) utilizam modelos *Autoregressive Integrated Moving Average with eXogenous variables* (ARIMAX) não sazonais que consideram dados relativos à temperatura e a eventos de calendário para a previsão de vendas de bebidas alcoólicas e concluem que a sensibilidade das vendas à temperatura varia entre regiões e categorias de bebidas, enquanto que as épocas festivas influenciam as vendas de todas as categorias de bebidas consideradas. Verstraete et al. (2019) propõem uma metodologia que lida com a incerteza das previsões das condições meteorológicas de curto e de longo prazo para prever as vendas de produtos do setor do retalho. Ainda assim, Badorf e Hoberg (2020) mostram que a melhoria verificada nas previsões de vendas em resultado da integração das previsões das condições meteorológicas diminui à medida que o horizonte de previsão aumenta.

A partir de dados de vendas de uma empresa de refrigerantes, Ramanathan e Muyldermans (2011) mostram que a informação promocional e os fatores sazonais representam fatores com forte influência nas vendas, enquanto que os eventos de calendário têm um impacto limitado. De facto, a generalidade dos modelos que integram fatores determinantes nas vendas destaca a importância da informação promocional, dada a elevada frequência da realização de atividades promocionais pelos retalhistas e o forte impacto dessas atividades no comportamento dos clientes. Ali et al. (2009) demonstram que os métodos de previsão mais simples exibem um bom desempenho na ausência de promoções. No entanto, para períodos com atividade promocional, a utilização de métodos mais sofisticados, que integrem informação adicional sobre as promoções, permite melhorar de forma significativa a precisão das previsões. Arunraj e Ahrens (2015) apresentam dois modelos de regressão linear que utilizam as previsões geradas por um modelo ARIMA sazonal e variáveis como promoções, eventos de calendário, sazonalidade mensal e condições meteorológicas como regressores. Aplicando os modelos na previsão de vendas, conclui-se que estes superam o método naïve sazonal, o modelo ARIMA sazonal sem integração de regressores e as *Multi-Layer Perceptron Neural Networks* (MLPNN). Arunraj et al. (2016) desenvolvem um modelo ARIMAX sazonal que considera promoções e eventos de

calendário para a previsão de vendas diárias de uma loja de retalho e os resultados indicam melhorias na precisão das previsões, em comparação com o modelo ARIMA sazonal tradicional. Abolghasemi, Beh, et al. (2020) estudam a previsão de séries de procura caracterizada por volatilidade em consequência da prática de promoções. Os autores propõem uma estratégia que combina o modelo ARIMA para a previsão de séries sem promoções e regressão segmentada para prever os aumentos na procura provocados pelas promoções. Esta estratégia híbrida garante menor erro de previsão quando a volatilidade da procura é elevada, enquanto que o modelo ARIMAX exibe melhor desempenho, quando a volatilidade da procura é baixa ou moderada. Abolghasemi, Hurley, et al. (2020) desenvolvem um modelo de previsão da procura que integra o efeito de eventos sistemáticos, como as promoções, com o objetivo de aumentar a precisão das previsões e minimizar o esforço de ajustamento manual das mesmas. Huber e Stuckenschmidt (2020) recorrem a algoritmos de *Machine Learning* para a previsão de vendas com base em informações externas relativas a eventos de calendário, e os resultados permitem confirmar a viabilidade dos métodos considerados e a perda de relevância de ajustamentos manuais após a geração das previsões.

De um modo geral, os modelos de previsão multivariados conseguem melhorar as previsões geradas pelos modelos univariados. No entanto, a integração de fatores determinantes das vendas no processo de previsão requer o uso de métodos de previsão e de estratégias de seleção de variáveis mais sofisticados, de forma a lidar com o número elevado de variáveis explicativas e a heterogeneidade dos dados.

Guo et al. (2013) desenvolvem um modelo de previsão multivariado baseado em NNs que incorpora a abordagem *Harmony Search-Wrapper-based variable selection* para identificar o conjunto de variáveis de entrada mais relevantes para a previsão das vendas. Os autores comprovam a eficácia da estratégia de seleção de variáveis adotada para reduzir o número de parâmetros do método, permitindo alcançar uma maior precisão das previsões. Huang et al. (2014) propõem um modelo de previsão de vendas ao nível do SKU que envolve primeiramente a seleção de variáveis explicativas competitivas relevantes, como preço e promoções, e a subsequente integração dessa informação num modelo *Autoregressive Distributed Lag* (ADL). Os resultados superam os obtidos através do método S*imple Exponential Smoothing* (SES), mesmo quando as previsões geradas por este método são posteriormente ajustadas por efeitos passados das promoções nas vendas. Ma et al. (2016)

utilizam um modelo idêntico ao de Huang et al. (2014), contudo consideram interações promocionais entre produtos de diferentes categorias, além das observadas entre produtos da mesma categoria. Além disso, para selecionar as variáveis explicativas relevantes recorrem a uma estratégia mais eficiente baseada em regressão *Least Absolute Shrinkage and Selection Operator* (LASSO). Trapero et al. (2015) desenvolvem um modelo de regressão múltipla que integra variáveis promocionais e resolve algumas limitações associadas a este tipo de modelos, nomeadamente, o elevado número de variáveis e a multicolinearidade, através da utilização de *Principal Component Analysis* (PCA). Os autores observam que o modelo proposto permite gerar previsões mais precisas do que o método naïve, o método SES e o método multivariado *Last like promotion* desenvolvido por Ali et al. (2009).

3. Metodologia

3.1 Previsão hierárquica

3.1.1 Séries temporais hierárquicas

A título de exemplo e de forma a ilustrar uma estrutura hierárquica simples, considere-se a Figura 3.1. No nível superior da hierarquia (nível 0), está representado o Total que corresponde à série temporal mais agregada. Esta série divide-se nas séries A e B que correspondem ao nível 1 da hierarquia e são, por sua vez, desagregadas em três séries no nível hierárquico inferior (nível 2). Suponhamos que $y_{i,t}$ representa a observação no instante t da série que corresponde ao nó i da estrutura hierárquica, onde $t = 1, \dots, T$. Por exemplo, $y_{AA,t}$ representa o valor da série correspondente ao nó AA, observado no instante t. Sendo n o número total de séries da hierarquia e m o número de séries do nível hierárquico mais desagregado, para a hierarquia representada na Figura 3.1, temos que $n = 9$ e $m = 6$.

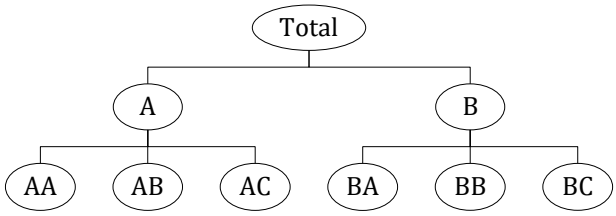

Figura 3.1 – Representação de uma estrutura hierárquica com três níveis

Para cada instante t, a soma das observações das séries do nível inferior da hierarquia é igual às observações da série do nível acima. Desta forma,

$$y_{Total,t} = y_{A,t} + y_{B,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t} + y_{BC,t}, \tag{3.1}$$

visto que,

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \text{ e } y_{B,t} = y_{BA,t} + y_{BB,t} + y_{BC,t}. \tag{3.2}$$

Considere-se o vetor $\boldsymbol{y}_t$ que contém as observações no instante t de todas as séries da hierarquia e o vetor $\boldsymbol{b}_t$ que inclui apenas as observações das séries do nível hierárquico mais desagregado. Neste sentido, as Equações 3.1 e 3.2 podem ser representadas, em termos matriciais, da seguinte forma (Hyndman & Athanasopoulos, 2021):

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t. \tag{3.3}$$

Para o caso da hierarquia representada na Figura 3.1, a Equação 3.3 traduz-se em:

$$\begin{bmatrix} y_{Total,t} \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \end{bmatrix}. \tag{3.4}$$

A matriz $\boldsymbol{S}$ corresponde à matriz soma de dimensão $n \times m$, em que as três primeiras linhas correspondem às restrições de agregação 3.1 e 3.2, e as restantes formam a matriz identidade $\boldsymbol{I}_6$ que assegura que cada observação do nível inferior da hierarquia do lado direito da equação seja igual ao seu valor do lado esquerdo.

3.1.2 Estratégias de reconciliação de previsões

Pretende-se gerar previsões coerentes para cada série da hierarquia, ou seja, previsões que reflitam a estrutura hierárquica e respeitem, por isso, as restrições de agregação da mesma.

Considere-se o vetor $\widehat{\boldsymbol{y}}_{t+h|t}$ que contém as previsões de todas as séries da hierarquia, geradas para o instante $t + h$ a partir das respetivas observações até ao instante t com $h = 1,2,\ldots$. Estas previsões são habitualmente denominadas previsões base, sendo que são geradas de forma independente para cada série da hierarquia. No entanto, dificilmente esta solução garantirá a coerência entre as previsões dos diferentes níveis da hierarquia e, portanto, é necessário recorrer a um método para reconciliar essas previsões, para que reflitam a estrutura de agregação.

De um modo geral, qualquer estratégia de reconciliação pode ser representada por:

$$\widetilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{P}\widehat{\boldsymbol{y}}_{t+h|t}, \tag{3.5}$$

onde $\boldsymbol{P}$ é uma matriz de dimensão $m \times n$ que mapeia as previsões base $\widehat{\boldsymbol{y}}_{t+h|t}$ em previsões reconciliadas para o nível mais desagregado da hierarquia, que são agregadas pela matriz soma $\boldsymbol{S}$, tendo em consideração a estrutura hierárquica, com o objetivo de gerar um conjunto de previsões coerentes $\widetilde{\boldsymbol{y}}_{t+h|t}$. A matriz $\boldsymbol{P}$ depende do método de previsão escolhido e é especificada de seguida para cada caso.

Método *bottom-up*

O método *bottom-up* consiste, primeiramente, em gerar previsões para o nível mais desagregado da hierarquia e, de seguida, agregá-las sucessivamente para obter previsões para os níveis superiores. Neste sentido, para a estrutura hierárquica representada na Figura 3.1, são geradas inicialmente as previsões $\hat{y}_{AA,t+h|t}$, $\hat{y}_{AB,t+h|t}$, $\hat{y}_{AC,t+h|t}$, $\hat{y}_{BA,t+h|t}$, $\hat{y}_{BB,t+h|t}$ e $\hat{y}_{BC,t+h|t}$, que são posteriormente agregadas, tendo em conta as seguintes restrições de agregação:

$$\tilde{y}_{Total,t+h|t} = \hat{y}_{AA,t+h|t} + \hat{y}_{AB,t+h|t} + \hat{y}_{AC,t+h|t} + \hat{y}_{BA,t+h|t} + \hat{y}_{BB,t+h|t} + \hat{y}_{BC,t+h|t}, \quad (3.6a)$$

$$\tilde{y}_{A,t+h|t} = \hat{y}_{AA,t+h|t} + \hat{y}_{AB,t+h|t} + \hat{y}_{AC,t+h|t}, \quad (3.6b)$$

$$\tilde{y}_{B,t+h|t} = \hat{y}_{BA,t+h|t} + \hat{y}_{BB,t+h|t} + \hat{y}_{BC,t+h|t}. \quad (3.6c)$$

Em notação matricial, o conjunto das previsões reconciliadas pode ser representado por:

$$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\hat{\boldsymbol{b}}_{t+h|t}, \quad (3.7)$$

onde o vetor $\hat{\boldsymbol{b}}_{t+h|t}$ inclui as previsões base das séries do nível hierárquico mais desagregado, geradas para o instante $t+h$ a partir das respetivas observações até ao instante t. Estas previsões são agregadas pela matriz soma $\boldsymbol{S}$, tendo em consideração a estrutura hierárquica, a fim de gerar as previsões coerentes $\tilde{\boldsymbol{y}}_{t+h|t}$.

Para a estrutura hierárquica representada na Figura 3.1, a matriz $\boldsymbol{P}$ é dada por:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.8)$$

Generalizando, para qualquer hierarquia tem-se $\boldsymbol{P} = [\boldsymbol{0}_{m\times(n-m)} | \boldsymbol{I}_m]$ onde $\boldsymbol{0}_{m\times(n-m)}$ é a matriz nula de ordem $m \times (n-m)$ e $\boldsymbol{I}_m$ é a matriz identidade de ordem m.

Método *top-down*

O método *top-down* consiste, inicialmente, em gerar previsões para o nível mais agregado e, posteriormente, desagregá-las com base em proporções adequadas. Seja $\boldsymbol{p} = [p_1, \dots, p_m]'$ o vetor de proporções a partir das quais as previsões base são desagregadas, originando as previsões das séries do nível inferior da hierarquia.

Por exemplo, para a estrutura hierárquica representada na Figura 3.1, utilizando as proporções $p_1, \dots, p_6$, temos que:

$$\begin{aligned} &\tilde{y}_{AA,t+h|t} = p_1\hat{y}_{Total,t+h|t},\ \tilde{y}_{AB,t+h|t} = p_2\hat{y}_{Total,t+h|t}, \tilde{y}_{AC,t+h|t} = p_3\hat{y}_{Total,t+h|t}, \\ &\tilde{y}_{BA,t+h|t} = p_4\hat{y}_{Total,t+h|t}, \tilde{y}_{BB,t+h|t} = p_5\hat{y}_{Total,t+h|t}, \tilde{y}_{BC,t+h|t} = p_6\hat{y}_{Total,t+h|t}. \end{aligned} \tag{3.9}$$

Em termos matriciais, as previsões reconciliadas para o nível inferior da hierarquia podem ser representadas da seguinte forma:

$$\widehat{\boldsymbol{b}}_{t+h|t} = \boldsymbol{p}\hat{y}_{Total,t+h|t}. \tag{3.10}$$

Uma vez geradas as previsões $\widehat{\boldsymbol{b}}_{t+h|t}$, estas são agregadas através da matriz $\boldsymbol{S}$, tendo em consideração a estrutura da hierarquia, de modo a obter as previsões para os restantes níveis hierárquicos. Em geral, podemos representar as previsões coerentes obtidas a partir do método *top-down* por:

$$\widetilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{SP}\widehat{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\widehat{\boldsymbol{b}}_{t+h|t}. \tag{3.11}$$

Para a estrutura hierárquica representada na Figura 3.1, a matriz $\boldsymbol{P}$ é dada por:

$$\boldsymbol{P} = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.12}$$

Generalizando, para qualquer hierarquia tem-se $\boldsymbol{P} = [\boldsymbol{p} \mid \boldsymbol{0}_{m\times(n-1)}]$.

Relativamente à definição das proporções de desagregação das previsões base, existem várias abordagens. As mais usadas foram apresentadas por Gross e Sohl (1990) e baseiam-se em dados históricos. A primeira abordagem estabelece que cada proporção p_i corresponde à média das proporções históricas da série do nível inferior da hierarquia, $y_{i,t}$, relativamente à série do nível hierárquico mais agregado, $y_{Total,t}$:

$$p_i = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t}}{y_{Total,t}}, \quad i = 1, \dots, m. \tag{3.13}$$

A segunda determina que cada proporção p_i corresponde à média dos valores históricos da série do nível inferior da hierarquia, $y_{i,t}$, relativa à média dos valores históricos da série do nível hierárquico mais agregado, $y_{Total,t}$:

$$p_i = \sum_{t=1}^{T} \frac{y_{i,t}}{T} / \sum_{t=1}^{T} \frac{y_{Total,t}}{T}, \quad i = 1, \dots, m. \tag{3.14}$$

Contudo, as abordagens que se baseiam em proporções históricas para a desagregação das previsões base não consideram as possíveis variações que essas proporções podem sofrer ao longo do tempo e geram, por isso, previsões com menor precisão nos níveis hierárquicos mais desagregados.

Athanasopoulos et al. (2009) apresentam um método *top-down* que utiliza proporções com base em previsões. Este método gera inicialmente previsões de forma independente para cada série da hierarquia e, de seguida, calcula a proporção da previsão base de cada série de um certo nível da hierarquia na soma das previsões base de todas as séries desse nível:

$$p_i = \prod_{l=0}^{k-1} \frac{\hat{y}_{i,t+h|t}^{(l)}}{\hat{S}_{i,t+h|t}^{(l+1)}}, \quad i = 1, \dots, m, \tag{3.15}$$

onde k é o número de níveis da hierarquia, $\hat{y}_{i,t+h|t}^{(l)}$ é a previsão base, gerada para o instante $t+h$, da série que corresponde ao nó que está l níveis acima do nó i e $\hat{S}_{i,t+h|t}^{(l+1)}$ é a soma das previsões base, geradas para o instante $t+h$, que correspondem aos nós que se encontram abaixo daquele que está $l+1$ níveis acima do nó i.

Estratégias de reconciliação ótima

Hyndman et al. (2011) propõem uma abordagem de reconciliação ótima baseada num modelo de regressão, em que as previsões base são dadas por:

$$\hat{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{\beta}_{t+h|t} + \boldsymbol{\varepsilon}_h, \tag{3.16}$$

onde $\boldsymbol{\beta}_{t+h|t}$ é a média da série mais desagregada da hierarquia e $\boldsymbol{\varepsilon}_h$ representa o erro de reconciliação com média zero e matriz de covariâncias $\sum_h$. Se a matriz $\sum_h$ for conhecida, o estimador de mínimos quadrados generalizados (GLS – *Generalized Least Squares*) de $\boldsymbol{\beta}_{t+h|t}$ pode ser usado para gerar as previsões reconciliadas:

$$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\hat{\boldsymbol{\beta}}_{t+h|t} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}(\boldsymbol{S}'\textstyle\sum_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\textstyle\sum_h^{-1}\hat{\boldsymbol{y}}_{t+h|t}. \tag{3.17}$$

Segundo Hyndman et al. (2011), para que previsões base $\hat{\boldsymbol{y}}_{t+h|t}$ não enviesadas garantam previsões coerentes $\tilde{\boldsymbol{y}}_{t+h|t}$ também não enviesadas, é necessário que $\boldsymbol{SPS} = \boldsymbol{S}$. Esta

condição é satisfeita para esta estratégia de reconciliação ótima, assim como, no caso do método *bottom-up*. No entanto, nenhum método *top-down* permite satisfazer esta condição, o que significa que gerará sempre previsões reconciliadas enviesadas.

Wickramasuriya et al. (2019) demonstram que a estratégia de reconciliação proposta por Hyndman et al. (2011) não é exequível, visto que, em geral, a matriz $\sum_h$ não é conhecida e não pode ser identificada. Segundo os autores, a matriz de covariâncias dos erros das previsões reconciliadas é dada por:

$$\mathrm{Var}\left[y_{t+h} - \tilde{y}_{t+h|t}\right] = \boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}', \tag{3.18}$$

onde $\boldsymbol{W}_h = \mathrm{Var}\left[y_{t+h} - \hat{y}_{t+h|t}\right]$ representa a matriz de variâncias e covariâncias dos erros das previsões base.

Pretende-se encontrar a matriz $\boldsymbol{P}$ que minimiza a variância dos erros das previsões reconciliadas, que se encontram na diagonal da matriz $\mathrm{Var}\left[y_{t+h} - \tilde{y}_{t+h|t}\right]$. A estratégia de reconciliação ótima proposta por Wickramasuriya et al. (2019), referida como MinT, determina, assim, que a matriz $\boldsymbol{P}$ que minimiza o traço da matriz $\mathrm{Var}\left[y_{t+h} - \tilde{y}_{t+h|t}\right]$, tal que $\boldsymbol{SPS} = \boldsymbol{S}$, é dada por:

$$\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}. \tag{3.19}$$

Deste modo, as previsões reconciliadas ótimas são dadas por:

$$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_{t+h|t}. \tag{3.20}$$

De forma a aplicar esta estratégia, é ainda necessário estimar a matriz $\boldsymbol{W}_h$. Wickramasuriya et al. (2019) apresentam algumas hipóteses:

1. $\boldsymbol{W}_h = k_h\boldsymbol{I}_n$ para qualquer valor de h, onde $k_h > 0$ é uma constante de proporcionalidade. O estimador de $\boldsymbol{\beta}_{t+h|t}$ para obter as previsões reconciliadas é o estimador OLS. Esta é a estimação mais simples, onde a matriz $\boldsymbol{P}$ é independente dos dados. Porém, este método de reconciliação não considera as diferenças de escala entre os níveis hierárquicos e as correlações entre as séries.

2. $\boldsymbol{W}_h = k_h\mathrm{diag}(\widehat{\boldsymbol{W}}_1)$ para qualquer valor de h, onde $k_h > 0$ e $\widehat{\boldsymbol{W}}_1 = \frac{1}{T}\sum_{t=1}^{T} e_t e_t'$ é o estimador de covariâncias amostrais dos erros das previsões base para $h = 1$, onde o vetor e_t inclui os erros dos modelos que geraram essas previsões. Deste modo, esta

hipótese utiliza a variância dos erros para escalar as previsões base, sendo o estimador MinT um estimador WLS (*var*).

3. $\boldsymbol{W}_h = k_h \boldsymbol{\Lambda}$ para qualquer valor de h, onde $k_h > 0$, $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{S}\mathbf{1})$ e $\mathbf{1}$ é um vetor unitário de dimensão m. Esta hipótese assume que os erros das previsões base do nível inferior da hierarquia não estão correlacionados entre os nós desse nível e têm variância k_h. O estimador usado, WLS (*struct*), é independente dos dados, estando apenas dependente das restrições de agregação da estrutura hierárquica. Neste sentido, ganha particular interesse quando os resíduos não estão disponíveis.

4. $\boldsymbol{W}_h = k_h \widehat{\boldsymbol{W}}_1$ para qualquer valor de h, onde $k_h > 0$, sendo o estimador de covariâncias amostrais para $h = 1$. Este é facilmente obtido, contudo não é o mais apropriado, quando o número de séries do nível inferior da hierarquia é superior à dimensão da série, isto é, quando $m > T$. Este estimador é designado por MinT (*sample*).

5. $\boldsymbol{W}_h = k_h \widehat{\boldsymbol{W}}^*_{1,D}$ para qualquer valor de h, onde $k_h > 0$. $\widehat{\boldsymbol{W}}^*_{1,D} = \lambda_D \widehat{\boldsymbol{W}}_{1,D} + (1 - \lambda_D)\widehat{\boldsymbol{W}}_1$ é um estimador *shrinkage* que reduz os elementos fora da diagonal da matriz $\widehat{\boldsymbol{W}}_1$ para zero e mantém inalterados os restantes elementos (variâncias), onde $\widehat{\boldsymbol{W}}_{1,D}$ é uma matriz diagonal que inclui as entradas diagonais da matriz $\widehat{\boldsymbol{W}}_1$ e λ_D é o parâmetro de intensidade de *shrinkage*. Schäfer e Strimmer (2005) propõem um estimador *shrinkage*, tal que:

$$\hat{\lambda}_D = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}(\hat{r}_{ij})}}{\sum_{i \neq j} \hat{r}_{ij}^2}, \tag{3.21}$$

onde $\hat{r}_{ij}$ é o elemento na posição ij da matriz $\widehat{\boldsymbol{R}}_1$ que corresponde à matriz de correlações amostrais dos erros das previsões base para $h = 1$. Este estimador é referido como MinT (*shrink*).

3.2 Modelos de previsão de séries temporais

No presente estudo, para gerar as previsões base para os diferentes níveis da hierarquia, consideram-se os modelos ARIMA, pelo bom desempenho demonstrado na previsão de séries temporais de vendas. Em alternativa, considera-se uma DNN, dado o recente sucesso na adoção de métodos de *Deep Learning* para a previsão de vendas. Estes dois modelos são descritos nas secções que se seguem.

3.2.1 Modelos ARIMA

Os modelos ARIMA constituem uma das abordagens tradicionais mais utilizadas para a previsão de séries temporais, dados os bons resultados da sua aplicação e a ampla variedade existente de classes de modelos apropriados para lidar com diferentes características das séries temporais. Contudo, os modelos ARIMA aplicam-se apenas a séries estacionárias, ou seja, a séries cujas propriedades estatísticas (como a média e a variância) não variam com o instante da observação (Hyndman & Athanasopoulos, 2021). Desta forma, séries não estacionárias têm de ser transformadas em séries estacionárias. Para estabilizar a variância pode ser utilizada uma transformação Box-Cox (Box & Cox, 1964):

$$y_t' = \begin{cases} \ln(y_t)\,, & \lambda = 0 \\ (y_t^{\lambda} - 1)/\lambda, & \lambda \neq 0 \end{cases} \tag{3.22}$$

Por outro lado, para estabilizar a média, é utilizado um processo de diferenciação. A diferenciação ordinária consiste em calcular a diferença entre observações consecutivas:

$$y_t' = y_t - y_{t-1}. \tag{3.23}$$

Ao tornar a série estacionária pode ocasionalmente ainda ser necessária uma diferenciação ordinária de 2ª ordem que se obtém da forma seguinte:

$$y_t'' = y_t' - y_{t-1}' = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}. \tag{3.24}$$

Para além disso, se a série temporal for caracterizada por um padrão de sazonalidade, pode ainda ser aplicada uma diferenciação sazonal, que consiste em calcular a diferença entre observações separadas pelo período de sazonalidade m:

$$y_t' = y_t - y_{t-m}. \tag{3.25}$$

O operador atraso B (de *backshift*) é bastante útil para descrever o processo de diferenciação, tendo um efeito de desfasamento na série sobre a qual é aplicado:

$$By_t = y_{t-1}. \tag{3.26}$$

Desta forma, as Equações 3.23 e 3.24 equivalem, respetivamente, às expressões seguintes:

$$y_t' = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t, \tag{3.27}$$

$$y_t'' = y_t - 2y_{t-1} + y_{t-2} = (1 - 2B + B^2)y_t = (1 - B)^2 y_t. \tag{3.28}$$

De forma genérica, uma diferenciação de ordem d pode ser definida por:

$$(1-B)^d y_t. \tag{3.29}$$

Assim, por exemplo, uma diferenciação sazonal seguida de uma diferenciação de 1ª ordem pode ser definida por:

$$\begin{aligned}(1-B)(1-B^m)y_t &= (1-B-B^m+B^{m+1})y_t \\ &= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}.\end{aligned} \tag{3.30}$$

As análises dos gráficos da Função de Autocorrelação (ACF – *AutoCorrelation Function*) e da Função de Autocorrelação Parcial (PACF – *Partial Autocorrelation Function*) podem ajudar a esclarecer se existe necessidade de aplicar alguma diferenciação à série temporal. Em alternativa, podem ser utilizados testes de raiz unitária (*unit root tests)* que consistem em testes estatísticos concebidos para decidir se existe necessidade de diferenciação. É relevante salientar que a diferenciação sazonal deve ser realizada em primeiro lugar, especialmente, quando a série temporal for caracterizada por fortes padrões de sazonalidade, visto que, por vezes, essa diferenciação já torna a série estacionária, não havendo necessidade de aplicar diferenciação ordinária. De entre os vários testes de raiz unitária existentes, podem ser usados os testes *Canova-Hansen* (Canova & Hansen, 1995) para determinar o grau de diferenciação sazonal. Após escolher esse valor, podem ser usados sucessivamente os testes *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) (Kwiatkowski et al., 1992) para selecionar o grau de diferenciação ordinária adequado.

Os modelos ARIMA envolvem uma parte autorregressiva (que gera previsões de determinada variável com base em valores passados da mesma) de ordem p, uma diferenciação ordinária de ordem d e uma parte de médias móveis (que gera previsões com base em erros de previsão passados) de ordem q. Deste modo, os modelos ARIMA não sazonais, designados por ARIMA(p,d,q), são representados por (Brockwell & Davis, 2016):

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \tag{3.31}$$

Utilizando o operador atraso B, pode-se reescrever a Equação 3.31 da seguinte forma:

$$\left(1-\phi_1 B - \cdots - \phi_p B^p\right)(1-B)^d y_t = c + (1+\theta_1 B + \cdots + \theta_q B^q)\varepsilon_t, \tag{3.32}$$

onde d é o grau de diferenciação ordinária, $\phi_1, \dots, \phi_p$ são os parâmetros autorregressivos, $\theta_1, \dots, \theta_q$ são os parâmetros de médias móveis, $c = \mu(1 - \phi_1 - \cdots - \phi_p)$, sendo μ a média de $(1 - B)^d y_t$, e ε_t representa erros de média zero e variância constante σ^2.

Os modelos ARIMA podem também ser aplicados a séries temporais caracterizadas por padrões de sazonalidade. Nesse caso, os modelos ARIMA sazonais, designados por ARIMA$(p, d, q)(P, D, Q)_m$ são representados por:

$$\phi_p(B)\Phi_P(B^m)(1 - B)^d(1 - B^m)^D y_t = c + \theta_q(B)\Theta_Q(B^m)\varepsilon_t, \tag{3.33}$$

onde

$$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \text{ e } \Phi_P(B^m) = 1 - \Phi_1 B^m - \cdots - \Phi_P B^{Pm},$$

$$\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q \text{ e } \Theta_Q(B^m) = 1 + \Theta_1 B^m + \cdots + \Theta_Q B^{Qm},$$

m é o período de sazonalidade, D é o grau de diferenciação sazonal, e $c = \mu(1 - \phi_1 - \cdots - \phi_p)(1 - \Phi_1 - \cdots - \Phi_p)$, sendo μ a média de $(1 - B)^d(1 - B^m)^D y_t$. Os restantes parâmetros têm o significado identificado na Equação 3.32.

Relativamente aos graus de diferenciação (d e D) dos modelos ARIMA, podemos obter valores apropriados para os mesmos, como já mencionado, através de testes de raiz unitária. No caso dos parâmetros p, q, P e Q, estes podem ser selecionados através de critérios de informação, nomeadamente através do critério de informação de Akaike (AIC – *Akaike's Information Criterion*), critério de informação de Akaike corrigido e critério de informação Bayesiano (BIC – *Bayesian Information Criterion*).

O AIC é definido por:

$$\text{AIC} = -2\log(L) + 2(p + q + P + Q + k + 1), \tag{3.34}$$

onde L é a verosimilhança dos dados, $k = 1$ se $c \neq 0$ e $k = 0$, caso contrário. O último termo dentro de parênteses corresponde ao número total de parâmetros do modelo, incluindo a variância dos erros (σ^2).

O AIC corrigido, adequado para amostras de dimensão (T) pequena, é dado por:

$$\text{AICc} = \text{AIC} + \frac{2(p + q + P + Q + k + 1)(p + q + P + Q + k + 2)}{T - p - q - P - Q - k - 2}. \tag{3.35}$$

O BIC é definido por:

$$\text{BIC} = \text{AIC} + [\log(T) - 2](p + q + P + Q + k + 1). \tag{3.36}$$

Deste modo, podem ser encontrados valores adequados para os parâmetros p, q, P e Q através da minimização de AIC, AICc ou BIC. De notar que os valores destes critérios apenas são comparáveis entre modelos com as mesmas ordens de diferenciação.

Após a identificação do modelo ARIMA$(p, d, q)(P, D, Q)_m$, resta obter valores para os parâmetros c, $\phi_1, \dots, \phi_p$, $\Phi_1, \dots, \Phi_P$, $\theta_1, \dots, \theta_q$, $\Theta_1, \dots, \Theta_Q$ que são geralmente estimados por máxima verosimilhança.

Modelos ARIMAX

Os modelos ARIMA, como anteriormente mencionado, utilizam valores passados da série temporal para efeitos de previsão da mesma. No entanto, sendo modelos univariados, não admitem a inclusão de variáveis externas que podem ser determinantes para a explicação de variações da série temporal. No caso do setor do retalho, fatores como promoções, preço, eventos de calendário e condições meteorológicas podem explicar parte das variações das vendas e melhorar a precisão das previsões. Nesse sentido, os modelos ARIMA podem ser estendidos de forma a incluir fatores externos, obtendo-se o modelo de regressão com erros ARIMA, designado por ARIMAX:

$$\begin{gathered} y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \eta_t, \\ \phi_p(B)\Phi_P(B^m)(1 - B)^d(1 - B^m)^D \eta_t = \theta_q(B)\Theta_Q(B^m)\varepsilon_t, \end{gathered} \tag{3.37}$$

onde y_t corresponde à série temporal, $x_{1,t}, \dots, x_{k,t}$ correspondem às séries temporais das k variáveis explicativas, $\beta_0, \dots, \beta_k$ são os parâmetros do modelo de regressão e η_t é a série dos erros que é modelizada por um modelo ARIMA.

Da mesma forma que nos modelos ARIMA, os parâmetros do modelo ARIMAX podem ser estimados por máxima verosimilhança. Alternativamente, podem-se obter estimativas idênticas minimizando a soma dos quadrados dos erros ε_t. Contudo, para que estas estimativas sejam consistentes, é necessário, primeiramente, garantir que a série temporal e a série das variáveis explicativas são estacionárias. No sentido de manter a relação entre y_t e $x_{1,t}, \dots, x_{k,t}$, deve ser aplicada a mesma diferenciação a todas as variáveis.

3.2.2 *Dense Neural Networks*

Os algoritmos de *Deep Learning* consistem numa classe de algoritmos de *Machine Learning* com estrutura lógica semelhante à do cérebro humano, sendo capazes de considerar várias camadas de transformação de dados para obter a representação ótima dos mesmos. A aprendizagem geralmente ocorre através de *multi-layer neural networks*. A estrutura mais simples de uma NN, representada na Figura 3.2, inclui uma camada de entrada (que corresponde às variáveis originais), uma camada oculta (onde ocorre a maior parte do processo de aprendizagem) e uma camada de saída (que fornece as previsões).

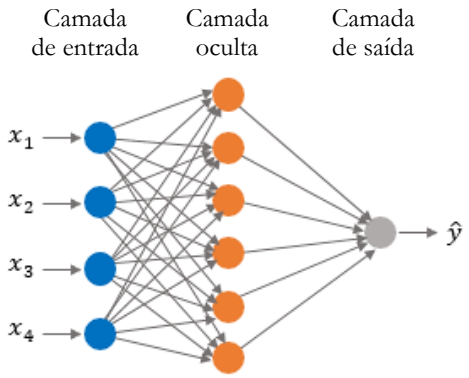

Figura 3.2 – Representação de uma NN simples

Numa DNN, as camadas estão densamente conectadas, o que significa que cada camada contém um certo número de nós que se encontram ligados a cada nó da camada anterior. Qualquer DNN em que se verifique que as entradas influenciam de forma sucessiva as camadas ocultas e estas, por sua vez, influenciam a camada de saída, designa-se por *feedforward neural network*.

Não está estabelecida nenhuma abordagem para definir o número de camadas e nós de uma DNN. No entanto, normalmente, o número de nós em cada camada oculta é fortemente influenciado pelo número de variáveis originais que compõem a camada de entrada. Ainda assim, é relevante notar que quanto maior for a complexidade das DNNs, ou seja, o número de camadas ocultas e respetivos nós, maior será a capacidade das DNNs para aprender novas características dos dados. Por outro lado, a especificação da camada de saída está dependente da natureza do problema em questão (regressão, classificação binária ou classificação multidimensional), dados os diferentes formatos de resultado final originados. No caso de um problema de regressão, em que o propósito é a previsão de um

valor contínuo, a camada de saída é composta por apenas um nó que fornece a previsão final.

Além disso, é necessário clarificar o conceito de ativação para a construção das DNNs. Como já mencionado, os nós de uma camada estão ligados a todos os nós da camada anterior. A cada uma dessas conexões é atribuído um peso, o que significa que cada nó soma as saídas da camada anterior ponderadas pelo respetivo peso de conexão (w) e adiciona ainda um termo extra de enviesamento (b). Por sua vez, as somas realizadas por cada nó constituem as entradas de uma função de ativação. Esta função determina, assim, se a informação num nó é suficiente para que este provoque um sinal para a camada seguinte. Dentro das várias funções de ativação existentes, as mais usadas são a *Rectified Linear Unit* (ReLU), a sigmoide e a tangente hiperbólica (tanh):

$$\mathrm{ReLU}(x) = \max(0, x) \tag{3.38}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.39}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.40}$$

Assim, para uma função de ativação $f(.)$, em cada nó i de uma certa camada é realizada a seguinte operação:

$$a_i = f(Wx + b_i), \tag{3.41}$$

onde x representa o conjunto de saídas da camada anterior, W representa o conjunto de pesos do nó i e b_i é o termo de enviesamento do nó i.

Para treinar as DNNs, inicialmente são geradas previsões, com base numa parte dos dados (denominada por *batch*) e em valores aleatórios atribuídos aos pesos das conexões entre os nós e aos termos de enviesamento. Os erros de previsão são calculados através de uma função objetivo. No caso dos problemas de regressão, a função objetivo mais comum é o *Mean Squared Error* dado por:

$$L = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2, \tag{3.42}$$

onde y_t é o conjunto de valores observados no instante t e $\hat{y}_t$ é o conjunto das previsões geradas para o instante t.

Posteriormente, as DNNs utilizam o algoritmo *backpropagation* que calcula os gradientes da função objetivo em relação aos parâmetros, $\frac{\partial L}{\partial w}$ e $\frac{\partial L}{\partial b}$, de forma a os ajustar no sentido oposto ao do gradiente (isto é, no sentido de maior redução da função objetivo) e consequentemente a melhorar as previsões. Os novos parâmetros são utilizados numa segunda tentativa de previsão que utiliza outro *batch*. Este processo, designado por *Stochastic Gradient Descent* (SGD), é realizado de forma sucessiva, até que a função objetivo seja minimizada. Quando a DNN já tiver sido treinada com base em todos os *batches*, completa-se uma época. O número de épocas necessárias para minimizar a função objetivo está dependente da taxa de aprendizagem do otimizador SGD, ou seja, da velocidade com que se desce os gradientes e se atinge o mínimo da função objetivo.

Durante o processo SGD, os parâmetros de cada camada são atualizados com a expectativa de que as entradas nessa camada têm uma certa distribuição, ou seja, que os parâmetros das camadas anteriores se mantêm. No entanto, como os parâmetros de todas as camadas são atualizados simultaneamente, ocorrem alterações na distribuição das entradas de cada camada, o que afeta negativamente as taxas de aprendizagem. Para solucionar este problema, pode ser aplicado o processo *batch normalization* de redimensionamento das entradas de cada camada, para que as estatísticas (média e variância) das mesmas assumam determinados valores. Deste modo, as expectativas que cada camada tem sobre a distribuição das saídas da camada anterior não serão alteradas. Esta técnica torna, assim, o processo de treino das DNNs mais eficiente.

Outra questão relevante está relacionada com a capacidade de generalização das DNNs, ou seja, a capacidade de terem um bom desempenho ao utilizarem novos dados de entrada, distintos dos que foram usados durante o treino. Quando uma DNN não aprende de forma suficiente as relações entre as entradas, tanto do conjunto de treino como de teste, e a variável a ser explicada, estamos perante um cenário de *underfitting*. Este problema pode ser facilmente solucionado através de um aumento da complexidade da DNN, ou seja, do número de camadas e de nós. Por outro lado, quando uma DNN é complexa e aprende bastante bem as relações entre as entradas do conjunto de treino e a variável a ser explicada, mas o mesmo não acontece ao usar novas entradas do conjunto de teste, estamos perante um cenário de *overfitting*. Isto significa que a DNN tem uma variância elevada, o que pode ser constatado pela elevada dimensão dos pesos que indica que pequenas alterações nas entradas podem levar a grandes alterações nas saídas da DNN.

Neste caso, a solução mais comum consiste na aplicação de um método de regularização para que os pesos da DNN se mantenham pequenos. A utilização deste método, para além de reduzir o erro de generalização, possibilita a atribuição de pesos menores a variáveis de entrada menos relevantes. Um método de regularização simples frequentemente utilizado corresponde a adicionar uma penalização à função objetivo que seja proporcional ao tamanho dos pesos, sendo que estes são usualmente calculados através da abordagem L1 (soma dos valores absolutos dos pesos) ou da L2 (soma dos quadrados dos pesos). Outras estratégias habitualmente utilizadas para resolver o problema de *overfitting* são o *early-stopping*, que corresponde a parar de treinar a DNN quando o erro de validação começa a aumentar, e o *dropout*, que consiste em desativar de forma aleatória um número fixo de nós da DNN durante o treino, de modo a evitar que a mesma se torne dependente de nós específicos para extrair resultados.

4. Caso de estudo

4.1 Conjunto de dados

O grupo Jerónimo Martins é uma empresa internacional com sede em Portugal, com grande experiência no setor do retalho. A sua atividade principal é a distribuição alimentar, sendo que, em Portugal, a cadeia de lojas Pingo Doce lidera o segmento de supermercados.

O presente caso de estudo pretende integrar informação adicional, nomeadamente informação de atividade promocional, nos métodos de previsão hierárquica, de forma a aumentar a precisão das previsões de vendas no setor do retalho. Para o efeito, foi disponibilizado pelo Grupo Jerónimo Martins um conjunto de dados de base diária relativa a vendas e a preços de produtos em lojas Pingo Doce. Os dados são relativos ao período entre 3 de janeiro de 2012 e 27 de abril de 2015, totalizando 1211 dias.

Do total de SKUs, foram selecionados 38 SKUs comuns a 10 lojas Pingo Doce e pertencentes a cinco áreas principais: perecíveis não especializados (14), mercearia (13), bebidas (6), perecíveis especializados (3) e detergentes e produtos de limpeza (2). Os SKUs em análise exibem padrões de vendas e comportamentos face a alterações de preço distintos. Esta diversidade dos dados permite alcançar uma maior fiabilidade nos resultados e, consequentemente, uma maior eficácia da avaliação do desempenho de previsão dos modelos considerados.

Para cada SKU, foi desenhada uma hierarquia com três níveis de agregação (Figura 4.1). No nível superior da hierarquia (nível 0), está representada a série do total de vendas que se divide nas séries de vendas das regiões A e B (nível 1). No nível mais desagregado da hierarquia (nível 2), estão representadas as séries de vendas das 10 lojas consideradas, sendo que duas pertencem à região A e as restantes oito pertencem à região B.

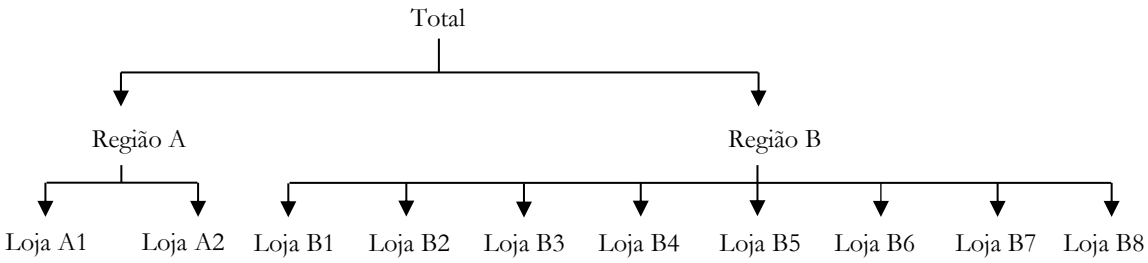

Figura 4.1 – Estrutura hierárquica de cada SKU

Desta forma, cada hierarquia é composta, no total, por 13 séries temporais. O número de séries por nível hierárquico está representado na Tabela 4.1.

Tabela 4.1 – Número de séries por nível de agregação

Nível hierárquico	**Número de séries**
Nível 0	1
Nível 1	2
Nível 2	10
Total	13

De forma a ilustrar como se processa a agregação dos dados, a Figura 4.2 apresenta as séries das vendas de um SKU representativo para os três níveis hierárquicos.

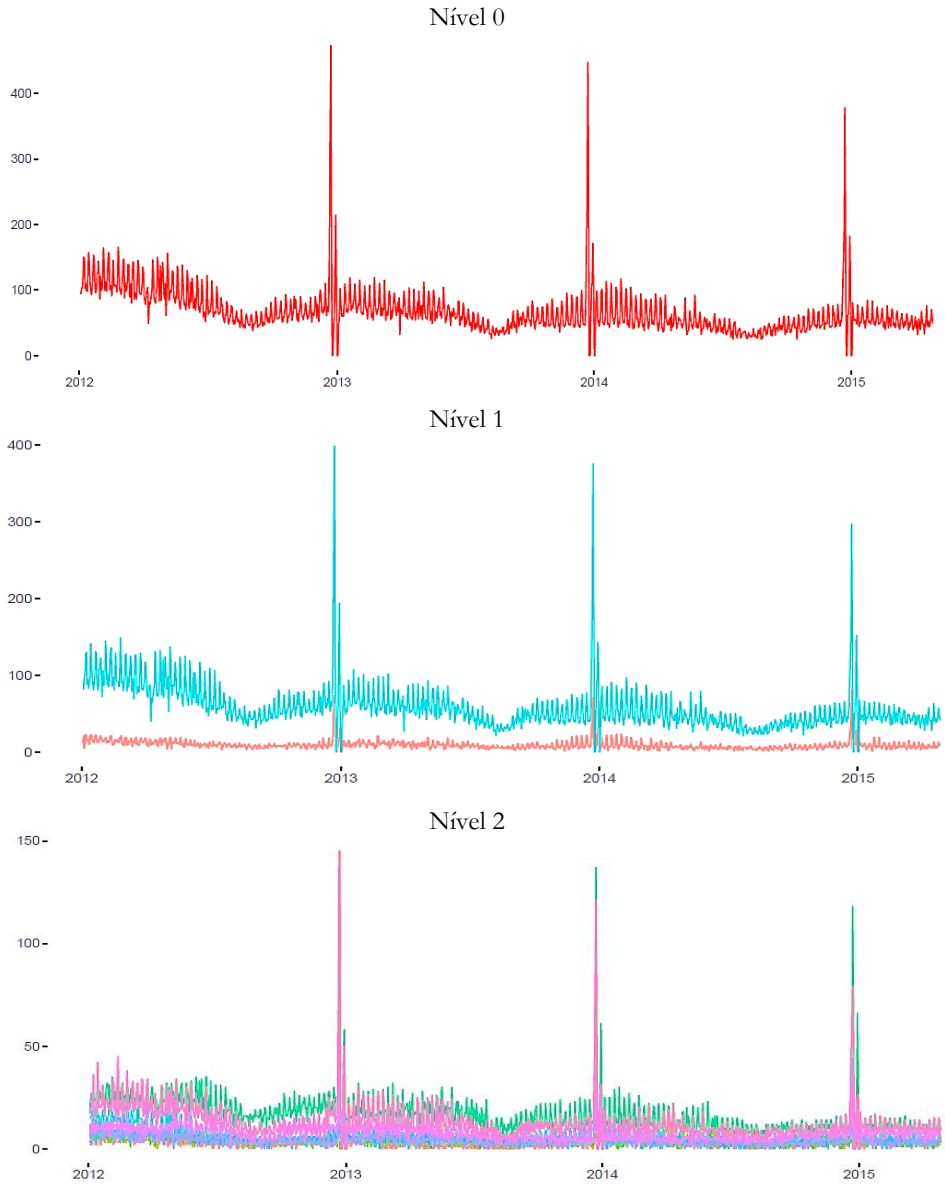

Figura 4.2 – Vendas por nível hierárquico de um SKU representativo

4.2 Procedimento metodológico

Após a definição da estrutura de agregação dos dados de cada SKU, são geradas previsões base para todas as séries das hierarquias. Para o efeito, são utilizados os modelos ARIMA e uma DNN. A estimação dos modelos ARIMA e a produção das respetivas previsões foram realizadas através das funções *ARIMA*() e *forecast*() dos pacotes *fable* e *fabletools* do *software* estatístico R. Para a construção da DNN, foram utilizados o *keras* e o *Tensorflow* da linguagem de programação Python. A DNN utilizada é formada pela camada de entrada, pelas camadas ocultas e pela camada de saída. Para o treino da DNN, foram utilizadas 100 épocas, uma janela deslizante de dimensão igual a 28 dias e a função de ativação ReLU. Através de uma estratégia de *grid search*, foram escolhidos os seguintes parâmetros de configuração: número de camadas ocultas igual a 1, número de nós igual a 80, dimensão de cada *batch* igual a 2, otimizador SGD com taxa de aprendizagem de 0.001 e função objetivo Huber. Para controlar o *overfitting*, foram utilizadas as estratégias de *early-stopping* e regularização L2 sobre a camada de entrada com parâmetro igual a 0.001.

Relativamente às variáveis consideradas no processo de previsão, numa primeira fase, utilizaram-se apenas dados históricos de vendas. Posteriormente, de forma a avaliar se a integração no processo de previsão de fatores determinantes das vendas permitem melhorar o desempenho de previsão, foram incorporados regressores nos dois modelos considerados: o preço de cada SKU; variáveis binárias que representam os dias da semana de segunda a domingo; e variáveis binárias que representam os eventos de calendário Natal e Páscoa. O preço constitui uma variável promocional visto que descidas de preço temporárias indicam a realização de promoções. Os dias da semana são incluídos devido à sazonalidade semanal dos dados, enquanto que os eventos de calendário são considerados por corresponderem a épocas em que se verifica um aumento nas vendas.

Além disso, dado o objetivo de gerar um conjunto de previsões coerentes, tendo em consideração a estrutura hierárquica adotada, as previsões base foram reconciliadas usando os métodos *bottom-up* e *top-down* (proposto por Athanasopoulos et al. (2009), que se baseia em proporções calculadas através de previsões) e usando estratégias de reconciliação ótima com os estimadores OLS e WLS (*struct*). Estas estratégias de reconciliação foram implementadas usando as funções *htsrec*() e *tdrec*() do pacote *FoReco* do *software* estatístico R.

Para poder avaliar o desempenho de um modelo de previsão é habitual dividir-se o conjunto de dados em duas partes: conjunto de treino e conjunto de teste. O conjunto de

treino é utilizado para estimar os parâmetros do modelo, enquanto que o conjunto de teste é usado para avaliar a precisão das previsões geradas por esse modelo. No caso em estudo, para os modelos ARIMA, o conjunto de treino foi composto por 1085 dias, consistindo no período entre 3 de janeiro de 2012 e 22 de dezembro de 2014. O conjunto de teste consistiu no período remanescente, isto é, entre 23 de dezembro de 2014 e 27 de abril de 2015, perfazendo um total de 126 dias. No caso da DNN o conjunto de treino foi composto apenas pelos primeiros 847 dias, visto que os 238 dias seguintes foram para o conjunto de validação, usado nos processos de *grid search* e *early-stopping*. O conjunto de teste foi o mesmo que se usou nos modelos ARIMA.

As previsões foram geradas com base numa abordagem de origem móvel, designada por *rolling*, tal como está ilustrado na Figura 4.3. Em cada iteração, o conjunto de treino está representado pelos círculos azuis, enquanto que os círculos a vermelho representam o conjunto de teste. Como se pode observar, o conjunto de treino é sucessivamente atualizado sendo-lhe adicionado um dia em cada iteração. Deste modo, este processo altera em cada iteração a origem da previsão a partir da qual são geradas previsões para um horizonte de sete dias.

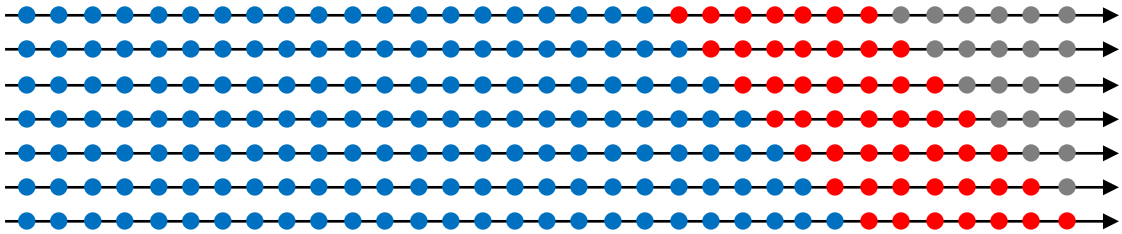

Figura 4.3 – Abordagem *rolling*

Dado um conjunto de treino $(y_1, y_2, \dots, y_T)$ e um conjunto de teste $(y_{T+1}, y_{T+2}, \dots)$, para um dado instante $T + h$, onde $h = 1,2, \dots$ representa o horizonte temporal de previsão, o erro de previsão corresponde à diferença entre o valor observado e o valor previsto:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}. \tag{4.1}$$

Visto que as escalas dos produtos diferem entre si, por forma a poderem-se comparar os erros de previsão, o desempenho dos vários modelos considerados é avaliado a partir de medidas de precisão baseadas em erros escalados nomeadamente o *Mean Absolute Scaled Error* (MASE), proposto por Hyndman e Koehler (2006), e o *Root Mean Squared Scaled Error* (RMSSE).

Como os dados são diários e existe sazonalidade semanal é natural escalar os erros através do *Mean Absolute Error* (MAE) do método naïve sazonal no conjunto de treino:

$$q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}, \tag{4.2}$$

onde m representa o período de sazonalidade que, neste caso, é igual a sete.

A partir de q_j, obtêm-se as medidas de precisão das previsões:

$$\begin{aligned} \text{MASE} &= média(|q_j|) \\ \text{RMSSE} &= \sqrt{média({q_j}^2)} \end{aligned} \tag{4.3}$$

4.3 Resultados

A Figura 4.4 mostra a evolução do treino da DNN com e sem regressores para um SKU típico. Pode observar-se a diminuição progressiva do valor da função objetivo e do MAE nos conjuntos de treino e teste à medida que o número de épocas aumenta, sendo essa diminuição mais acentuada no início do treino. Atende-se que no caso da DNN o treino atingiu o número máximo de épocas, enquanto que no caso da DNN com regressores o processo de *early-stopping* parou o treino ao fim de 29 épocas. Tal como é espectável, o valor da função objetivo e do MAE relativos ao conjunto de treino são inferiores aos respetivos valores no conjunto de validação uma vez que a rede desconhece as observações deste último conjunto.

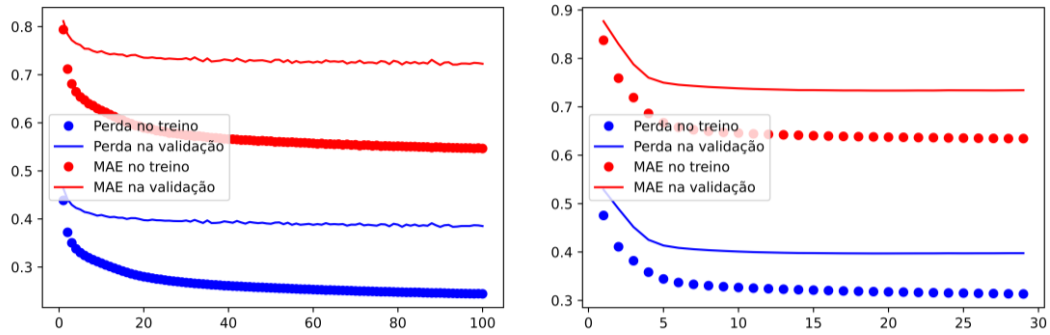

Figura 4.4 – Valor da função objetivo (a azul) e do MAE (a vermelho) nos conjuntos de treino e validação para a DNN (à esquerda) e a DNN com regressores (à direita) de um SKU típico de uma das lojas

A Figura 4.5 mostra a série das vendas de um SKU típico numa das lojas e as respetivas previsões para o conjunto de teste obtidas usando os modelos ARIMA, ARIMAX, DNN e DNN com regressores. Como se pode observar todos os modelos têm alguma dificuldade em acompanhar a elevada volatilidade da série, sendo que o ajustamento das previsões é ligeiramente melhor no caso dos modelos com regressores.

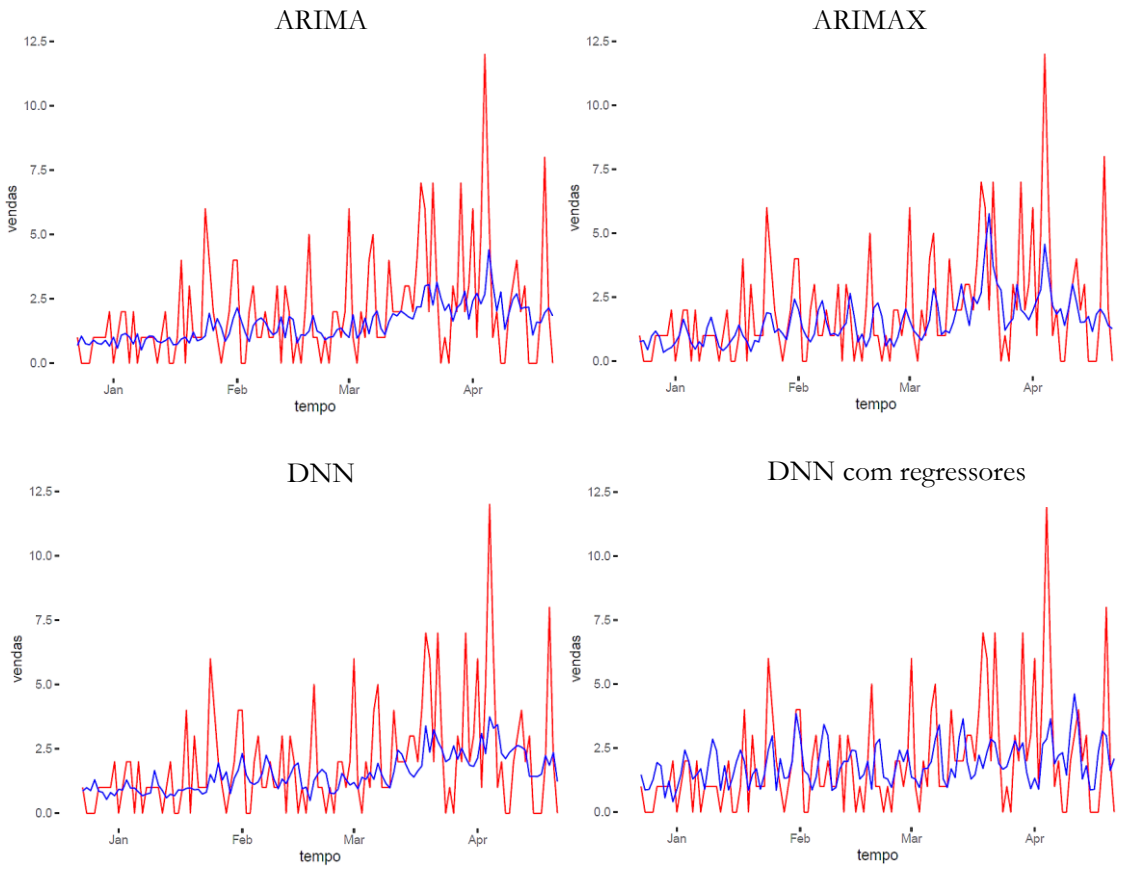

Figura 4.5 – Série das vendas (a vermelho) e das vendas previstas (a azul) pelos modelos ARIMA, ARIMAX, DNN e DNN com regressores para um SKU típico de umas das lojas

As Tabelas 4.2 a 4.5 apresentam os valores das medidas de precisão das previsões base e das previsões reconciliadas por nível hierárquico para os quatro modelos considerados. Em cada tabela, as sete primeiras colunas correspondem às medidas de precisão das previsões para horizontes entre um e sete dias, enquanto que a última coluna apresenta a média dessas medidas de precisão. Os erros dos modelos ARIMAX são notoriamente menores do que os erros dos modelos ARIMA, para todos os horizontes de previsão e em todos os níveis hierárquicos. Deste modo, conclui-se que a integração dos regressores nos modelos ARIMA permite aumentar a precisão das previsões.

Comparando os resultados dos modelos ARIMA e da DNN, verifica-se que, para horizontes de previsão superiores a um dia, a DNN supera os modelos ARIMA. Este é um resultado importante para o setor do retalho, dada a necessidade de os retalhistas realizarem o planeamento dos seus *stocks* com alguma antecedência. Além disso, para horizontes de previsão superiores a um dia, o ganho em termos de precisão das previsões da DNN sobre os modelos ARIMA é maior quando as previsões base são reconciliadas, revelando, assim, um impacto positivo das estratégias de reconciliação.

A integração de regressores na DNN, tal como nos modelos ARIMA, tinha como objetivo uma melhoria da precisão das previsões, contudo tal não se comprovou. Na realidade, enquanto os modelos ARIMAX utilizam o conhecimento dos valores dos regressores nos instantes para os quais geram as previsões, as DNNs utilizam esses valores só nos instantes passados. Esta desigualdade na forma como os regressores são utilizados constitui uma possível explicação para a DNN não ter beneficiado da incorporação dos regressores, ao contrário dos modelos ARIMA.

Outras arquiteturas de NNs mais complexas podem conseguir alcançar melhores resultados com a integração de regressores no processo de previsão. No entanto, provavelmente isso acarreta um custo de processamento mais elevado e um tempo de execução maior, o que não é o ideal para a área do retalho, onde existe um elevado número de SKUs e uma necessidade de obtenção imediata de previsões.

Relativamente às estratégias de reconciliação, em geral, a estratégia de reconciliação ótima com estimador WLS (*struct*) é a que exibe melhor desempenho, sendo superada em alguns casos pelo método *bottom-up*, especialmente no nível inferior da hierarquia.

Tabela 4.2 – MASE e RMSSE das previsões geradas pelos modelos ARIMA

				ARIMA				
MASE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
				Nível 0: Total				
Base	0.9811	1.0442	1.0650	1.0773	1.0950	1.1047	1.1041	1.0673
Bottom-up	1.0090	1.0438	1.0398	1.0327	**1.0336**	**1.0292**	**1.0228**	1.0301
Top-down (prev)	0.9811	1.0442	1.0650	1.0773	1.0950	1.1047	1.1041	1.0673
OLS	0.9671	1.0260	1.0463	1.0531	1.0670	1.0743	1.0735	1.0439
WLS (*struct*)	**0.9609**	**1.0131**	**1.0256**	**1.0288**	1.0376	1.0407	1.0382	**1.0207**
				Nível 1: Região				
Base	0.8975	0.9381	0.9496	0.9502	0.9570	0.9592	0.9590	0.9444
Bottom-up	0.9157	0.9439	0.9427	**0.9388**	**0.9409**	**0.9377**	**0.9337**	0.9362
Top-down (prev)	0.9118	0.9595	0.9733	0.9814	0.9921	0.9987	0.9979	0.9735
OLS	0.8976	0.9455	0.9605	0.9660	0.9761	0.9811	0.9812	0.9583
WLS (*struct*)	**0.8875**	**0.9274**	**0.9367**	0.9399	0.9471	0.9490	0.9479	**0.9336**
				Nível 2: Loja				
Base	**0.7974**	**0.8113**	**0.8106**	**0.8109**	**0.8120**	**0.8107**	**0.8083**	**0.8088**
Bottom-up	**0.7974**	**0.8113**	**0.8106**	**0.8109**	**0.8120**	**0.8107**	**0.8083**	**0.8088**
Top-down (prev)	0.8136	0.8412	0.8523	0.8627	0.8724	0.8796	0.8810	0.8575
OLS	0.8115	0.8402	0.8523	0.8611	0.8697	0.8762	0.8775	0.8555
WLS (*struct*)	0.7981	0.8221	0.8307	0.8369	0.8433	0.8470	0.8473	0.8322

RMSSE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
				Nível 0: Total				
Base	1.6757	1.7551	1.7809	1.7963	1.8277	1.8445	1.8332	1.7876
Bottom-up	1.7602	1.7792	1.7691	1.7734	1.7792	1.7709	1.7613	1.7705
Top-down (prev)	1.6757	1.7551	1.7809	1.7963	1.8277	1.8445	1.8332	1.7876
OLS	**1.6607**	1.7311	1.7557	1.7629	1.7869	1.8006	1.7921	1.7557
WLS (*struct*)	1.6637	**1.7140**	**1.7264**	**1.7329**	**1.7505**	**1.7567**	**1.7481**	**1.7275**
				Nível 1: Região				
Base	1.4903	1.5373	1.5538	1.5519	1.5621	1.5670	1.5639	1.5466
Bottom-up	1.5476	1.5609	1.5555	1.5617	1.5673	1.5619	1.5547	1.5585
Top-down (prev)	1.5004	1.5575	1.5737	1.5865	1.6105	1.6230	1.6152	1.5810
OLS	1.4836	1.5377	1.5567	1.5645	1.5854	1.5944	1.5885	1.5587
WLS (*struct*)	**1.4825**	**1.5185**	**1.5273**	**1.5351**	**1.5496**	**1.5538**	**1.5480**	**1.5307**
				Nível 2: Loja				
Base	1.2645	1.2757	**1.2723**	**1.2764**	**1.2796**	**1.2785**	**1.2736**	**1.2744**
Bottom-up	1.2645	1.2757	**1.2723**	**1.2764**	**1.2796**	**1.2785**	**1.2736**	**1.2744**
Top-down (prev)	1.2591	1.2999	1.3147	1.3285	1.3455	1.3578	1.3541	1.3228
OLS	1.2494	1.2897	1.3071	1.3173	1.3310	1.3415	1.3387	1.3107
WLS (*struct*)	**1.2381**	**1.2672**	1.2771	1.2860	1.2963	1.3030	1.2998	1.2811

Tabela 4.3 – MASE e RMSSE das previsões geradas pelos modelos ARIMAX

				ARIMAX				
MASE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
				Nível 0: Total				
Base	0.7795	0.7863	0.7875	0.7799	0.7795	0.7756	0.7720	0.7800
Bottom-up	0.8141	0.8221	0.8221	0.8157	0.8161	0.8163	0.8151	0.8173
Top-down (prev)	0.7795	0.7863	0.7875	0.7799	0.7795	0.7756	0.7720	0.7800
OLS	0.7671	0.7744	0.7763	0.7693	0.7692	0.7659	0.7625	0.7692
WLS (*struct*)	**0.7587**	**0.7671**	**0.7697**	**0.7636**	**0.7643**	**0.7630**	**0.7609**	**0.7639**
				Nível 1: Região				
Base	0.7406	0.7463	0.7457	0.7423	0.7433	0.7408	0.7377	0.7424
Bottom-up	0.7625	0.7679	0.7682	0.7643	0.7664	0.7666	0.7657	0.7659
Top-down (prev)	0.7532	0.7585	0.7579	0.7544	0.7545	0.7510	0.7483	0.7540
OLS	0.7422	0.7482	0.7473	0.7439	0.7445	0.7417	0.7395	0.7439
WLS (*struct*)	**0.7329**	**0.7386**	**0.7391**	**0.7361**	**0.7372**	**0.7361**	**0.7340**	**0.7363**
				Nível 2: Loja				
Base	**0.7268**	**0.7301**	**0.7291**	**0.7301**	**0.7318**	**0.7309**	**0.7295**	**0.7298**
Bottom-up	**0.7268**	**0.7301**	**0.7291**	**0.7301**	**0.7318**	**0.7309**	**0.7295**	**0.7298**
Top-down (prev)	0.7479	0.7513	0.7506	0.7511	0.7523	0.7505	0.7489	0.7504
OLS	0.7471	0.7502	0.7487	0.7492	0.7505	0.7490	0.7470	0.7488
WLS (*struct*)	0.7330	0.7362	0.7352	0.7360	0.7376	0.7363	0.7344	0.7355

RMSSE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
				Nível 0: Total				
Base	1.3318	1.3396	1.3441	1.3459	1.3519	1.3415	1.3320	1.3410
Bottom-up	1.4153	1.4075	1.4088	1.4093	1.4165	1.4141	1.4077	1.4113
Top-down (prev)	1.3318	1.3396	1.3441	1.3459	1.3519	1.3415	1.3320	1.3410
OLS	1.3207	1.3258	1.3298	1.3309	1.3375	1.3288	1.3187	1.3274
WLS (*struct*)	**1.3196**	**1.3205**	**1.3239**	**1.3249**	**1.3323**	**1.3260**	**1.3169**	**1.3234**
				Nível 1: Região				
Base	1.2312	1.2279	1.2288	1.2304	1.2387	1.2344	1.2273	1.2312
Bottom-up	1.2811	1.2714	1.2710	1.2741	1.2814	1.2794	1.2755	1.2763
Top-down (prev)	1.2395	1.2404	1.2430	1.2468	1.2544	1.2473	1.2418	1.2447
OLS	1.2244	1.2244	1.2258	1.2293	1.2368	1.2305	1.2247	1.2280
WLS (*struct*)	**1.2208**	**1.2171**	**1.2182**	**1.2217**	**1.2297**	**1.2253**	**1.2196**	**1.2218**
				Nível 2: Loja				
Base	1.1414	1.1400	1.1375	1.1417	1.1455	1.1442	1.1411	1.1416
Bottom-up	1.1414	1.1400	1.1375	1.1417	1.1455	1.1442	1.1411	1.1416
Top-down (prev)	1.1440	1.1476	1.1488	1.1515	1.1563	1.1529	1.1496	1.1501
OLS	1.1329	1.1355	1.1328	1.1371	1.1412	1.1382	1.1341	1.1360
WLS (*struct*)	**1.1212**	**1.1226**	**1.1200**	**1.1245**	**1.1287**	**1.1264**	**1.1224**	**1.1237**

Tabela 4.4 – MASE e RMSSE das previsões geradas pela DNN

DNN								
MASE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
Nível 0: Total								
Base	1.0199	1.0363	1.0260	1.0077	1.0131	1.0192	1.0110	1.0190
Bottom-up	1.0046	1.0109	1.0003	0.9931	0.9966	0.9980	0.9900	0.9991
Top-down (prev)	1.0199	1.0363	1.0260	1.0077	1.0131	1.0192	1.0110	1.0190
OLS	0.9999	1.0132	1.0071	0.9915	0.9948	1.0007	0.9931	1.0001
WLS (*struct*)	**0.9898**	**1.0005**	**0.9938**	**0.9816**	**0.9846**	**0.9881**	**0.9815**	**0.9886**
Nível 1: Região								
Base	0.9263	0.9349	0.9340	0.9223	0.9263	0.9287	0.9225	0.9279
Bottom-up	0.9210	0.9249	**0.9176**	0.9150	0.9171	0.9180	0.9139	0.9182
Top-down (prev)	0.9383	0.9508	0.9457	0.9328	0.9392	0.9413	0.9371	0.9407
OLS	0.9299	0.9402	0.9360	0.9242	0.9301	0.9329	0.9275	0.9315
WLS (*struct*)	**0.9130**	**0.9217**	0.9177	**0.9092**	**0.9127**	**0.9143**	**0.9100**	**0.9141**
Nível 2: Loja								
Base	**0.8101**	**0.8118**	**0.8085**	**0.8090**	**0.8109**	**0.8112**	**0.8086**	**0.8100**
Bottom-up	**0.8101**	**0.8118**	**0.8085**	**0.8090**	**0.8109**	**0.8112**	**0.8086**	**0.8100**
Top-down (prev)	0.8306	0.8355	0.8325	0.8286	0.8319	0.8339	0.8284	0.8316
OLS	0.8311	0.8340	0.8321	0.8276	0.8313	0.8338	0.8280	0.8311
WLS (*struct*)	0.8195	0.8221	0.8200	0.8174	0.8200	0.8217	0.8171	0.8197

RMSSE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
Nível 0: Total								
Base	1.7521	1.7263	1.7094	1.7024	1.7084	1.7157	1.7050	1.7170
Bottom-up	1.7832	1.7510	1.7314	1.7253	1.7332	1.7333	1.7282	1.7408
Top-down (prev)	1.7521	1.7263	1.7094	1.7024	1.7084	1.7157	1.7050	1.7170
OLS	**1.7423**	**1.7089**	**1.6953**	**1.6879**	**1.6927**	**1.7014**	**1.6896**	**1.7026**
WLS (*struct*)	1.7466	1.7114	1.6970	1.6892	1.6945	1.7015	1.6918	1.7046
Nível 1: Região								
Base	1.5545	1.5221	1.5145	1.5096	1.5141	1.5203	1.5109	1.5209
Bottom-up	1.5723	1.5443	1.5293	1.5290	1.5350	1.5346	1.5315	1.5394
Top-down (prev)	1.5549	1.5348	1.5239	1.5201	1.5275	1.5289	1.5238	1.5306
OLS	1.5514	1.5279	1.5160	1.5131	1.5204	1.5218	1.5155	1.5237
WLS (*struct*)	**1.5466**	**1.5185**	**1.5070**	**1.5047**	**1.5099**	**1.5131**	**1.5072**	**1.5153**
Nível 2: Loja								
Base	1.2776	1.2669	1.2582	1.2605	1.2642	1.2638	1.2608	1.2646
Bottom-up	1.2776	1.2669	1.2582	1.2605	1.2642	1.2638	1.2608	1.2646
Top-down (prev)	1.2753	1.2682	1.2597	1.2620	1.2666	1.2673	1.2613	1.2658
OLS	1.2749	1.2645	1.2564	1.2578	1.2622	1.2640	1.2574	1.2625
WLS (*struct*)	**1.2703**	**1.2589**	**1.2509**	**1.2528**	**1.2564**	**1.2580**	**1.2525**	**1.2571**

Tabela 4.5 – MASE e RMSSE das previsões geradas pela DNN com regressores

DNN com regressores								
MASE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
Nível 0: Total								
Base	1.2115	1.1893	1.1752	1.1636	1.1617	1.1567	1.1527	1.1729
Bottom-up	**1.0954**	**1.0709**	**1.0555**	**1.0439**	**1.0418**	**1.0377**	**1.0337**	**1.0541**
Top-down (prev)	1.2115	1.1893	1.1752	1.1636	1.1617	1.1567	1.1527	1.1729
OLS	1.1962	1.1738	1.1596	1.1483	1.1467	1.1424	1.1390	1.1580
WLS (*struct*)	1.1580	1.1349	1.1203	1.1089	1.1072	1.1029	1.0994	1.1188
Nível 1: Região								
Base	1.0562	1.0382	1.0278	1.0202	1.0196	1.0167	1.0139	1.0275
Bottom-up	**0.9910**	**0.9714**	**0.9603**	**0.9524**	**0.9511**	**0.9480**	**0.9447**	**0.9598**
Top-down (prev)	1.0665	1.0484	1.0380	1.0298	1.0282	1.0241	1.0203	1.0365
OLS	1.0578	1.0397	1.0292	1.0212	1.0198	1.0157	1.0122	1.0279
WLS (*struct*)	1.0288	1.0102	0.9995	0.9915	0.9904	0.9873	0.9842	0.9988
Nível 2: Loja								
Base	**0.8559**	**0.8454**	**0.8389**	**0.8365**	**0.8365**	**0.8347**	**0.8324**	**0.8401**
Bottom-up	**0.8559**	**0.8454**	**0.8389**	**0.8365**	**0.8365**	**0.8347**	**0.8324**	**0.8401**
Top-down (prev)	0.8927	0.8827	0.8767	0.8740	0.8738	0.8718	0.8694	0.8773
OLS	0.8933	0.8834	0.8774	0.8748	0.8748	0.8727	0.8703	0.8781
WLS (*struct*)	0.8776	0.8675	0.8613	0.8588	0.8588	0.8570	0.8548	0.8623

RMSSE	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=1-7
Nível 0: Total								
Base	1.9013	1.8255	1.7965	1.7830	1.7824	1.7774	1.7739	1.8057
Bottom-up	**1.8506**	**1.7698**	**1.7369**	**1.7247**	**1.7246**	**1.7218**	**1.7194**	**1.7497**
Top-down (prev)	1.9013	1.8255	1.7965	1.7830	1.7824	1.7774	1.7739	1.8057
OLS	1.8896	1.8140	1.7851	1.7724	1.7727	1.7696	1.7672	1.7958
WLS (*struct*)	1.8705	1.7934	1.7633	1.7509	1.7513	1.7487	1.7465	1.7750
Nível 1: Região								
Base	1.6630	1.6004	1.5786	1.5713	1.5726	1.5702	1.5679	1.5891
Bottom-up	**1.6326**	**1.5651**	**1.5404**	**1.5327**	**1.5332**	**1.5304**	**1.5279**	**1.5517**
Top-down (prev)	1.6670	1.6028	1.5804	1.5715	1.5711	1.5659	1.5621	1.5887
OLS	1.6599	1.5969	1.5745	1.5661	1.5660	1.5611	1.5577	1.5832
WLS (*struct*)	1.6431	1.5779	1.5548	1.5469	1.5476	1.5447	1.5423	1.5653
Nível 2: Loja								
Base	**1.3141**	**1.2834**	**1.2698**	**1.2684**	**1.2693**	**1.2677**	**1.2651**	**1.2768**
Bottom-up	**1.3141**	**1.2834**	**1.2698**	**1.2684**	**1.2693**	**1.2677**	**1.2651**	**1.2768**
Top-down (prev)	1.3296	1.3005	1.2882	1.2862	1.2867	1.2842	1.2811	1.2938
OLS	1.3268	1.2981	1.2858	1.2840	1.2847	1.2823	1.2794	1.2916
WLS (*struct*)	1.3180	1.2884	1.2757	1.2741	1.2750	1.2732	1.2706	1.2821

5. Conclusões

A incerteza associada à procura implica que os retalhistas apoiem as suas decisões na previsão de vendas, de modo a garantir uma gestão eficiente dos seus *stocks* e a qualidade dos serviços que prestam aos clientes. A previsão de vendas geralmente baseia-se em dados históricos, mas o uso de outras variáveis pode ajudar a diminuir os erros de previsão. Desta forma, é importante que sejam incorporados no processo de previsão fatores que influenciam as vendas dos retalhistas, nomeadamente, a atividade promocional. Ainda assim, realça-se que a consideração de tais fatores aumenta a complexidade do processo de previsão, o que exige o uso de métodos de previsão mais sofisticados.

No caso do setor do retalho, são necessárias previsões para um vasto conjunto de séries de vendas que se encontram organizadas segundo uma determinada estrutura hierárquica. Neste sentido, a abordagem mais adequada consiste na previsão hierárquica que permite gerar previsões de vendas para diferentes níveis de agregação. Porém, não existe um consenso sobre o modelo de previsão que deve ser utilizado em cada nível hierárquico ou sobre a melhor estratégia de reconciliação das previsões.

Esta dissertação teve como objetivo a integração de informação adicional à própria série das vendas na estratégia de previsão hierárquica, de forma a aumentar a precisão das previsões no setor do retalho e contribuir para que, principalmente, grandes retalhistas aumentem a eficiência das suas decisões. Para o efeito, foi realizado um estudo com base em dados disponibilizados pelo Grupo Jerónimo Martins relativos a vendas e a preços de produtos em lojas Pingo Doce, observados no período entre 3 de janeiro de 2012 e 27 de abril de 2015. Para gerar as previsões base para os vários níveis hierárquicos, foram considerados os modelos ARIMA e uma DNN que utilizaram numa primeira fase de previsão, apenas dados de vendas observadas e, posteriormente, variáveis adicionais como o preço e eventos de calendário. A consideração da DNN para a previsão de vendas foi motivada por estudos recentes que indicam a eficácia dos métodos de *Deep Learning* em modelar relações complexas e não lineares e conferem vantagem a estas abordagens de previsão sobre os modelos de previsão tradicionais. Para reconciliar as previsões base, foram implementados os métodos *bottom-up* e *top-down* e as estratégias de reconciliação ótima com estimadores OLS e WLS (*struct*).

Os desempenhos dos modelos de previsão foram avaliados a partir das medidas de precisão MASE e RMSSE. Os resultados permitiram concluir que, para horizontes de

previsão superiores a um dia, a DNN permite obter previsões mais precisas do que os modelos ARIMA. Além disso, a integração da atividade promocional nos modelos ARIMA permite aumentar a precisão das previsões, ainda que o mesmo não suceda para o caso da DNN. A desigualdade na forma como os regressores são integrados no processo de previsão pode explicar este resultado. Neste sentido, a utilização de uma NN mais complexa poderia ter permitido alcançar melhores resultados com a integração dos regressores. Embora, NNs mais complexas, à partida, gerem melhores previsões e apresentem uma maior capacidade de generalização, acarretam custos de processamentos mais elevados e tempos de execução maiores. Estes são fatores a ter em consideração na escolha do método de previsão, visto que a grande escala de previsão das vendas no setor do retalho implica a utilização de métodos de previsão eficientes.

Investigação futura poderá incluir o desenvolvimento de modelos híbridos que beneficiem simultaneamente de técnicas estatísticas e da inteligência artificial. Além disso, seria relevante a consideração no processo de previsão de mais variáveis promocionais ou, até mesmo, de novas fontes de informação analisadas pela literatura recentemente como as redes sociais e o histórico de pesquisas *online*. Relativamente à estratégia de previsão hierárquica, poderá ser experimentada a agregação temporal, em alternativa às hierarquias *cross-sectional*.

6. Referências

Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering*, *143*, Article 106435. https://doi.org/10.1016/j.cie.2020.106435

Abolghasemi, M., Beh, E., Tarr, G., & Gerlach, R. (2020). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering*, *142*, Article 106380. https://doi.org/10.1016/j.cie.2020.106380

Abolghasemi, M., Hurley, J., Eshragh, A., & Fahimnia, B. (2020). Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. *International Journal of Production Economics*, *230*, Article 107892. https://doi.org/10.1016/j.ijpe.2020.107892

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, *7*(1), 136-144. https://doi.org/10.1016/j.asoc.2005.06.001

Ali, Ö. G., Sayın, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, *36*(10), 12340-12348. https://doi.org/10.1016/j.eswa.2009.04.052

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, *8*(3), 147-156. https://doi.org/10.1016/S0969-6989(00)00011-4

Aras, S., Deveci Kocakoç, İ., & Polat, C. (2017). Comparative study on retail sales forecasting between single and combination methods. *Journal of Business Economics and Management*, *18*(5), 803-832. https://doi.org/10.3846/16111699.2017.1367324

Arunraj, N., & Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *International Journal of Production Economics*, *170*, 321-335. https://doi.org/10.1016/j.ijpe.2015.09.039

Arunraj, N., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX model to forecast daily sales in food retail industry. *International Journal of Operations Research and Information Systems*, *7*(2), 1-21. https://doi.org/10.4018/IJORIS.2016040101

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, *25*(1), 146-166. https://doi.org/10.1016/j.ijforecast.2008.07.004

Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J., & Affan, M. (2020). Hierarchical forecasting. In P. Fuleky (Ed.), *Macroeconomic forecasting in the era of big data: Theory and practice* (pp. 689-719). Springer. https://doi.org/10.1007/978-3-030-31150-6_21

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60-74. https://doi.org/10.1016/j.ejor.2017.02.046

Badorf, F., & Hoberg, K. (2020). The impact of daily weather on retail sales: An empirical study in brick-and-mortar stores. *Journal of Retailing and Consumer Services*, *52*, Article 101921. https://doi.org/10.1016/j.jretconser.2019.101921

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In T. Gedeon, K. W. Wong, & M. Lee (Eds.), *Neural Information Processing* (pp. 462-474). Springer. https://doi.org/10.1007/978-3-030-36718-3_39

Beutel, A.-L., & Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, *140*(2), 637-645. https://doi.org/10.1016/j.ijpe.2011.04.017

Boone, T., Boylan, J. E., Fildes, R., Ganeshan, R., & Sanders, N. (2019). Perspectives on supply chain forecasting. *International Journal of Forecasting*, *35*(1), 121-127. https://doi.org/10.1016/j.ijforecast.2018.11.002

Boone, T., Ganeshan, R., Hicks, R. L., & Sanders, N. R. (2018). Can Google trends improve your sales forecast? *Production and Operations Management*, *27*(10), 1770-1774. https://doi.org/10.1111/poms.12839

Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, *35*(1), 170-180. https://doi.org/10.1016/j.ijforecast.2018.09.003

Box, G. E., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer.

Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics*, *13*(3), 237-252. https://doi.org/10.1080/07350015.1995.10524598

Chen, K. (2020). An online retail prediction model based on AGA-LSTM neural network. In *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)* (pp. 145-149). IEEE. https://doi.org/10.1109/MLBDBI51377.2020.00032

Chern, C.-C., Wei, C.-P., Shen, F.-Y., & Fan, Y.-N. (2015). A sales forecasting model for consumer products based on the influence of online word-of-mouth. *Information Systems and e-Business Management*, *13*(3), 445-473. https://doi.org/10.1007/s10257-014-0265-0

Chu, C.-W., & Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, *86*(3), 217-231. https://doi.org/10.1016/S0925-5273(03)00068-9

Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management*, *27*(10), 1749-1769. https://doi.org/10.1111/poms.12707

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, *29*(3), 510-522. https://doi.org/10.1016/j.ijforecast.2012.09.002

Divakar, S., Ratchford, B. T., & Shankar, V. (2005). CHAN4CAST: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Marketing Science*, *24*(3), 334-350. https://doi.org/10.1287/mksc.1050.0135

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1), Article 100058. https://doi.org/10.1016/j.jjimei.2022.100058

Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive analytics for demand forecasting – A comparison of SARIMA and LSTM in retail SCM. *Procedia Computer Science*, *200*, 993-1003. https://doi.org/10.1016/j.procs.2022.01.298

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3-23. https://doi.org/10.1016/j.ijforecast.2008.11.010

Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2019.06.004

Fisher, M., & Raman, A. (2018). Using data and big data in retailing. *Production and Operations Management*, *27*(9), 1665-1669. https://doi.org/10.1111/poms.12846

Fisher, M., & Vaidyanathan, R. (2014). A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science*, *60*(10), 2401-2415. https://doi.org/10.1287/mnsc.2014.1904

Franses, P. H., & Legerstee, R. (2010). Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? *Journal of Forecasting*, *29*(3), 331-340. https://doi.org/10.1002/for.1129

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, *9*(3), 233-254. https://doi.org/10.1002/for.3980090304

Guo, Z. X., Wong, W. K., & Li, M. (2013). A multivariate intelligent decision-making model for retail sales forecasting. *Decision Support Systems*, *55*(1), 247-255. https://doi.org/10.1016/j.dss.2013.01.026

He, Q.-Q., Wu, C., & Si, Y.-W. (2022). LSTM with particle swam optimization for sales forecasting. *Electronic Commerce Research and Applications*, *51*, Article 101118. https://doi.org/10.1016/j.elerap.2022.101118

Hirche, M., Haensch, J., & Lockshin, L. (2021). Comparing the day temperature and holiday effects on retail sales of alcoholic beverages – A time-series analysis. *International Journal of Wine Business Research*, *33*(3), 432-455. https://doi.org/10.1108/IJWBR-07-2020-0035

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hosoda, T., Naim, M. M., Disney, S. M., & Potter, A. (2008). Is there a benefit to sharing market sales information? Linking theory and practice. *Computers & Industrial Engineering*, *54*(2), 315-326. https://doi.org/10.1016/j.cie.2007.07.014

Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, *237*(2), 738-748. https://doi.org/10.1016/j.ejor.2014.02.022

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, *36*(4), 1420-1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, *55*(9), 2579-2589. https://doi.org/10.1016/j.csda.2011.03.006

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., Lee, A. J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, *97*, 16-32. https://doi.org/10.1016/j.csda.2015.11.007

Jain, A., Rudi, N., & Wang, T. (2015). Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, *63*(1), 134-150. https://doi.org/10.1287/opre.2014.1326

Kahn, K. B. (2014). Solving the problems of new product forecasting. *Business Horizons*, *57*(5), 607-615. https://doi.org/10.1016/j.bushor.2014.05.003

Kaunchi, P., Jadhav, T., Dandawate, Y., & Marathe, P. (2021). Future sales prediction for indian products using convolutional neural network-long short term memory. In *2021 2nd Global Conference for Advancement in Technology (GCAT)* (pp. 1-5). IEEE. https://doi.org/10.1109/GCAT52182.2021.9587668

Kim, S., Kim, H., & Lu, J.-C. (2019). A practical approach to measuring the impacts of stockouts on demand. *Journal of Business & Industrial Marketing*, *34*(4), 891-901. https://doi.org/10.1108/JBIM-04-2018-0126

Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, *30*(2), 291-302. https://doi.org/10.1016/j.ijforecast.2013.09.006

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1-3), 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lau, R. Y. K., Zhang, W., & Xu, W. (2018). Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Production and Operations Management*, *27*(10), 1775-1794. https://doi.org/10.1111/poms.12737

LeCun, Y., Boser, B., Denker, J. S., Howard, R. E., Habbard, W., Jackel, L. D., & Henderson, D. (1990). Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems 2* (pp. 396–404). Morgan Kaufmann Publishers Inc.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science*, *43*(4), 546-558. https://doi.org/10.1287/mnsc.43.4.546

Lee, H. L., So, K. C., & Tang, C. S. (2000). The value of information sharing in a two-Level supply chain. *Management Science*, *46*(5), 626-643. https://doi.org/10.1287/mnsc.46.5.626.12047

Lee, H. L., & Whang, S. (2000). Information sharing in a supply chain. *International Journal of Manufacturing Technology and Management*, *1*(1), 79-93. https://doi.org/10.1504/ijmtm.2000.001329

Liu, X., & Ichise, R. (2017). Food sales prediction with meteorological data — A case study of a japanese chain supermarket. In Y. Tan, H. Takagi, & Y. Shi (Eds.), *Data Mining and Big Data* (pp. 93-104). Springer. https://doi.org/10.1007/978-3-319-61845-6_10

Liu, Y., Lan, K., Huang, F., Cao, X., Feng, B., & Zhu, B. (2021). An aggregate store sales forecasting framework based on ConvLSTM. In *2021 The 5th International Conference on Compute and Data Analysis* (pp. 67–72). Association for Computing Machinery. https://doi.org/10.1145/3456529.3456540

Ma, S., & Fildes, R. (2021). Retail sales forecasting with meta-learning. *European Journal of Operational Research*, *288*(1), 111-128. https://doi.org/10.1016/j.ejor.2020.05.038

Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*, *249*(1), 245-257. https://doi.org/10.1016/j.ejor.2015.08.029

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*, Article 1. https://doi.org/10.1186/s40537-014-0007-7

Nelson, M., Hill, T., Remus, W., & O'Connor, M. (1999). Time series forecasting using neural networks: Should the data be deseasonalized first? *Journal of Forecasting*, *18*(5), 359-367. https://doi.org/10.1002/(SICI)1099-131X(199909)18:5<359::AID-FOR746>3.0.CO;2-P

Nithin, S. S. J., Rajasekar, T., Jayanthy, S., Karthik, K., & Rithick, R. R. (2022). Retail demand forecasting using CNN-LSTM model. In *2022 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1751-1756). IEEE. https://doi.org/10.1109/ICEARS53579.2022.9752283

Oliveira, J. M., & Ramos, P. (2019). Assessing the performance of hierarchical forecasting methods on the retail sector. *Entropy*, *21*(4), 436. https://doi.org/10.3390/e21040436

Pan, H., & Zhou, H. (2020). Study on convolutional neural network and its application in data mining and sales forecasting for e-commerce. *Electronic Commerce Research*, *20*(2), 297-320. https://doi.org/10.1007/s10660-020-09409-0

Pennings, C. L. P., & van Dalen, J. (2017). Integrated hierarchical forecasting. *European Journal of Operational Research*, *263*(2), 412-418. https://doi.org/10.1016/j.ejor.2017.04.047

Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, *58*(16), 4964-4979. https://doi.org/10.1080/00207543.2020.1735666

Ramanathan, U., & Muyldermans, L. (2010). Identifying demand factors for promotional planning and forecasting: A case of a soft drink company in the UK. *International Journal of Production Economics*, *128*(2), 538-545. https://doi.org/10.1016/j.ijpe.2010.07.007

Ramanathan, U., & Muyldermans, L. (2011). Identifying the underlying structure of demand during promotions: A structural equation modelling approach. *Expert Systems with Applications*, *38*(5), 5544-5552. https://doi.org/10.1016/j.eswa.2010.10.082

Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). https://doi.org/10.2202/1544-6115.1175

Seaman, B. (2018). Considerations of a retail forecasting practitioner. *International Journal of Forecasting*, *34*(4), 822-829. https://doi.org/10.1016/j.ijforecast.2018.03.001

Spiliotis, E., Abolghasemi, M., Hyndman, R. J., Petropoulos, F., & Assimakopoulos, V. (2021). Hierarchical forecast reconciliation with machine learning. *Applied Soft Computing*, *112*, Article 107756. https://doi.org/10.1016/j.asoc.2021.107756

Steinker, S., Hoberg, K., & Thonemann, U. W. (2017). The value of weather information for e-commerce operations. *Production and Operations Management*, *26*(10), 1854-1874. https://doi.org/10.1111/poms.12721

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1-26. https://doi.org/10.1016/j.ejor.2015.11.010

Trapero, J. R., Kourentzes, N., & Fildes, R. (2012). Impact of information exchange on supplier forecasting performance. *Omega*, *40*(6), 738-747. https://doi.org/10.1016/j.omega.2011.08.009

Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, *66*(2), 299-307. https://doi.org/10.1057/jors.2013.174

Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, *29*(2), 234-243. https://doi.org/10.1016/j.ijforecast.2012.10.002

Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications*, *201*, Article 116993. https://doi.org/10.1016/j.eswa.2022.116993

Verstraete, G., Aghezzaf, E.-H., & Desmet, B. (2019). A data-driven framework for predicting weather impact on high-volume low-margin retail products. *Journal of Retailing and Consumer Services*, *48*, 169-177. https://doi.org/10.1016/j.jretconser.2019.02.019

Villegas, M. A., & Pedregal, D. J. (2018). Supply chain decision support systems based on a novel hierarchical forecasting approach. *Decision Support Systems*, *114*, 29-36. https://doi.org/10.1016/j.dss.2018.08.003

Vulcano, G., van Ryzin, G., & Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, *60*(2), 313-334. https://doi.org/10.1287/opre.1110.1012

Wang, J., Liu, G. Q., & Liu, L. (2019). A selection of advanced technologies for demand forecasting in the retail industry. In *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)* (pp. 317-320). IEEE. https://doi.org/10.1109/ICBDA.2019.8713196

Weng, T., Liu, W., & Xiao, J. (2020). Supply chain sales forecasting based on lightGBM and LSTM combination model. *Industrial Management & Data Systems*, *120*(2), 265-279. https://doi.org/10.1108/IMDS-03-2019-0170

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, *114*(526), 804-819. https://doi.org/10.1080/01621459.2018.1448825

Yu, Z., Yan, H., & Edwin Cheng, T. C. (2001). Benefits of information sharing with supply chain partnerships. *Industrial Management & Data Systems*, *101*(3), 114-121. https://doi.org/10.1108/02635570110386625

Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501-514. https://doi.org/10.1016/j.ejor.2003.08.037

Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, *93-94*, 479-491. https://doi.org/10.1016/j.ijpe.2004.06.044