

Predicting product return volume using machine learning methods	Прогнозирование объема возврата товара с использованием методов машинного обучения
Hailong Cui ^{a, *} , Sampath Rajagopalan ^a , Amy R. Ward ^{b a} <i>Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA</i> ^b <i>Booth School of Business, University of Chicago, Chicago, IL 60637, USA</i>	
a b s t r a c t	Аннотация
In 2015, U.S. consumers returned goods worth \$261 billion and the return rates for online sales sometimes exceeded 30%. Manufacturers and retailers have an interest in predicting return volume to address operational challenges in managing product returns. In this paper, we develop and test data-driven models for predicting return volume at the retailer, product type and period levels using a rich data set comprised of detailed operations on each product, and retailer information. The goal is to achieve a good prediction accuracy out of sample. We consider main effects and detailed interaction effects models using various machine learning methods. We find that Least Absolute Shrinkage and Selection Operator (LASSO) yields a predictive model achieving the best prediction accuracy for future return volume due to its ability to select informative interaction terms out of more than one thousand possible combinations. The LASSO model also turns in consistent performance based on several robustness tests and is easy to implement in practice. Our work provides a general predictive model framework for manufacturers to track product returns.	В 2015 году потребители в США возвратили товаров на сумму 261 млрд долларов, а показатели возврата при онлайн-продажах иногда превышали 30%. Производители и розничные продавцы заинтересованы в прогнозировании объемов возврата для решения операционных проблем управления возвратами продуктов. В этой статье мы разрабатываем и тестируем управляемые данными модели для прогнозирования объема возврата на уровне продавца, типа продукта и периода, используя богатый набор данных, состоящий из подробных операций над каждым продуктом и информации о продавце. Цель состоит в том, чтобы добиться хорошей точности прогноза из выборки. Мы рассматриваем основные эффекты и детализированные модели эффектов взаимодействия, используя различные методы машинного обучения. Мы считаем, что оператор наименьшей абсолютной усадки и выбора (LASSO) дает прогностическую модель, обеспечивающую наилучшую точность прогнозирования будущего объема возврата благодаря его способности выбирать информативные термины взаимодействия из более чем тысячи возможных комбинаций. Модель LASSO также обеспечивает стабильную производительность, основанную на нескольких тестах на надежность, и ее легко реализовать на практике. Наша работа предоставляет производителям общую модель прогнозирования для отслеживания возврата продукции.
1.Introduction	1. Введение
In the retail industry consumer returns create a significant and costly issue for manufacturers and retailers. In 2015, consumers returned goods worth \$261 billion out of \$3.3 trillion sold in the U.S. (Economist, 2014) and returned goods worth \$642.6 billion out of \$14.53 trillion sold worldwide. Furthermore, more than half of all returns may not be resold for full price, which results in substantial financial losses (Cheng, 2015). The rise in online sales (Insider, 2017) has led to increasing return rates, sometimes exceeding	В розничной торговле потребительские доходы создают значительную и дорогостоящую проблему для производителей и розничных продавцов. В 2015 году потребители вернули товары на сумму 261 млрд долларов из 3,3 трлн долларов, проданных в США (Economist, 2014), и вернули товары на сумму 642,6 млрд долларов из 14,53 трлн долларов, проданных по всему миру. Кроме того, более половины всех доходов не могут быть перепроданы по полной цене, что приводит к существенным

<p>30% (Rudolph, 2016). The issue is particularly problematic for online retailers that offer extensive product variety and customization options, making it more difficult to resell the returned items.</p>	<p>финансовым потерям (Cheng, 2015). Рост онлайн-продаж (Insider, 2017) привел к увеличению доходности, иногда превышающей 30% (Rudolph, 2016). Эта проблема особенно проблематична для интернет-магазинов, которые предлагают широкий ассортимент товаров и возможности индивидуальной настройки, что затрудняет перепродажу возвращенных товаров.</p>
<p>Firms have an interest in developing a model for predicting return volume and understanding the underlying factors associated with it for several reasons. These can be broadly grouped into two categories: (i) operational issues and (ii) financial issues. Returns can cause significant operational and logistics challenges because firms have to devote resources such as staff and space to process returns, identify if an item should be resold or disposed, etc. Moreover, if returned items have to be repaired, then this may negatively impact the workflow in the production process depending on the product type and workload levels. Financially, return volume helps provide an estimate of the cost or loss due to returns. So, from a financial and operational perspective, understanding return volume is beneficial not just to the firm in our study, but also any firm facing a large return volume in the e-commerce era. Finally, anticipating return volume of a product type at a retailer in each period may be valuable in taking actions to reduce returns. For example, Jet.com offers a discount to customers who opt out of free returns (https://jet.com/help-center/faq). Rather than offering such discounts for all product types, a manufacturer may be able to optimize its discounting strategy by targeting such discounts to specific product types offered at certain retailers that are likely to have higher returns. Such discounts may also be restricted to certain periods, for instance during the Christmas season when returns are higher.</p>	<p>Фирмы заинтересованы в разработке модели для прогнозирования объема прибыли и понимания основных факторов, связанных с ним, по нескольким причинам. Их можно сгруппировать в две категории: (i) операционные вопросы и (ii) финансовые проблемы. Возврат может вызвать значительные эксплуатационные и логистические проблемы, потому что фирмы должны выделять ресурсы, такие как персонал и пространство, для обработки возвратов, определения необходимости перепродажи или утилизации предмета и т. Д. Кроме того, если подлежащие ремонту предметы должны быть отремонтированы, это может оказать негативное влияние рабочий процесс в производственном процессе в зависимости от типа продукта и уровней нагрузки. В финансовом отношении объем возврата помогает оценить стоимость или убыток в результате возврата. Таким образом, с финансовой и операционной точек зрения понимание объема возврата выгодно не только для фирмы в нашем исследовании, но и для любой, кто сталкивается с большим объемом возврата в эпоху электронной коммерции. Наконец, предвидение объема возврата продукта определенного типа в розничном магазине в каждый период может быть полезным при принятии мер по снижению возврата. Например, Jet.com предлагает скидки для клиентов, которые отказываются от бесплатных возвратов (https://jet.com/help-center/faq). Вместо того, чтобы предлагать такие скидки для всех типов продуктов, производитель может оптимизировать свою стратегию дисконтирования, ориентируя такие скидки на конкретные типы продуктов, предлагаемые в определенных розничных магазинах, которые могут иметь более высокую доходность. Такие скидки также могут быть ограничены определенными периодами, например, во время рождественского сезона, когда доходы выше.</p>
<p>The primary objective of this paper is to build a good data driven model to predict return volume in the future. Our study was done in collaboration with a leading manufacturer of auto motive accessories in the U.S., referred to as Company <i>A</i> hereafter, whose real name is</p>	<p>Основная цель этой статьи создать хорошую модель, основанную на данных, для прогнозирования объема возврата в будущем. Наше исследование было проведено в сотрудничестве с ведущим производителем аксессуаров для автомобилей в США, далее</p>

<p>withheld due to confidentiality reasons. Company <i>A</i> primarily sells three types of products (seat cover, dash cover, car cover) through various online retailers. Customers can choose a car accessory matched to a specific car model year, color, fabric type and also customize the product with logos, specialized prints, etc. Customers place orders online via the retailer's website and the firm makes each item to order and ships it in around a week. Customer return policies, whether strict or liberal, are dictated by the retailer. However, all returns are handled by Company <i>A</i>, not the retailers, who serve only as a sales channel. Due to the extensive variety offered, the odds are low that a product returned by one customer will be demanded by another one within, say, a few months. However, the firm checks each returned item to assess whether to put it back in stock (if it is not defective and is likely to be resold) or to dispose it. Many items are disposed or sold at a steep discount and cause a significant financial loss to the company, because extracting value from returned product is difficult in the focal firm of our study. This is unlike for other product categories such as appliances where a substantial part of the value may be recovered from a resale.</p>	<p>именуемым Компанией А, настоящее имя которого не разглашается по соображениям конфиденциальности. Компания А в основном продает три вида продукции (чехлы на сиденья, панели приборов, автомобильные чехлы) через различные интернет-магазины. Клиенты могут выбрать автомобильный аксессуар, соответствующий конкретному году выпуска автомобиля, цвету, типу ткани, а также настроить продукт с помощью логотипов, специализированных принтов и т. Д. Клиенты размещают заказы онлайн через веб-сайт продавца, и фирма делает каждый товар на заказ и отправляет его. примерно через неделю. Политика возврата клиентов, будь то строгая или либеральная, диктуется продавцом. Тем не менее, все возвраты обрабатываются компанией А, а не продавцами, которые служат только в качестве канала продаж. Из-за большого разнообразия предлагаемых товаров маловероятно, что продукт, возвращаемый одним покупателем, будет востребован другим в течение, скажем, нескольких месяцев. Однако фирма проверяет каждый возвращенный товар, чтобы определить, следует ли вернуть его на склад (если он не имеет дефектов и, вероятно, будет перепродан) или утилизировать его. Многие товары утилизируются или продаются с большой скидкой, что приводит к значительным финансовым потерям для компании, поскольку извлечение ценности из возвращенного продукта является сложной задачей в нашем исследовании. Это непохоже на другие категории продуктов, такие как бытовая техника, где значительная часть стоимости может быть восстановлена после перепродажи.</p>
<p>Our predictive models are built using a large, detailed data set from Company A on every item that was sold and/or returned over 39 months. For each item, the firm provided us with data on order date, retailer, product type, production process details (including dates when each process step was completed and who worked on it), ship date, and return date. We also obtained aggregate information on production levels, workers, inspection policies, and return policies of retailers.</p>	<p>Наши прогностические модели построены с использованием большого подробного набора данных компании А по каждому товару, который был продан и / или возвращен в течение 39 месяцев. Для каждой позиции фирма предоставила нам данные о дате заказа, розничном продавце, типе продукта, деталях производственного процесса (включая даты, когда каждый шаг процесса был завершен и кто работал над ним), дату отгрузки и дату возврата. Мы также получили сводную информацию об уровнях производства, работниках, политике проверок и политике возврата розничных продавцов.</p>
<p>We first develop a baseline main effects model for predicting return volume using four factors deemed to be important based on our initial understanding of returns at Company A: sales volume, time, retailer and product type. Return volume is generally proportional to sales</p>	<p>Сначала мы разработали базовую модель основных эффектов для прогнозирования объема прибыли, используя четыре фактора, которые считаются важными на основе нашего первоначального понимания доходности в компании А: объем продаж, время, розничный продавец и</p>

<p>volume although the proportion (return rate) may vary by product type, retailer and over time. The retailer variable captures factors such as return policies, type of consumers who visit a retailer, which could influence return volume. For example, more impulsive purchases may generate higher return volume and such impulsive consumers may be more likely to visit certain retailers. Return volume may also vary by product type because certain product types may have greater fit issues (e.g., seat cover) or certain product types may be prone to defects in the manufacturing process. Finally, return volume may vary over time; for example, consumers may be more impulsive during holiday purchases. We then explore how additional variables such as manufacturing workload levels, process quality checks, production process complexity, and product personalization (e.g., logos) improve prediction performance—this leads to a full main effects model with a much larger set of independent variables.</p>	<p>тип продукта. Объем возврата, как правило, пропорционален объему продаж, хотя пропорция (норма возврата) может варьироваться в зависимости от типа продукта, продавца и времени. Переменная розничного продавца учитывает такие факторы, как политика возврата, тип потребителей, которые посещают розничного продавца, что может повлиять на объем возврата. Например, более импульсивные покупки могут генерировать более высокий объем возврата, и такие импульсивные потребители могут с большей вероятностью посетить определенные розничные продавцы. Объем возврата также может варьироваться в зависимости от типа продукта, поскольку некоторые типы продуктов могут иметь более серьезные проблемы (например, чехол для сиденья) или некоторые типы продуктов могут быть подвержены дефектам в процессе производства. Наконец, объем возврата может меняться со временем; например, потребители могут быть более импульсивными во время праздничных покупок. Затем мы исследуем, как дополнительные переменные, такие как уровни рабочей нагрузки производства, проверки качества процесса, сложность производственного процесса и персонализация продукта (например, логотипы), улучшают производительность прогнозирования это приводит к полной модели основных эффектов с гораздо большим набором независимых переменных.</p>
<p>In addition to the main effects models discussed above, we also investigate whether adding second and third order interaction effects improves prediction. For instance, returns may be increasing over time at a particular retailer. Alternatively, the presence of personalized logos may impact returns for some products but not others or inspection may be valuable in reducing defects/returns for some products but not others. The incorporation of aforementioned interaction effects, however, results in a large number of predictor variables and warrants the need for a robust variable/model selection methodology. The traditional methods such as best subset selection, forward/backward selection (see Hocking, 1976 for a review) are not applicable in our study, because we are in a high dimensional setting due to having more predictors than observations as will be clear later.</p>	<p>В дополнение к моделям основных эффектов, рассмотренным выше, мы также исследуем, улучшает ли предсказание добавление эффектов взаимодействия второго и третьего порядка. Например, доход может увеличиваться с течением времени у конкретного продавца. В качестве альтернативы, наличие персонализированных логотипов может повлиять на возврат для некоторых продуктов, но не для других, или проверка может быть полезна для уменьшения дефектов / возврата для некоторых продуктов, но не для других. Однако включение вышеупомянутых эффектов взаимодействия приводит к большому количеству предикторных переменных и оправдывает необходимость в надежной методологии выбора переменных / моделей. Традиционные методы, такие как выбор наилучшего подмножества, выбор вперед / назад (см. Hocking, 1976 для обзора), не применимы в нашем исследовании, потому что мы находимся в условиях высокой размерности из-за наличия большего количества предикторов, чем наблюдений, что будет ясно позже.</p>

<p>To address this issue, we use four high-dimensional machine learning methods: Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani, 1996), LARS-OLS hybrid (Efron, Hastie, Johnstone, & Tibshirani, 2004), Smoothly Clipped Absolute Deviation (SCAD, Fan & Li, 2001), and Elastic Net (Zou & Hastie, 2005) that can yield sparse models. In addition, we explore two notable tree-based machine learning methods, Random Forest (Breiman, 2001) and Gradient Boosting (Friedman, 2001) to capture possible complex non-linear structure in the data to improve prediction accuracy.</p>	<p>Для решения этой проблемы мы используем четыре многоплановых метода машинного обучения: оператор наименьшего абсолютного сжатия и усадки (LASSO, Tibshirani, 1996), гибрид LARS-OLS (Efron, Hastie, Johnstone & Tibshirani, 2004), плавное ограничение абсолютного отклонения (SCAD, Fan & Li, 2001) и Elastic Net (Zou & Hastie, 2005), которые могут дать разреженные модели. Кроме того, мы исследуем два известных метода машинного обучения на основе дерева: «Случайный лес» (Breiman, 2001) и «Повышение градиента» (Friedman, 2001), чтобы зафиксировать возможную сложную нелинейную структуру в данных для повышения точности прогнозирования.</p>
<p>Our contributions can be summarized as follows. We use various machine learning methods to build and test predictive models for return volume using a real data set that is comprehensive and includes retailer, product type, and process related variables. We show that in our setting retailer effects are stronger than product effects in predicting returns, and the baseline main effects model utilizing sales, time, product and retailer effects achieves a fair prediction performance. We consider higher order interaction effects (e.g., product type and retailer) and apply various convex and concave regularization methods for variable/model selection to derive a sparse predictive model that strongly improves prediction accuracy. In our study, we find that the optimal Elastic Net model coincides with the LASSO model, which achieves smaller prediction errors in the test data compared to all the alternative methods. It shows robust performance with similar prediction accuracy both in the training and test data.</p>	<p>Наш вклад можно резюмировать следующим образом. Мы используем различные методы машинного обучения для построения и тестирования прогностических моделей для возвращаемого объема с использованием реального набора данных, который является всеобъемлющим и включает в себя данные о розничном продавце, типе продукта и переменных, связанных с процессом. Мы показываем, что в наших условиях эффекты розничного предприятия сильнее, чем эффекты продукта при прогнозировании доходности, а базовая модель основных эффектов, использующая эффекты продаж, времени, продукта и розничного продавца, обеспечивает справедливую эффективность прогнозирования. Мы рассматриваем эффекты взаимодействия более высокого порядка (например, тип продукта и розничный продавец) и применяем различные выпуклые и вогнутые методы регуляризации для выбора переменной / модели, чтобы получить разреженную модель прогнозирования, которая значительно повышает точность прогнозирования. В нашем исследовании мы обнаружили, что оптимальная модель эластичной сети совпадает с моделью LASSO, которая обеспечивает меньшие ошибки прогнозирования в тестовых данных по сравнению со всеми альтернативными методами. Он показывает надежную производительность с одинаковой точностью прогнозирования как в данных обучения, так и в тестовых данных.</p>
<p>The remainder of the paper is organized as follows. In Section 2 we review the relevant literature. Section 3 discusses the empirical setting and the predictor variables derived from the data. In Section 4 we sequentially build main effects models. Section 5 explores higher-order interaction models with highdimensional variable/model selection methods, and tree-based statistical</p>	<p>Оставшаяся часть статьи организована следующим образом. В разделе 2 мы рассматриваем соответствующую литературу. В разделе 3 обсуждаются эмпирические параметры и предикторные переменные, полученные из данных. В разделе 4 мы последовательно строим модели основных эффектов. В разделе 5 рассматриваются модели взаимодействия более высокого порядка с методами выбора</p>

<p>machine learning methods. We provide results from robustness checks in Section 6, and discuss key take-aways from our study and future research direction in Section 7. Additional details can be found in the Online Appendix.</p>	<p>многомерных переменных / моделей и статистическими методами машинного обучения на основе дерева. Мы предоставляем результаты проверок надежности в Разделе 6 и обсуждаем основные выводы из нашего исследования и направления будущих исследований в Разделе 7. Дополнительные подробности можно найти в Онлайн-приложении.</p>
<p>2.Literature review</p>	<p>2. Литературный обзор</p>
<p>Our study is related to several streams of literature, one of them being the marketing literature on consumer product returns. Hess and Mayhew (1997) are among the earliest to empirically study product returns, and provide a model to predict the timing of returns for an apparel direct marketer. Janakiramana, Syrdalb, and Freling (2016) show that return policy leniency overall increases purchases more than returns. Anderson, Karsten, and Duncan (2009) identify a considerable variation in the value of returns across customers and product categories at a mail-order catalog company. Petersen and Kumar (2009) determine the firm-customer exchange process factors that help explain product return behavior and the consequences of product returns on future customer and firm behavior. Urbanke, Kranz, and Kolbe (2015) propose a returns prediction system for an online retailer to explore the impact of different levers on the likelihood of returns, for example, by artificially increasing delivery time to deter a consumer from purchase. The above-mentioned research is mostly concerned with factors impacting consumer's decision to keep or return a product, and studies returns from the perspective of retailers. In contrast, our paper develops data-driven predictive models for a manufacturer, which sells online an extensive range of product variants and is primarily concerned with the cost and logistics of handling returns, and is interested in predicting consumer returns due to various reasons (e.g. defect in product, consumer's change of mind)</p>	<p>Наше исследование связано с несколькими потоками литературы, одним из которых является маркетинговая литература по возврату потребительских товаров. Хесс и Мэйхью (1997) являются одними из первых, кто эмпирически изучает возврат товаров, и предоставляют модель для прогнозирования сроков возврата товаров для прямого продавца одежды. Janakiramana, Syrdalb и Freling (2016) показывают, что смягчение политики возврата в целом увеличивает покупки больше, чем доходы. Андерсон, Карстен и Дункан (2009) идентифицируют существенные различия в стоимости возвратов по клиентам и категориям продуктов в компании, производящей каталоги по почте. Petersen и Kumar (2009) определяют факторы процесса обмена фирма-клиент, которые помогают объяснить поведение возврата продукта и последствия возврата продукта для будущего клиента и поведения фирмы. Urbanke, Kranz и Kolbe (2015) предлагают систему прогнозирования прибыли для интернет-магазина для изучения влияния различных рычагов на вероятность возврата, например, путем искусственного увеличения времени доставки, чтобы удержать потребителя от покупки. Вышеупомянутое исследование в основном касается факторов, влияющих на решение потребителя оставить или вернуть товар, и исследования возвращаются с точки зрения розничных продавцов. В отличие от этого, в нашей статье разрабатываются прогнозирующие модели, основанные на данных, для производителя, который продает онлайн широкий спектр вариантов продукта и в первую очередь занимается стоимостью и логистикой обработки доходов, а также интересуется прогнозированием потребительских доходов по различным причинам (например, дефект товара, изменение мнения потребителя)</p>
<p>Another relevant stream is the operations literature on returns, which is primarily focused on analytical studies of product returns, for example, returned component reuse (DeCroix G, Song, & Zipkin, 2009), restocking fees (Shulman, Coughlan, & Savaskan,</p>	<p>Другим важным потоком является литература по операциям с возвратами, которая в основном сосредоточена на аналитических исследованиях возвратов продуктов, например, повторное использование возвращенных компонентов (DeCroix G, Song & Zipkin,</p>

<p>2011), return policies (Su, 2009, Altug, 2016, Shang, Pekgü7n, Ferguson, & Galbreth, 2017a, Ülkü & Gürler, 2017), remanufacturing (Calmon & Graves, 2017; Cerag, Ferguson, & Toktay, 2016), optimal retail assortment under consumer returns (Alptekinog̃lu & Grasas, 2017), return strategy and pricing in a dual-channel supply chain (Li, Li, Sethi, & Guan, 2017). In contrast, our work is empirical in nature and focused on predicting returns. There are some works related to predicting returns for the purpose of remanufacturing. For example, Toktay (2003) reviews a few forecasting methods (e.g., using past sales and returns), and Tsiliyannis (2018) presents a stochastic method for real-time forecasting of product returns in remanufacturing. Our study focuses on data-driven prediction models for returns and is not concerned with remanufacturing, because the products in our study are made of fabrics that are not reused to make new products.</p>	<p>2009), сборы за пополнение запасов (Shulman, Coughlan, & Savaskan, 2011).), политика возврата (Su, 2009, Altug, 2016, Shang, Pekgü7n, Ferguson & Galbreth, 2017a, Ülkü & Gürler, 2017), ремоделирование (Calmon & Graves, 2017; Cerag, Ferguson & Toktay, 2016), оптимальный розничный ассортимент по доходности потребителей (Alptekinog̃lu & Grasas, 2017), стратегия возврата и ценообразование в двухканальной цепочке поставок (Li, Li, Sethi & Guan, 2017). Напротив, наша работа носит эмпирический характер и ориентирована на прогнозирование прибыли. Есть некоторые работы, связанные с прогнозированием прибыли с целью восстановления производства. Например, Toktay (2003) рассматривает несколько методов прогнозирования (например, с использованием прошлых продаж и прибыли), а Tsiliyannis (2018) представляет стохастический метод для прогнозирования возврата продукта в реальном времени при восстановлении. Наше исследование сфокусировано на моделях прогнозирования доходности, основанных на данных, и не касается восстановления, потому что продукты, которые мы изучаем, сделаны из тканей, которые не используются повторно для создания новых продуктов.</p>
<p>Some recent works in operations also conduct empirical studies on returns from the perspective of retailers and consumers. For example, Shang, Ghosh, and Galbreth (2017b) analyze both the return policy drivers from the retailer's perspective and the return policy value from the consumer's perspective, and show that the value of a full refund policy to consumers may not be as large as one might expect. Akturk, Ketzenberg, and Heim (2018) study omnichannel retailing in the context of a national jewelry retailer and suggest that introducing ship-to-store increased cross-channel customer returns of online purchases to physical stores. In contrast to these works, we study product returns from the perspective of a manufacturer.</p>	<p>Некоторые недавние работы также проводят эмпирические исследования доходности с точки зрения розничной торговли и потребителей. Например, Шан, Гош и Галбрет (2017b) анализируют как драйверы политики возврата с точки зрения розничного продавца, так и значение политики возврата с точки зрения потребителя и показывают, что значение политики полного возврата для потребителей может быть не таким большим, как можно ожидать. Akturk, Ketzenberg и Heim (2018) изучают розничную торговлю через многоканальную сеть в контексте национального ритейлера ювелирных изделий и предполагают, что введение доставки между магазинами увеличило отдачу от покупок по нескольким каналам онлайн-покупок в физических магазинах. В отличие от этих работ, мы изучаем возврат продукта с точки зрения производителя.</p>
<p>Our paper is also related to the literature on predictive analytics, which are applied to a variety of settings. In the operations research literature, however, there is still a relatively low volume of analytics-orientated studies (Mortenson, Doherty, & Robinson, 2015). Here are two examples: Oztekin, Al-Ebbini, Sevkli, and</p>	<p>Наша статья также связана с литературой по прогнозной аналитике, которая применяется в различных условиях. Однако в литературе по операционным исследованиям объем аналитических исследований все еще остается относительно небольшим (Mortenson, Doherty, & Robinson, 2015). Вот два примера: Озтекин, Аль-Эббини, Севкли и Делен (2018)</p>

<p>Delen (2018) develop a hybrid methodology to predict and explain the quality of life for patients undergoing a lung transplant, and Sevim, Oztekin, Bali, Gumus, and Guresen (2014) develop an early warning system to predict currency crises using artificial neural networks, decision trees, and logistic regression.</p>	<p>разработали гибридную методологию для прогнозирования и объяснения качества жизни пациентов, перенесших трансплантацию легкого, и Севим, Озтекин, Бали, Гумус и Гюресен (2014).) разработать систему раннего предупреждения для прогнозирования валютных кризисов с использованием искусственных нейронных сетей, деревьев решений и логистической регрессии.</p>
<p>We utilize high-dimensional machine learning methods, which have become increasingly popular and widely used in areas such as genomics, neuroscience, social media analysis and high-frequency finance. They are not only utilized to obtain a good prediction accuracy, but also to address variable/model selection problems associated with various challenges such as noise accumulation, spurious correlation, scalability and stochastic errors. We refer readers to Fan, Lv, and Qi (2011); Varian (2014) for excellent reviews, and Hastie, Tibshirani, and Friedman (2009); James, Witten, Hastie, and Tibshirani (2017) for in-depth treatments of high-dimensional methods. One of the best known high-dimensional methods is LASSO or otherwise known as $L1$ regularization, which has recently received some attention in the operations management literature for predictive models, and here are some recent examples. Ma, Fildes, and Huang (2016) apply LASSO to select key explanatory variables from a high dimensional data set for demand forecasting for SKU retail sales. Martinez, Schmuck, Pereverzyev, Pirkerc, and Haltmeier (2018) use LASSO as one of the methods to develop a machine learning framework for customer purchase prediction in a non-contractual setting. Bertsimas, O’Hair, Relyea, and Silberholz (2016) use LASSO as one of the methods to predict the outcomes of clinical trials. In the context of a Red Cross fund-raising campaign, Ryzhov, Han, and Bradic (2016) employ LASSO to logistic regression models to identify key interactions between designs (e.g., the presence or absence of a free gift) and various donor segments. In contrast to the aforementioned literature, we study product returns, and extract features common to many product manufacturers from our data set. Furthermore, we consider higher-order interaction terms (e.g., product type and retailer) and initialize multiple machine learning methods (LASSO, SCAD,</p>	<p>Мы используем многомерные методы машинного обучения, которые становятся все более популярными и широко используются в таких областях, как геномика, нейробиология, анализ социальных сетей и высокочастотное финансирование. Они используются не только для получения хорошей точности прогнозирования, но и для решения проблем выбора переменных / моделей, связанных с различными проблемами, такими как накопление шума, ложная корреляция, масштабируемость и стохастические ошибки. Мы отсылаем читателей к Fan, Lv и Qi (2011); Varian (2014) за отличные отзывы и Hastie, Tibshirani и Friedman (2009); James, Witten, Hastie и Tibshirani (2017) за углубленное рассмотрение многомерных методов. Одним из наиболее известных многомерных методов является LASSO или другой известный как регуляризация $L1$, который недавно получил некоторое внимание в литературе по управлению операциями для прогнозирующих моделей, и вот несколько недавних примеров. Ma, Fildes и Huang (2016) применяют LASSO, чтобы выбрать ключевые объясняющие переменные из набора многомерных данных для прогнозирования спроса на розничные продажи SKU. Martinez, Schmuck, Pereverzyev, Pirkerc и Haltmeier (2018) используют LASSO в качестве одного из методов разработки системы машинного обучения для прогнозирования покупок в условиях, не связанных с договором. Bertsimas, O’Hair, Relyea и Silberholz (2016) используют LASSO в качестве одного из методов прогнозирования результатов клинических испытаний. В контексте кампании по сбору средств Красного Креста Ryzhov, Han и Bradic (2016) используют LASSO для моделей логистической регрессии, чтобы идентифицировать ключевые взаимодействия между проектами (например, наличие или отсутствие бесплатного дара) и различными сегментами доноров. В отличие от вышеупомянутой литературы, мы изучаем возврат продукта и извлекаем из нашего набора данных характеристики, общие для многих производителей продукта. Кроме того, мы рассматриваем термины</p>

Elastic Net, Random Forest, Gradient Boosting) to derive and calibrate predictive models.	взаимодействия более высокого порядка (например, тип продукта и розничный продавец) и инициализируем несколько методов машинного обучения (LASSO, SCAD, Elastic Net, Random Forest, Gradient Boosting) для получения и калибровки прогностических моделей.
3. Empirical setting	3. Эмпирическая настройка
In this section, we first introduce the operational context and process in our study in Section 3.1. Section 3.2 describes the data, and Section 3.3 discusses how we derive predictor variables from the data.	В этом разделе мы впервые представим операционный контекст и процесс в нашем исследовании в разделе 3.1. В разделе 3.2 описываются данные, а в разделе 3.3 обсуждается, как мы получаем предикторные переменные из данных.
3.1. Operational context and process	3.1. Операционный контекст и процесс
Company A manufactures car accessories (seat covers, car covers and dash covers) at two factory locations – in California and Mexico. Consumers place orders at an online retailer for a specific vehicle (make, model, year, trim level) and a specific product type. A consumer can choose the fabric, color, design and whether to have a personalized logo. Sometimes, she may request multiple products in the same order. The order information is transferred to Company A which manufactures and ships the product to the consumer in about a week.	Компания А производит автомобильные аксессуары (чехлы на сиденья, автомобильные чехлы и приборные панели) на двух заводах - в Калифорнии и Мексике. Потребители размещают заказы в онлайн-магазине на конкретное транспортное средство (марка, модель, год выпуска, уровень комплектации) и конкретный тип продукта. Потребитель может выбрать ткань, цвет, дизайн и выбрать индивидуальный логотип. Иногда она может запросить несколько товаров в одном и том же порядке. Информация о заказе передается компании А, которая производит и отгружает продукт потребителю в течение недели.
Once Company A receives an order, it is released into production immediately unless there is substantial pending work. The entire manufacturing process can be broadly categorized into three stages. The pre-sewing stage comprises of printing the order label, which contains detailed instructions for each of the subsequent operations, cutting of fabrics by computerized cutting machines, and placement of cut pieces in a plastic bag with the order label to be routed for sewing. During the sewing stage, sewers perform tasks such as joining, binding, embroidering and adding logos, which vary with the product types. In the last post-sewing stage, finished goods are inspected, packed and shipped to the consumer. There are some random inspections by the floor supervisor in the pre-sewing and post-sewing stages. Most products are exclusively made in the U.S. while some are produced jointly in both factories. Returns. A product may be returned within the return period for refund, or beyond the return period for repair or replacement under warranty. The return policy is set by each retailer and return	Как только Компания А получает заказ, он сразу же запускается в производство, если не будет выполнено значительных незавершенных работ. Весь производственный процесс можно разделить на три этапа. Этап предварительного шитья состоит из печати этикетки заказа, которая содержит подробные инструкции для каждой из последующих операций, резки тканей компьютеризированными отрезными машинами и помещения отрезанных кусочков в пластиковый пакет с этикеткой заказа, которая будет направлена для шитья. На этапе шитья швея выполняет такие задачи, как соединение, переплет, вышивание и добавление логотипов, которые зависят от типа продукта. На последнем этапе после шитья готовая продукция проверяется, упаковывается и отправляется потребителю. Существует несколько выборочных проверок, проводимых надзором за полами на этапах до и после шитья. Большинство продуктов производится исключительно в США, а некоторые производятся совместно на обоих заводах. Возвраты. Продукт может быть возвращен в течение периода возврата для возврата денег или после периода возврата для ремонта или замены по

<p>policies determine aspects such as time within which a return has to be made for a refund, allowing returns to physical stores, etc. All the returns are handled by Company A, not the retailers. If a returned product is confirmed to be in a resalable condition, it is first kept in the warehouse depending on the available space, and then disposed if it is not sold within a certain time window. If a returned product is for repair or replacement, the manufacturer needs to allocate resources to meet the request. Therefore, the returns handling increases operational overhead in staffing and resources.</p>	<p>гарантии. Политика возврата устанавливается каждым розничным продавцом, и политика возврата определяет такие аспекты, как время, в течение которого необходимо произвести возврат для возврата, разрешение возврата в физические магазины и т. Д. Все возвраты обрабатываются компанией А, а не продавцами. Если подтверждено, что возвращенный продукт находится в состоянии допускающем перепродажу, он сначала хранится на складе в зависимости от доступного пространства, а затем утилизируется, если он не продается в течение определенного периода времени. Если возвращенный продукт предназначен для ремонта или замены, изготовителю необходимо выделить ресурсы для удовлетворения запроса. Таким образом, обработка возвратов увеличивает операционные издержки на управление и ресурсы.</p>
<p>3.2.Data description</p>	<p>3.2. Описание данных</p>
<p>In our study, we focus on three main product types that comprise over 90% of sales volume for Company A: seat cover, car cover and dash cover. They are sold to consumers in the United States and most sales are through 13 retailers. There was a small fraction (0.17%) of the products that were returned and resold during our study period, but they were excluded during data cleaning because they comprised of a negligible fraction of the sales. This gives us a data set containing 331,390 products sold between July 2012 and September 2015, for a total of 39 months denoted as periods $t = 1, \dots, 39$. In Fig. 1, we show how sales, returns (in units) and return rates change over 39 periods in the focal firm in our study. One can observe that both returns as well as return rates fluctuate over time.</p>	<p>В нашем исследовании мы фокусируемся на трех основных типах продуктов, которые составляют более 90% объема продаж для компании А: чехлы на сиденья, чехлы на автомобиль и панель приборов. Они продаются потребителям в Соединенных Штатах, и большинство продаж осуществляется через 13 магазинов. Была небольшая доля (0,17%) продуктов, которые были возвращены и перепроданы в течение нашего периода исследования, но они были исключены во время очистки данных, поскольку они составляли незначительную долю продаж. Это дает нам набор данных, содержащий 331 390 продуктов, проданных в период с июля 2012 г. по сентябрь 2015 г., за 39 месяцев, обозначенных как периоды $t = 1, \dots, 39$. На рис. 1 мы показываем, как продажи, доходы (в единицах) и нормы прибыли меняются в течение 39 периодов в фокусе нашего исследования. Можно наблюдать, что как доходность, так и возвраты колеблются во времени.</p>
<p>Though aggregate return volume in each period is of first-order interest, Company A is also interested in predicting return volume by each product type and each retailer for the following reasons. First, the revenue loss varies among different product types, for example, the revenue loss due to refund of seat cover is often higher than one for dash cover. Second, the ability to resell the returned product may also vary with the product type. Third, since different product types have different manufacturing processes, when products are returned for repair or replacement under</p>	<p>Хотя совокупный объем прибыли в каждом периоде представляет интерес первого порядка, Компания А также заинтересована в прогнозировании объема возврата по каждому типу продукта и каждому розничному продавцу по следующим причинам. Во-первых, потеря дохода варьируется среди разных типов продуктов, например, потеря дохода из-за возврата покрытия сиденья часто выше, чем потеря покрытия приборной панели. Во-вторых, возможность перепродать возвращенный товар также может варьироваться в зависимости от типа товара. В-третьих, поскольку различные типы продуктов имеют разные</p>

warranty, the increased operational workload and costs may also be quite different. Fourth, the same seat cover may lead to a larger revenue loss due to return for refund through a retailer with a liberal re-turn policy, but may result in less costly return for repair through a retailer with a 30-day return policy.

производственные процессы, когда продукты возвращаются для ремонта или замены по гарантии, повышенная рабочая нагрузка и затраты также могут быть весьма разными. В-четвертых, то же самое покрытие сидений может привести к большей потере дохода из-за возврата средств через розничную торговлю с либеральной политикой возврата, но может привести к менее дорогостоящему возврату для ремонта через розничную торговлю с 30-дневной политикой возврата.

Fig. 1. Sales, returns and return rates over 39 months, between July 2012 and September 2015.

Рис. 1. Продажи, возвраты и нормы возвратов за 39 месяцев, с июля 2012 года по сентябрь 2015 года.



Table 1 A sample operational data set.

Таблица 1 Образец рабочего набора данных.

ReleaseDate	ScanDate	OrderNo	SerialNo	Product	Retailer	Logo	Operation	EmplID	Loc	FName	LName	ReturnDate
5-13-2015	5-13-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Printing labels	xx01	US	xxxx	xxxx	7-10-2015
5-13-2015	5-14-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Cutting fabrics	xx12	US	xxxx	xxxx	7-10-2015
5-13-2015	5-15-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Joining	xx53	US	xxxx	xxxx	7-10-2015
5-13-2015	5-15-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Binding	xx53	US	xxxx	xxxx	7-10-2015
5-13-2015	5-18-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Logo	xx26	US	xxxx	xxxx	7-10-2015
5-13-2015	5-18-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Random Inspection	xx45	US	xxxx	xxxx	7-10-2015
5-13-2015	5-18-2015	149038902	L49325331	SCxxxx	xxxx.com	USC001	Packing	xx34	US	xxxx	xxxx	7-10-2015

There are many possible reasons why a particular product may be returned and understanding them is helpful when building a predictive model. Next, we provide a sample operational level data set obtained from Company A (see Table 1) to provide a context to discuss these reasons. Some values in the table are hidden or modified to preserve confidentiality. The data used in the study includes both retailer level and operational level data sets obtained from the company. The retailer level data contains the return policy and name of each retailer.


Есть много возможных причин, по которым конкретный продукт может быть возвращен, и понимание его полезно при построении прогнозной модели. Далее мы предоставляем примерный набор данных операционного уровня, полученный от Компании А (см. Таблицу 1), чтобы предоставить контекст для обсуждения этих причин. Некоторые значения в таблице скрыты или изменены, чтобы сохранить конфиденциальность. Данные, использованные в исследовании, включают в себя наборы данных как на уровне розничной торговли, так и на уровне эксплуатации, полученные от компании. Данные уровня розничной торговли содержат политику возврата и имя каждого розничного продавца.

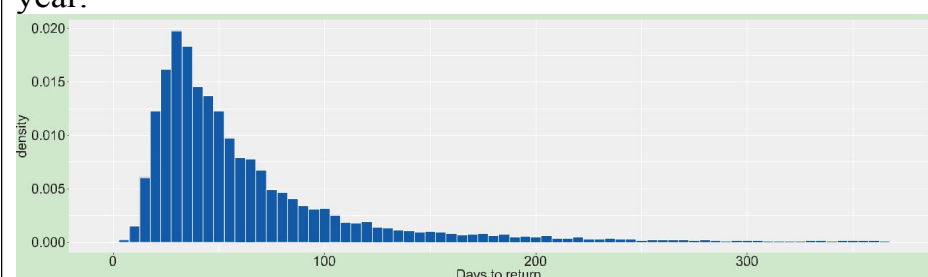
This example shows 5 different operations performed on a product with serial number L49325331 in order number

В этом примере показаны 5 различных операций, выполненных с продуктом с серийным номером L49325331 в номере заказа 149038902.

<p>149038902. We can identify the product type—seat cover—by the prefix of the product, and observe that this particular product was released into production and the order label was printed on Wednesday, 5/13/2015. The next day, an employee read the label to find out which fabrics to pick up from inventory, and cut the fabric in preparation for sewing according to the specification. On Friday, 5/15/2015, an employee with ID xx53 performs two types of sewing operations—Joining and Binding. On Monday, 5/18/2015, a different sewer adds a logo and finishes this product. A floor supervisor (usually a sewer with more than 10 years of experience) performs a random inspection on the product, does not notice any defects, and she passes the product to an employee who packs the product. Depending on the time of day, this product is shipped to a consumer on the same day or next day. The product is returned to the factory on 7/10/2015 as shown in the ReturnDate column</p>	<p>Мы можем определить тип продукта - чехол для сиденья - по префиксу продукта и наблюдать, что этот конкретный продукт был выпущен в производство и на этикетке заказа. был напечатан в среду, 13.05.2015. На следующий день сотрудник прочитал этикетку, чтобы выяснить, какие ткани выбрать из инвентаря, и разрезать ткань при подготовке к пошиву в соответствии со спецификацией. В пятницу 15.05.2015 сотрудник с идентификатором xx53 выполняет два типа швейных операций - соединение и связывание. В понедельник, 18.05.2015, другой работник швейного цеха добавляет логотип и дополняет этот продукт. Куратор (обычно работник швейного цеха с опытом работы более 10 лет) проводит выборочную проверку продукта, не замечает никаких дефектов и передает продукт сотруднику, который упаковывает продукт. В зависимости от времени суток этот товар доставляется потребителю в тот же день или на следующий день. Продукт возвращается на фабрику 7/10/2015, как показано в столбце ReturnDate.</p>
<p>Which factors may have contributed to this return? This product was ordered just before the summer. There may be a seasonality effect for returns if the product fades during the summer heat. Product type may influence returns; for example, seat covers usually cost much more than dash covers, so a customer is more likely to go through the trouble of returning seat covers. Seat covers might also have higher defects because of more difficult sewing operations. Retailer xxxx.com in Table 1 is found to have a 60 days return policy in the retailer level data set, and the product is returned between 30 days and 60 days, so the return may have been because the consumer changed his/her mind or due to a defect in the product. The production processes and resources used in manufacturing the product may cause defects and in turn result in returns. The example in Table 1 shows that two sewers performed three sewing tasks on this product, and we may ask if single tasking (three sewers, each doing a single task) leads to lower defects and thus lower returns. It is also possible that the consumer may have simply changed her mind, and historical returns may partly capture this factor. We could also check whether an order of multiple products lead to higher (or lower) returns. In the next subsection, we explore the specific predictor variables used in the study in more detail.</p>	<p>Какие факторы могли способствовать этому возвращению? Этот товар был заказан незадолго до лета. Может быть эффект сезонности для возвращений, если продукт выцветает во время летней жары. Тип продукта может влиять на возврат; Например, чехлы на сиденья обычно стоят намного дороже, чем чехлы для приборной панели, поэтому покупатель, скорее всего, столкнется с проблемой возврата чехлов на сиденья. Чехлы на сиденья также могут иметь более сильные дефекты из-за более сложных операций шитья. У розничного продавца xxxx.com в таблице 1 установлено, что в наборе данных уровня розничного продавца действует политика возврата в течение 60 дней, а продукт возвращается в течение 30–60 дней, поэтому возврат мог быть вызван тем, что потребитель изменил свое мнение или из-за дефекта в продукте. Производственные процессы и ресурсы, используемые при изготовлении продукта, могут вызвать дефекты и, в свою очередь, привести к возврату. Пример в Таблице 1 показывает, что два работника швейного цеха выполнили три задачи по пошиву этого продукта, и мы можем спросить, приводит ли одна задача (три работника швейного цеха, каждый из которых выполняет одну задачу) к более несущественным дефектам и, следовательно, к более низкой отдаче. Также возможно, что потребитель мог просто передумать, и исторические доходы могут частично охватить этот фактор. Мы также можем проверить, приводит ли заказ нескольких продуктов к более высокой (или более низкой)</p>

	доходности. В следующем подразделе мы рассмотрим конкретные переменные предиктора, используемые в исследовании более подробно.																																																		
The predicted variable in the study is $Return_{stij}$, that is, how many units of product type i will be returned out of sales through retailer j in period t . As is clear from Table 1, every item that is returned can be tracked, and the details about retailer, product and production process characteristics for the item are available. While we collected data until October 2016, we analyze returns only for sales until Sep 2015 in our study so that almost all returns, which occur within one year, are captured.	Прогнозируемой переменной в исследовании является $Return_{stij}$, то есть сколько единиц продукта типа i будет возвращено из продаж через розничного продавца j в период t . Как видно из таблицы 1, каждый возвращаемый товар можно отслеживать, и для него доступны подробные данные о розничном продавце, характеристиках продукта и производственного процесса. Несмотря на то, что мы собирали данные до октября 2016 года, в нашем исследовании мы анализируем только выручку от продаж до сентября 2015 года, так что фиксируются почти все результаты, которые происходят в течение одного года.																																																		
3.3.Predictor variables	3.3. Независимые переменные																																																		
Our focus in this subsection is to identify the variables that can help predict the volume of returns defined as $Return_{stij}$, where the subscripts t, i, j denote the period, product type and retailer respectively. As potential predictors of returns, we consider production process and resources, multi-product order and historical returns in addition to aforementioned factors—sales, time, product type, retailer. We list the predictor variables reflecting these factors in Table 2 along with their definitions. In the following, we discuss each of these predictor variables and the justification for using them in the prediction models.	Наша цель в этом подразделе - определить переменные, которые могут помочь предсказать объем возвратов, определенных как $Return_{stij}$, где индексы t, i, j обозначают период, тип продукта и розничного продавца соответственно. В качестве потенциальных предикторов доходности мы рассматриваем производственный процесс и ресурсы, заказы из нескольких продуктов и исторические доходы в дополнение к вышеупомянутым факторам: объем продаж, время, тип продукта, розничный продавец. Перечислим предикторные переменные, отражающие эти факторы, в таблице 2 вместе с их определениями. Далее мы обсудим каждую из этих переменных-предикторов и обоснование их использования в моделях прогнозирования.																																																		
Table 2 Definition of response and predictor variables for $t = 1, \dots, 39, i = 1, 2, 3, j = 1, \dots, 13$.	Таблица 2 Определение переменных ответа и предиктора для $t = 1, \dots, 39, i = 1, 2, 3, j = 1, \dots, 13$																																																		
<table><tr><th>Category</th><th>Variable name</th><th>Definition</th></tr><tr><td>Response variable</td><td>$Return_{stij}$</td><td>Number of returns out of sales of product type i via retailer j in period t.</td></tr><tr><td>Sales effect</td><td>$Sales_{stij}$</td><td>Number of sales of product type i via retailer j in period t.</td></tr><tr><td>Time effect</td><td>$Year_t$</td><td>Sales year from 2012, 2013, 2014 to 2015, used to capture time trend.</td></tr><tr><td></td><td>$Month_t$</td><td>Dummy variable for the month corresponding to period t, used to capture seasonality.</td></tr><tr><td>Product effect</td><td>$Product_i$</td><td>Dummy variable for product type i.</td></tr><tr><td>Retailer effect</td><td>$Retailer_j$</td><td>Dummy variable for retailer j ranked by volume.</td></tr><tr><td rowspan="11">Production process and resources</td><td>$SewerCnt_{stij}$</td><td>Average number of sewers per product among $Sales_{stij}$.</td></tr><tr><td>$SewingTaskCnt_{stij}$</td><td>Average number of sewing tasks per product among $Sales_{stij}$.</td></tr><tr><td>$SewingDays_{stij}$</td><td>Average number of days in sewing stage per product among $Sales_{stij}$.</td></tr><tr><td>$BacklogDays_{stij}$</td><td>Average number of backlog days before production per product among $Sales_{stij}$.</td></tr><tr><td>$ProductionDays_{stij}$</td><td>Average number of days in entire production process per product among $Sales_{stij}$.</td></tr><tr><td>$Workload_{stij}$</td><td>Average number of finished products per employee in period t.</td></tr><tr><td>$JointProduction\%_{it}$</td><td>Fraction of products jointly manufactured in period t.</td></tr><tr><td>$CustomFabric\%_{it}$</td><td>Fraction of products using customized fabrics among $Sales_{stij}$.</td></tr><tr><td>$Logo\%_{it}$</td><td>Fraction of products ordered with special logos among $Sales_{stij}$.</td></tr><tr><td>$InspectionPreSewing\%_{it}$</td><td>Random inspection rate before sewing among $Sales_{stij}$.</td></tr><tr><td>$InspectionPostSewing\%_{it}$</td><td>Random inspection rate after sewing among $Sales_{stij}$.</td></tr><tr><td>Multi-product effect</td><td>$MultiProduct\%_{it}$</td><td>Fraction of orders with two or more products among $Sales_{stij}$.</td></tr><tr><td>Historical returns</td><td>$LaggedReturns_{stij}$</td><td>Observed returns in period t from sales in periods $t - 4, t - 5, t - 6$.</td></tr></table>	Category	Variable name	Definition	Response variable	$Return_{stij}$	Number of returns out of sales of product type i via retailer j in period t .	Sales effect	$Sales_{stij}$	Number of sales of product type i via retailer j in period t .	Time effect	$Year_t$	Sales year from 2012, 2013, 2014 to 2015, used to capture time trend.		$Month_t$	Dummy variable for the month corresponding to period t , used to capture seasonality.	Product effect	$Product_i$	Dummy variable for product type i .	Retailer effect	$Retailer_j$	Dummy variable for retailer j ranked by volume.	Production process and resources	$SewerCnt_{stij}$	Average number of sewers per product among $Sales_{stij}$.	$SewingTaskCnt_{stij}$	Average number of sewing tasks per product among $Sales_{stij}$.	$SewingDays_{stij}$	Average number of days in sewing stage per product among $Sales_{stij}$.	$BacklogDays_{stij}$	Average number of backlog days before production per product among $Sales_{stij}$.	$ProductionDays_{stij}$	Average number of days in entire production process per product among $Sales_{stij}$.	$Workload_{stij}$	Average number of finished products per employee in period t .	$JointProduction\%_{it}$	Fraction of products jointly manufactured in period t .	$CustomFabric\%_{it}$	Fraction of products using customized fabrics among $Sales_{stij}$.	$Logo\%_{it}$	Fraction of products ordered with special logos among $Sales_{stij}$.	$InspectionPreSewing\%_{it}$	Random inspection rate before sewing among $Sales_{stij}$.	$InspectionPostSewing\%_{it}$	Random inspection rate after sewing among $Sales_{stij}$.	Multi-product effect	$MultiProduct\%_{it}$	Fraction of orders with two or more products among $Sales_{stij}$.	Historical returns	$LaggedReturns_{stij}$	Observed returns in period t from sales in periods $t - 4, t - 5, t - 6$.	
Category	Variable name	Definition																																																	
Response variable	$Return_{stij}$	Number of returns out of sales of product type i via retailer j in period t .																																																	
Sales effect	$Sales_{stij}$	Number of sales of product type i via retailer j in period t .																																																	
Time effect	$Year_t$	Sales year from 2012, 2013, 2014 to 2015, used to capture time trend.																																																	
	$Month_t$	Dummy variable for the month corresponding to period t , used to capture seasonality.																																																	
Product effect	$Product_i$	Dummy variable for product type i .																																																	
Retailer effect	$Retailer_j$	Dummy variable for retailer j ranked by volume.																																																	
Production process and resources	$SewerCnt_{stij}$	Average number of sewers per product among $Sales_{stij}$.																																																	
	$SewingTaskCnt_{stij}$	Average number of sewing tasks per product among $Sales_{stij}$.																																																	
	$SewingDays_{stij}$	Average number of days in sewing stage per product among $Sales_{stij}$.																																																	
	$BacklogDays_{stij}$	Average number of backlog days before production per product among $Sales_{stij}$.																																																	
	$ProductionDays_{stij}$	Average number of days in entire production process per product among $Sales_{stij}$.																																																	
	$Workload_{stij}$	Average number of finished products per employee in period t .																																																	
	$JointProduction\%_{it}$	Fraction of products jointly manufactured in period t .																																																	
	$CustomFabric\%_{it}$	Fraction of products using customized fabrics among $Sales_{stij}$.																																																	
	$Logo\%_{it}$	Fraction of products ordered with special logos among $Sales_{stij}$.																																																	
	$InspectionPreSewing\%_{it}$	Random inspection rate before sewing among $Sales_{stij}$.																																																	
	$InspectionPostSewing\%_{it}$	Random inspection rate after sewing among $Sales_{stij}$.																																																	
Multi-product effect	$MultiProduct\%_{it}$	Fraction of orders with two or more products among $Sales_{stij}$.																																																	
Historical returns	$LaggedReturns_{stij}$	Observed returns in period t from sales in periods $t - 4, t - 5, t - 6$.																																																	
Fig. 2. Sales, returns and return rates by product type (left) and by	Рис. 2. Продажи, возвраты и нормы возврата по типу продукта (слева) и																																																		

<p>retailer (right) over all periods.</p> 	<p>по ритейлеру (справа) за все периоды.</p>
<p>Sales. As sales volume goes up, return volume is likely to increase. So, we include sales as a predictor.</p> <p>Time. Similar to Anderson et al. (2009) we consider two types of time effects: trend effect and monthly fixed effect (e.g. consumers may purchase products more impulsively during holiday seasons resulting in higher returns.) The trend effect is captured by the predictor variable $Year_t \in \{2012, 2013, 2014, 2015\}$, which depends on the period t; for example, $t = 1, \dots, 6$ implies 2012, $t = 7, \dots, 18$ implies 2013, and so on, recalling that the dataset begins in July 2012. The monthly fixed effect is captured by the dummy variable $Month_t \in \{January, \dots, November, December\}$, which depends on the month the period t corresponds to; for instance, $Month_1$ indicates July, $Month_2$ equals August, and so on. We provided the returns pattern over the 39 periods in Fig. 1. As an alternative to using year and month variables, we de-trended and de-seasonalized the data before estimating the models but we found little difference (see Online Appendix A1 for details).</p>	<p>Продажи. По мере того, как объем продаж растет, объем возврата может увеличиться. Итак, мы включаем продажи в качестве предиктора.</p> <p>Время. Аналогично Anderson et al. (2009) мы рассматриваем два типа временных эффектов: эффект тренда и ежемесячный фиксированный эффект (например, потребители могут покупать товары более импульсивно во время курортных сезонов, что приводит к более высокой доходности.) Эффект тренда фиксируется переменной-предиктором $Year_t \in \{2012, 2013, 2014, 2015\}$, который зависит от периода t; например, $t = 1, \dots, 6$ означает 2012, $t = 7, \dots, 18$ означает 2013 и т. д., Напомнив, что набор данных начинается в июле 2012 года. Фиксированный ежемесячный эффект фиксируется фиктивной переменной $Month_t \in \{Январь, \dots, ноябрь, декабрь\}$, которая зависит от месяца, которому период t соответствует; например, $Month_1$ указывает на июль, $Month_2$ равен августу и т. д. Мы представили модель доходности за 39 периодов на рис. 1. В качестве альтернативы использованию переменных года и месяца мы разобрали и сняли сезонность данных перед оценкой моделей, но обнаружили небольшую разницу (подробности см. В онлайн-приложении A1).</p>
<p>Product. The three product types in our study—seat cover, dash cover and vehicle cover—have distinct characteristics, and may have product specific fixed effects that influence returns. For example, a slightly larger vehicle cover may be considered fine by consumers, but a slightly loose seat cover may be regarded as defective and returned. We provided the returns for each product type in Fig. 2, and found that the return rate of car cover was much lower than that for the other two products. In a study of an online</p>	<p>Продукт. Три типа продуктов в нашем исследовании - чехол на сиденье, приборная панель и крышка транспортного средства - имеют разные характеристики и могут оказывать определенные эффекты продукта, которые влияют на возврат. Например, потребители могут считать несколько большую крышку транспортного средства более тонкой, но слегка ослабленную крышку сиденья можно считать дефектной и возвращенной. Мы представили данные о доходах для каждого типа продуктов на рис. 2 и обнаружили, что показатель возврата</p>

<p>retailer, Anderson et al. (2009) show variations in returns among different product categories. Therefore, we could expect that product effects are important in predicting return volume.</p>	<p>автомобильного покрытия был намного ниже, чем у двух других продуктов. В исследовании интернет-магазина Anderson et al. (2009) показывают различия в доходах между различными категориями продуктов. Следовательно, мы можем ожидать, что эффекты продукта важны для прогнозирования объема возврата.</p>
<p>Retailer. Based on previous research on the impact of return policies on sales and returns (e.g., Janakiramana et al., 2016 and references therein) and given different returns policies of retailers in our study, we would expect return volume to vary with retailers. In addition, there may be other important factors associated with a retailer that influence returns. Retailers with different return policies may attract different types of customers who exhibit different return behaviors. The number of brick-and-mortar stores of a retailer may have an impact on returns because more locations may make it easier to return a product: a warehouse club in our study operates hundreds of stores, whereas an auto specialty store has thousands of locations where a consumer may return a product. Therefore we expect that the retailer’s return policy and other characteristics may play an important role in returns, which are both captured by retailer indicator variables, as we showed earlier in Fig. 2. In our analysis, we also explored incorporating variables to reflect different dimensions of return policies such as return period, in-store return acceptance, in addition to the fixed retailer effect but we did not find the additional variables helpful in increasing the prediction accuracy.</p>	<p>Розничные продавцы. Основываясь на предыдущих исследованиях влияния политики возврата на продажи и прибыли (например, Janakiramana et al., 2016 и ссылки в них) и учитывая различные политики возврата розничных продавцов в нашем исследовании, мы ожидаем, что объем возврата будет варьироваться в зависимости от розничных продавцов. Кроме того, могут быть другие важные факторы, связанные с розничной торговлей, которые влияют на доходность. Ритейлеры с разной политикой возврата могут привлекать клиентов разных типов, которые демонстрируют разное поведение в отношении возврата. Количество обычных магазинов розничного предприятия может повлиять на возврат, поскольку большее количество мест может упростить возврат товара: в нашем исследовании складской клуб управляет сотнями магазинов, тогда как специализированный магазин по продаже автомобилей имеет тысячи местоположений, где потребитель может вернуть товар. Поэтому мы ожидаем, что политика возврата розничного предприятия и другие характеристики могут играть важную роль в доходах, которые оба фиксируются переменными показателями продавца, как мы показали ранее на рис. 2. В нашем анализе мы также исследовали включение переменных для отражения различных измерений политики возврата, такие как период возврата, принятие возврата в магазине, в дополнение к фиксированному эффекту розничного продавца, но мы не нашли дополнительные переменные, полезные для повышения точности прогнозирования.</p>
<p>Fig. 3. A histogram of days to return for products returned within a year.</p> 	<p>Рис. 3. Гистограмма дней возврата товаров, возвращенных в течение года.</p>

<p>Production processes and resources. Unlike prior studies of returns which considered only product and consumer purchase related factors, we consider various factors related to the production process that may contribute to returns due to defects. We measure most of the variables in this category as an average over each period, product and retailer combination. We track the average number of sewers and sewing tasks by <i>SewerCnttij</i> and <i>SewingTaskCnttij</i> for each product, to understand if more sewers or fewer sewing tasks per sewer (e.g. grouping some sewing operations into one task) lead to lower defects, and therefore are useful in predicting returns. We use <i>SewingDaystij</i> as a proxy to the actual sewing time, and we may expect this predictor to be negatively correlated with returns, that is, if sewers spend more time in sewing a product, the quality of the product may increase and the likelihood of returns may decrease. <i>BacklogDaystij</i> and <i>ProductionDaystij</i> reflect different aspects of the manufacturing process and may be useful in predicting returns. We are interested in whether the monthly workload of the factory and the fraction of items jointly produced in Mexico and US impact returns, and measure these effects by <i>Workloadt</i>, <i>JointProduction%</i> for each period. We investigate the effect of special customization on returns by <i>CustomFabric%</i> for the usage of in-house custom fabric, and <i>Logo%</i> for personalizing products with custom-logos, and the effect of random quality inspections by floor supervisor on returns before and after the sewing stage by <i>InspectionPreSewing%</i> and <i>InspectionPost Sewing%</i>.</p>	<p>Производственные процессы и ресурсы. В отличие от предыдущих исследований доходности, в которых рассматривались только факторы, связанные с покупкой продукта и потребителя, мы учитываем различные факторы, связанные с производственным процессом, которые могут способствовать возврату из-за дефектов. Мы измеряем большинство переменных в этой категории как среднее значение для каждой комбинации периода, продукта и розничного продавца. Мы отслеживаем среднее количество задач по шитью и шитью по <i>SewerCnttij</i> и <i>SewingTaskCnttij</i> для каждого продукта, чтобы понять, приводит ли большее количество швов или меньшее количество швейных задач на каждую швею (например, объединение некоторых швейных операций в одну задачу) и, следовательно, полезные в прогнозировании доходов. Мы используем <i>SewingDaystij</i> как прокси для фактического времени шитья, и мы можем ожидать, что этот предиктор будет отрицательно коррелировать с возвратами, то есть, если швея будет тратить больше времени на пошив продукта, качество продукта может возрасти и вероятность возврата может уменьшиться. <i>BacklogDaystij</i> и <i>ProductionDaystij</i> отражают различные аспекты производственного процесса и могут быть полезны при прогнозировании прибыли. Нас интересует, повлияют ли ежемесячная нагрузка на фабрику и доля товаров, совместно производимых в Мексике, и США, и измеряем эти эффекты по <i>Workloadt</i>, <i>JointProduction%</i> для каждого периода. Мы исследуем влияние специальной настройки на возвраты <i>CustomFabric%</i> для использования собственной нестандартной ткани и <i>Logo%</i> для персонализации продуктов с настраиваемыми логотипами, а также влияние случайных проверок качества, выполняемых куратором этажа, на возвраты до и после этап пошива с помощью <i>InspectionPreSewing%</i> и <i>InspectionPost Sewing%</i>.</p>
<p>Table 3 Summary statistics of response and predictor variables. Data size $N_{total} = 1360$ represents the number of data points in our model because we model the return volume per product type, per retailer, per time period.</p>	<p>Таблица 3 Сводная статистика ответов и предикторов. Размер данных $N_{total} = 1360$ представляет количество точек данных в нашей модели, потому что мы моделируем возвратный объем по типу продукта, по розничному продавцу, за период времени.</p>

Variable	Min	Mean	Median	Max	SD
<i>Returns</i>	0.00	21.05	6.00	509.00	49.53
<i>Sales</i>	1.00	243.70	99.00	2744.00	406.10
<i>SewerCnt</i>	1.00	3.28	3.27	6.75	1.14
<i>SewingTaskCnt</i>	1.00	3.61	3.86	8.75	1.33
<i>SewingDays</i>	1.00	1.99	1.80	8.00	0.77
<i>BacklogDays</i>	0.00	0.48	0.17	6.00	0.83
<i>ProductionDays</i>	2.00	7.07	6.97	15.00	3.23
<i>Workload</i>	410.00	1033.40	1117.40	1832.80	428.41
<i>JointProduction%</i>	0.00%	19.05%	0.42%	100.00%	29.50%
<i>CustomFabric%</i>	0.00%	5.41%	0.00%	73.84%	12.33%
<i>Logo%</i>	0.00%	2.74%	0.00%	72.73%	6.84%
<i>InspectionPreSewing%</i>	0.00%	0.07%	0.00%	26.07%	1.00%
<i>InspectionPost_Sewing%</i>	0.00%	11.97%	0.00%	116.67%	24.96%
<i>MultiProduct%</i>	0.00%	7.22%	2.97%	99.69%	16.82%
<i>LaggedReturns</i>	0.00	58.37	16.00	1490.00	148.17

Multi-product effect. Anderson et al. (2009) do not find a significant effect of multiple product purchase on returns in their study, but suggest that this may not always be true and in other applications, return rates may depend on whether a consumer purchased single or multiple products in an order. In our study, about 9% of products are sold as part of an order with two or more customized products. We utilize MultiProduct%_{tij} to capture and investigate this effect.

Мультипродуктовый эффект. Андерсон и соавт. (2009) не обнаруживают значительного влияния покупки нескольких продуктов на доходность в своем исследовании, но предполагают, что это не всегда может быть правдой, и в других приложениях нормы возврата могут зависеть от того, купил ли потребитель один или несколько продуктов в заказе. В нашем исследовании около 9% продуктов продаются как часть заказа с двумя или более индивидуальными продуктами. Мы используем MultiProduct%_{tij} для обнаружения и исследования этого эффекта.

Historical returns. We are interested in capturing new trends in returns that may not be captured by the aforementioned predictor variables. For example, recent returns data may include a change in consumer’s purchase and return behaviors, such as increased usage of smartphones to make purchases (Martin, 2018) which may be more impulsive and lead to higher returns than average online orders. In addition, using lagged values to predict current period’s value is not uncommon (for example, see Wilms, Basu, Bien, & Matteson, 2017). We use a predictor variable called *LaggedReturns*_{tij} to capture this effect, using historical returns data. To do this, we consider three criteria: (1) we want the returns status of historical data to be accurate, (2) we want to capture recent returns trend, 3) we want to reduce noise in low-volume returns. To satisfy these requirements,

Исторические возвращения. Мы заинтересованы в выявлении новых тенденций в доходах, которые могут не учитываться вышеупомянутыми переменными предиктора. Например, последние данные о возвратах могут включать изменение поведения покупателя и поведения при возврате, например, более широкое использование смартфонов для совершения покупок (Martin, 2018), что может быть более импульсивным и привести к более высокой прибыли, чем средние онлайн-заказы. Кроме того, использование запаздывающих значений для прогнозирования значения текущего периода не является редкостью (например, см. Wilms, Basu, Bien, & Matteson, 2017). Мы используем переменную-предиктор под названием *LaggedReturn*_{tij}, чтобы зафиксировать этот эффект, используя исторические возвращаемые данные. Для этого мы рассмотрим три критерия: (1) мы хотим, чтобы

<p>we define $LaggedReturns_{tij} := Returns_{t-6,ij} + Returns_{t-5,ij} + Returns_{t-4,ij}$ to predict $Returns_{tij}$. Next we discuss each of these points.</p>	<p>статус возвращаемых данных за прошлые периоды был точным, (2) мы хотим зафиксировать тенденцию недавних возвратов, 3) мы хотим уменьшить шум при возвратах с низким объемом. Чтобы удовлетворить эти требования, мы определяем $LaggedReturns_{tij} := Returns_{t-6,ij} + Returns_{t-5,ij} + Returns_{t-4,ij}$ для прогнозирования $Returns_{tij}$. Далее мы обсудим каждый из этих пунктов.</p>
<p>In the above Fig. 3, we observe that about 90% of returns occur within the first 4 months, thus the returns status largely becomes finalized within four months after sales of an item. For this reason, when we consider historical returns data, we want to go back at least 4 months, in other words, we want to use $Returns_{t-n,ij}$ for $n \geq 4$ for the prediction of $Returns_{tij}$ to satisfy our first criterion. Returns data from more than half a year ago may not reflect recent returns trend, and for this reason we propose to use $Returns_{t-n,i,j}$ for $n \leq 6$ to meet our second criterion. To reduce noise coming from low-volume returns, we aggregate historical returns over three months to satisfy our third criterion, instead of picking one of the variables $Returns_{t-4,ij}$, $Returns_{t-5,ij}$, $Returns_{t-6,ij}$. Please see Online Appendix A2 on details on how we construct the variable $LaggedReturns$ for the test periods.</p>	<p>На приведенном выше рисунке 3 мы видим, что около 90% возвратов происходит в течение первых 4 месяцев, таким образом, статус возвратов в основном становится окончательным в течение четырех месяцев после продажи товара. По этой причине, когда мы рассматриваем данные за прошлые периоды, мы хотим вернуться назад как минимум на 4 месяца, иными словами, мы хотим использовать $Returns_{t-n,ij}$ для $n \geq 4$ для прогнозирования $Returns_{tij}$, чтобы удовлетворить наш первый критерий. Данные о возвратах более полугода назад могут не отражать тенденцию недавних возвратов, и по этой причине мы предлагаем использовать $Returns_{t-n,i,j}$ для $n \leq 6$, чтобы соответствовать нашему второму критерию. Чтобы уменьшить шум от возвратов с малым объемом, мы агрегируем исторические возвращения за три месяца, чтобы удовлетворить наш третий критерий, вместо выбора одной из переменных $Returns_{t-4,ij}$, $Returns_{t-5,ij}$, $Returns_{t-6,ij}$. Пожалуйста, обратитесь к онлайн-приложению A2, где подробно описано, как мы строим переменную $LaggedReturns$ для тестовых периодов</p>
<p>We provide summary statistics of predictor variables in Table 3. One can notice a large variation among returns (and sales) ranging from 0 to 509 (and 1 to 2,744) depending on the period, product type and retailer. It requires an average of 3 to 4 sewers to finish a product and some sewers perform more than one task; the average time spent in the sewing stage is less than 2 days. Each employee, on average, finishes 1,033.40 products per period, as seen in <i>Workload</i>. When a product fails a post-sewing random inspection, it is sent back for rework and may be inspected again by the supervisor, which leads to a maximum possible 200% post-sewing inspection rate and explains the highest inspection rate of 116.67% for some period, product type and re-tailer. The $LaggedReturns$ are aggregated over three months and its summary statistics show a similar pattern as the ones for <i>Returns</i>, suggesting</p>	<p>Мы приводим сводную статистику переменных-предикторов в таблице 3. Можно заметить большие различия между доходами (и продажами) в диапазоне от 0 до 509 (и от 1 до 2744) в зависимости от периода, типа продукта и розничного продавца. Для отделки продукта требуется в среднем от 3 до 4 коллекторов, а некоторые коллекторы выполняют более одной задачи; среднее время, проведенное на стадии шитья, составляет менее 2 дней. В среднем каждый сотрудник готовит 1033,40 продуктов за период, как видно из раздела «Рабочая нагрузка». Когда изделие не проходит случайную проверку после шитья, оно отправляется обратно на доработку и может снова проверяться супервизором, что приводит к максимально возможной 200% проверке после шитья и объясняет самый высокий показатель проверки 116,67% для некоторый период, тип продукта и ретейлер. $LaggedReturns$ агрегируются за три месяца, и его сводные статистические данные</p>

<p><i>that historical returns may be an important predictor variable.</i></p>	<p>показывают аналогичную картину, что и для Returns, предполагая, что исторические возвраты могут быть важной переменной предиктора.</p>
<p>Recall from Section 3.2 that our data contains 331,390 products sold between July 2012 and September 2015, thus 331,390 represents the total number of transactions (items purchased). As we discussed at the beginning of this subsection, we are modeling $Return_{tij}$ –the volume of returns per product type, per store, per time period. Thus, the index “tij” represents one data point, therefore we have a much smaller set of data points, $N_{total} = 1,360$. Since $t = 39$ periods, $i = 3$ product types, and $j = 13$ retailers, one might assume that $N_{total} = 39 \times 3 \times 13 = 1,521$. The reason $N_{total} = 1,360 < 1,521$ is due to missing data points, that is, no sales and hence no returns for certain “tij” values (see Figure A1 in Online Appendix A5 for details).</p>	<p>Напомним из Раздела 3.2, что наши данные содержат 331 390 продуктов, проданных в период с июля 2012 года по сентябрь 2015 года, таким образом, 331 390 представляет общее количество транзакций (приобретенных товаров). Как мы уже говорили в начале этого подраздела, мы моделируем $Return_{tij}$ - объем возвратов по типу продукта, по магазину, за период времени. Таким образом, индекс «tij» представляет одну точку данных, поэтому мы имеем гораздо меньший набор точек данных, $N_{total} = 1,360$. Поскольку $t = 39$ периодов, $i = 3$ типа продуктов и $j = 13$ розничных продавцов, можно предположить, что $N_{total} = 39 \times 3 \times 13 = 1,521$. Причина $N_{total} = 1,360 < 1,521$ связана с отсутствием точек данных, то есть без продаж и, следовательно, без возврата для определенных значений «tij» (см. рисунок A1 в онлайн приложении A5 для деталей).</p>
<p>4.Main effects models</p>	<p>4. Основные модели эффектов</p>
<p>In this section, we explore various models to predict return volume using some or all of the variables listed in Table 2. We would ideally like a parsimonious model that has good predictive power and can be easily implemented in practice. Then, a natural question is—can we use only sales, time, product type and retailer information to build a model with reasonable prediction performance? We consider this model to be the baseline main effects model in our study, because these variables can be readily derived from a company’s ERP system and an operations manager can easily implement this predictive model using a spreadsheet, without the effort required to incorporate the remaining predictor variables. In Section 4.1, we sequentially add the predictor variables discussed above, and then present in Section 4.2 the results of model fit (training set) and prediction (test set) for all the models.</p>	<p>В этом разделе мы рассмотрим различные модели для прогнозирования возвратного объема, используя некоторые или все переменные, перечисленные в таблице 2. В идеале мы хотели бы экономную модель, которая обладает хорошей предсказательной способностью и может быть легко реализована на практике. Тогда возникает естественный вопрос: можем ли мы использовать только данные о продажах, времени, типе продукта и розничной торговле для построения модели с разумной эффективностью прогнозирования? Мы считаем эту модель базовой моделью основных эффектов в нашем исследовании, поскольку эти переменные могут быть легко получены из ERP-системы компании, и менеджер по операциям может легко внедрить эту прогнозную модель с помощью электронной таблицы, без усилий, необходимых для включения оставшегося предиктора. переменные. В разделе 4.1 мы последовательно добавляем переменные предиктора, рассмотренные выше, а затем представляем в разделе 4.2 результаты модели f (обучающий набор) и прогнозирования (набор тестов) для всех моделей.</p>
<p>4.1.Model specification</p>	<p>4.1. Спецификация модели</p>
<p>In the following simple model (1), we first use the predictor, <i>Sales</i>, to predict how many units out of such sales will be re turned.</p>	<p>В следующей простой модели (1) мы сначала используем предиктор Sales, чтобы предсказать, сколько единиц из таких продаж будет возвращено.</p>
<p>$Returns_{tij} = \beta_0 + \beta_1 Sales_{tij} + \epsilon_{tij} (1)$</p>	

<p>In the next model (2) we explore whether trend and seasonality effects help increase the prediction of return volume.</p>	<p>В следующей модели (2) мы исследуем, помогают ли трендовые и сезонные эффекты повысить прогнозируемый объем прибыли.</p>
$Returns_{tij} = \beta_0 + \beta_1 Sales_{ij} + \beta_2 Year_t + \beta^t Month_t + \epsilon_{tij} \quad (2)$	
<p>We now want to check the effects of product types and retailers on the prediction of return volume, but we don't know which effect is more important in predicting returns. Thus, we first investigate the product effect in model (3) and retailer effect in model (4) individually, and then both effects in the baseline main effects model Simultaneously.</p>	<p>Теперь мы хотим проверить влияние типов продуктов и розничных продавцов на прогнозирование объема прибыли, но мы не знаем, какой эффект более важен при прогнозировании доходности. Таким образом, мы сначала исследуем эффект продукта в модели (3) и эффект розничной торговли в модели (4) по отдельности, а затем оба эффекта в базовой модели основных эффектов одновременно.</p>
<div style="background-color: #e0f0e0; padding: 10px;"> $Returns_{tij} = \beta_0 + \beta_1 Sales_{ij} + \beta_2 Year_t + \beta_3 Month_t + \beta_4 Product_i + \epsilon_{tij} \quad (3)$ $Returns_{ij} = \beta_0 + \beta_1 Sales_{ij} + \beta_2 Year_t + \beta^t Month_t + \beta_5^j Retailer_j + \epsilon_{ij} \quad (4)$ $Returns_{ij} = \beta_0 + \beta_1 Sales_{ij} + \beta_2 Year_t + \beta^t Month_t + \beta^i Product_i + \beta_5^j Retailer_j + \epsilon_{ij} \quad (5)$ </div>	
<p>To simplify the exposition, we adopt column vector notation (shown in bold). For example, U consists of the remaining 13 variables shown in Table 2, and we add it to obtain the full main effects model (6) below.</p>	<p>Чтобы упростить изложение, мы используем векторное обозначение столбца (выделено жирным шрифтом). Например, U состоит из оставшихся 13 переменных, показанных в таблице 2, и мы добавляем его, чтобы получить полную модель основных эффектов (6) ниже.</p>
<p>For all the models (1) to (6) above, we considered a family of power transformations (Box&Cox,1964) for our response variable Returnstij to maximize normality before fitting models to the data; however, we did not find a need for a transformation</p>	<p>Для всех моделей (1) - (6), описанных выше, мы рассмотрели семейство степенных преобразований (Box & Cox, 1964) для нашей переменной отклика Returnstij, чтобы максимизировать нормальность до подгонки моделей к данным; Однако мы не нашли необходимости в преобразовании</p>
<p>4.2.Results from main effects models</p>	<p>4.2.Результаты моделей основных эффектов</p>
<p>To evaluate our models, we assign the first 33 periods (7/2012 to 3/2015) to be our training set, and the last 6 periods (4/2015 to 9/2015) to be our test set. We discuss alternative ways of splitting the data into training and test sets in our robustness checks in Section 6. For models (1) to (6) we present in Table 4 the model fit in training set, and the prediction performance in test set.</p>	<p>Для оценки наших моделей мы назначаем первые 33 периода (с 7/2012 по 3/2015) нашим обучающим набором, а последние 6 периодов (с 4 по 2015 по 9/2015) - нашим тестовым набором. Мы обсуждаем альтернативные способы разделения данных на обучающие и тестовые наборы в наших проверках надежности в Разделе 6. Для моделей (1) - (6) мы представляем в Таблице 4 модель соответствия в обучающем</p>

<p>In our study, we measure the model fit and prediction performance by MSE (equivalently, Prediction Error) and R^2. For the purpose of model comparison and selection for main effects models, we focus on adjusted R^2 and Akaike Information Criterion (AIC). Alternative measures may include Mallows's C_p, Bayesian Information Criterion (BIC), Risk Inflation Criterion (RIC), and predicted R^2. In the OLS setting, C_p and AIC are proportional to each other, therefore only one is needed. BIC and RIC generally place a heavier penalty on models with more variables compared to AIC (see Foster & George, 1994.) Predicted R^2 method is identical to leave-one-out cross-validation (CV, Stone, 1974), and is asymptotically the same method as AIC for model selection (Stone, 1977).</p>	<p>наборе и эффективность прогнозирования в тестовом наборе.</p> <p>В нашем исследовании мы измеряем модель и производительность прогнозирования по MSE (эквивалентно, ошибка прогнозирования) и R^2. В целях сравнения и выбора моделей для моделей с основными эффектами мы сосредоточены на скорректированном R^2 и информационном критерии Акаике (AIC). Альтернативные меры могут включать C_p Мэллоу, критерий байесовской информации (BIC), критерий инфляции риска (RIC) и прогнозируемый R^2. В настройке OLS C_p и AIC пропорциональны друг другу, поэтому необходим только один. BIC и RIC обычно накладывают более серьезные штрафы на модели с большим количеством переменных по сравнению с AIC (см. Foster & George, 1994.) Метод Predicted R^2 идентичен неконтролируемой перекрестной проверке (CV, Stone, 1974) и асимптотически тот же метод, что и в AIC для выбора модели (Stone, 1977).</p>
<p>We observe that the simplest model (1) yields $R^2 = 0.864$ for training set, and $R^2 = 0.836$ in test set for predicting returns over the future 6 months. This shows that the predictor variable sales by itself explains a large portion of variability in return volume, and that sales is a good predictor of return volume as expected. In the subsequent models (2) to (5), when we sequentially add time, product type, retailer, product type and retailer effects to model (1), we see that in both training and test sets, R^2 increases and MSE decreases indicating that we achieve not only an increasingly better model fit in the training set, but also a higher prediction accuracy in the test set. We also find that knowing the retailer is more important than knowing the product type when predicting returns, by comparing models (3) and (4). Our largest model (6)– with 13 additional predictors in U compared to model (5)–obtains the lowest $AIC = 9315.068$, the highest Adjusted $R^2 = 0.918$; and if we adopt conventional model selection methods, we may consider model (6) to be the best main effects model for training set. This model also has the best prediction performance in test set.</p>	<p>Мы видим, что самая простая модель (1) дает $R^2 = 0,864$ для обучающего набора и $R^2 = 0,836$ в тестовом наборе для прогнозирования доходности в течение следующих 6 месяцев. Это показывает, что прогнозируемые переменные продажи сами по себе объясняют большую часть изменчивости объема возврата, и что продажи, как и ожидалось, являются хорошим предиктором объема возврата. В последующих моделях (2) - (5), когда мы последовательно добавляем время, тип продукта, розничного продавца, тип продукта и эффекты розничного продавца к модели (1), мы видим, что как в обучающем, так и в тестовом наборе R^2 увеличивается, а MSE уменьшается, указывая что мы достигаем не только все более лучшей модели в учебном наборе, но и более высокой точности прогнозирования в наборе испытаний. Мы также находим, что знание розничного предприятия важнее, чем знание типа продукта при прогнозировании прибыли, путем сравнения моделей (3) и (4). Наша самая большая модель (6) - с 13 дополнительными предикторами в U по сравнению с моделью (5) - имеет самый низкий $AIC = 9315,068$, самый высокий скорректированный $R^2 = 0,918$; и если мы примем традиционные методы выбора модели, мы можем считать модель (6) лучшей моделью основных эффектов для учебного набора. Эта модель также имеет лучшую производительность прогнозирования в тестовом наборе.</p>

<p>Model (6) may have one potential drawback for use in practice, because the required overhead from data collection and analysis for many of the variables in U is substantial. By comparison, our second best model (5) may be preferred by Company A, because it only requires sales, time, product type and retailer information, thus it is much easier to implement by operations managers. For these reasons, we consider models (5) and (6) as the two best main effects models; and to address the aforementioned trade-off between models (5) and (6), we first provide in Table 5 on the next page the coefficient, standard deviation and significance levels for all the main effects models. We notice that two predictor variables—<i>MultiProduct%</i> and <i>LaggedReturns</i>—show up as significant at 0.001 level in model (6).</p>	<p>Модель (6) может иметь один потенциальный недостаток для использования на практике, потому что требуемые накладные расходы от сбора и анализа данных для многих переменных в U являются существенными. Для сравнения, наша вторая лучшая модель (5) может быть предпочтительна для компании А, потому что она требует только информацию о продажах, времени, типе продукта и информации о розничном продавце, поэтому ее гораздо легче внедрить операционным менеджерам. По этим причинам мы рассматриваем модели (5) и (6) как две лучшие модели основных эффектов; и для решения вышеупомянутого компромисса между моделями (5) и (6), мы сначала предоставим в таблице 5 на следующей странице коэффициенты, стандартное отклонение и уровни значимости для всех моделей основных эффектов. Мы замечаем, что две прогнозирующие переменные - <i>MultiProduct%</i> и <i>LaggedReturns</i> - отображаются как значимые на уровне 0,001 в модели (6).</p>
<p>A natural question that arises is whether one should choose one or both of these significant predictors from model (6) to add to model (5) to achieve prediction performance close to model yet without the added overhead for implementation. This question then further invites another question—what is the best model out of all possible models based on the 40 variables in model (6)? Finding the best subset of predictor variables can be onerous due to the potential model space size $2^p - 1$. The best subset selection method also suffers from a lack of stability (Breiman, 1996).</p>	<p>Естественный вопрос, который возникает, заключается в том, следует ли выбирать один или оба из этих значимых предикторов из модели (6) для добавления в модель (5) для достижения эффективности прогнозирования, близкой к модели, но без дополнительных затрат на реализацию. Затем этот вопрос еще раз вызывает другой вопрос - какая модель лучше всего подходит из всех возможных моделей, основанных на 40 переменных в модели (6)? Поиск лучшего подмножества переменных-предикторов может быть обременительным из-за потенциального размера пространства модели $2^p - 1$. Лучший метод выбора подмножества также страдает от отсутствия стабильности (Breiman, 1996).</p>
<p>Table 4 Prediction performance of 6 main effects models trained on 33 periods and tested on 6 periods</p>	<p>Таблица 4 Прогнозирование эффективности 6 моделей основных эффектов, обученных в 33 периодах и проверенных в 6 периодах</p>
<p>Table 5 OLS results for main effects models in the training set. Standard errors are shown in parentheses.</p>	<p>Таблица 5 Результаты OLS для моделей основных эффектов в тренировочном наборе. Стандартные ошибки указаны в скобках.</p>
<p>The model selection issue becomes even more challenging if we want to explore the effect of interaction terms to improve model performance, in which case we have a larger number of predictors in the model, and thus an even larger number of potential models. Even though the model performance is our primary concern, we are also interested in selecting a stable and sparse model for predicting</p>	<p>Проблема выбора модели становится еще более сложной, если мы хотим исследовать влияние терминов взаимодействия для повышения производительности модели, и в этом случае в модели имеется большее количество предикторов и, следовательно, еще большее количество потенциальных моделей. Несмотря на то, что производительность модели является нашей главной задачей, мы также заинтересованы в выборе стабильной и разреженной модели для прогнозирования</p>

returns and to implement it in practice. In the next Section 5 , we discuss multiple machine learning methods to address the aforementioned challenges in model selection and to achieve our goal of obtaining a parsimonious and stable model.	доходов и ее реализации на практике. В следующем разделе 5 мы обсудим несколько методов машинного обучения для решения вышеупомянутых проблем при выборе модели и для достижения нашей цели получения экономной и стабильной модели.
5. Incorporating interaction effects	5 Включение эффектов взаимодействия
The main objective of this section is to explore higher order interaction terms to see if adding such terms to the main effects model helps improve prediction performance. To do this, we first specify a model with a large number of interaction terms in Section 5.1 , which creates a challenge in fitting the traditional OLS model to the data. To overcome this challenge, we introduce LASSO in Section 5.2 as a method to select predictor variables, and present the prediction performance of two models selected by LASSO. Then, in Section 5.3 , we provide an interpretation of the “more sparse” LASSO model and its prediction performance for some high volume retailer product pairs. Finally, in Section 5.4 we employ alternative methods and show a comparison of prediction performance among such methods.	Основная цель этого раздела - изучить термины взаимодействия более высокого порядка, чтобы увидеть, помогает ли добавление таких терминов в модель основных эффектов улучшить производительность прогнозирования. Чтобы сделать это, мы сначала зададим модель с большим количеством терминов взаимодействия в Разделе 5.1, что создает проблему при установке традиционной модели OLS на данные. Чтобы преодолеть эту проблему, мы представляем LASSO в Разделе 5.2 в качестве метода выбора переменных предикторов и представляем эффективность прогнозирования двух моделей, выбранных LASSO. Затем в разделе 5.3 мы даем интерпретацию «более разреженной» модели LASSO и ее прогнозирующих характеристик для некоторых пар продуктов розничной торговли в больших объемах. Наконец, в разделе 5.4 мы используем альтернативные методы и показываем сравнение эффективности прогнозирования среди таких методов.
Table 6 2-way and 3-way interaction terms.	Таблица 6 Условия двустороннего и трехстороннего взаимодействия.
5.1. Model specification	5.1. Спецификация модели
We restrict our attention to the following three categories of higher order interaction terms that may help improve prediction performance of the full main effects in model (6). As in model (6), we use vector notations $\text{Month} = (\text{January}, \dots, \text{December})^T$, $\text{Product} = (\text{Product1}, \dots, \text{Product3})^T$ and $\text{Retailer} = (\text{Retailer1}, \dots, \text{Retailer13})^T$.	Мы ограничиваем наше внимание следующими тремя категориями взаимодействующих элементов более высокого порядка, которые могут помочь улучшить эффективность прогнозирования всех основных эффектов в модели (6). Как и в модели (6), мы используем векторные обозначения $\text{Month} = (\text{январь}, \dots, \text{декабрь})^T$, $\text{Product} = (\text{Product1}, \dots, \text{Product3})^T$ и $\text{Retailer} = (\text{Retailer1}, \dots, \text{Retailer13})^T$.
i. Quadratic effect: We are interested in the effect of Sales^2 because we have shown that sales is a strong predictor of return volume in all the main effects models, and our residual analysis indicated a possible quadratic effect of sales on returns.	Квадратичный эффект: мы заинтересованы в эффекте Sales^2 , потому что мы показали, что продажи являются сильным предиктором объема прибыли во всех моделях основных эффектов, а наш остаточный анализ показал возможное квадратичное влияние продаж на прибыль.
ii. 2-way interaction effects: The 2-way interactions of predictor variables may help predict returns more accurately. For example, we can ask: Does the effect of sales volume on	Эффекты двустороннего взаимодействия. Двустороннее взаимодействие переменных-предикторов может помочь более точно предсказать результаты. Например, мы можем спросить: варьируется ли влияние объема продаж на объем возврата в зависимости от продавца? Термин

<p>return volume vary depending on the retailer? The interaction term Sales · Retailer addresses this question. For another example, historical returns are shown to be significant and have positive correlation with returns (see Table 5). However, the effect of historical returns could increase or decrease over time and this would get discovered by including the interaction term LaggedReturns · Year.</p>	<p>взаимодействия Sales · Retailer решает этот вопрос. В другом примере показано, что исторические возвраты являются значимыми и имеют положительную корреляцию с доходами (см. Таблицу 5). Однако эффект исторических возвратов может увеличиваться или уменьшаться с течением времени, и это можно обнаружить, включив термин взаимодействия LaggedReturns · Year.</p>
<p>The question that arises is, which interaction terms should we consider? We begin with our original set of predictors Sales, Year, Month, Product, Retailer used in model (5), and additionally consider the two predictors found to be significant at the 0.001 level in model (6)—MultiProduct%, LaggedReturns; then, we utilize all the 2-way interactions among these predictor variables to create a new set of predictor variables shown in Table 6. Because Month, Product and Retailer are each a vector of 12, 3 and 13 dummy variables, we can see, for example, Product Retailer (here denotes Cartesian product) creates 39 interaction terms. Vector V includes all the 2-way interaction terms in Table 6, and one can easily verify there are 337 predictor variables in V.</p>	<p>Возникает вопрос: какие условия взаимодействия нам следует рассмотреть? Мы начнем с нашего исходного набора предикторов Sales, Year, Month, Product, Retailer, используемого в модели (5), и дополнительно рассмотрим два предиктора, которые оказались значимыми на уровне 0,001 в модели (6) - MultiProduct%, LaggedReturns; затем мы используем все двусторонние взаимодействия между этими переменными предикторами, чтобы создать новый набор переменных предикторов, показанных в таблице 6. Поскольку Month, Product и Retailer являются вектором из 12, 3 и 13 фиктивных переменных, мы можем видеть, Например, Product Retailer (здесь обозначается декартово произведение) создает 39 терминов взаимодействия. Вектор V включает в себя все члены двустороннего взаимодействия из Таблицы 6, и можно легко проверить, что в V. есть 337 переменных-предикторов.</p>
<p>3-way interaction effects. We use 3-way interactions to achieve higher accuracy in returns prediction, and also to address questions such as: Is return rate of any particular retailer predicted to increase over time? Because return rate equals returns divided by sales, the interaction term Sales · Retailer · Year addresses this question. We consider potentially meaningful 3 way interactions among the above-mentioned key predictors— Sales, Year, Month, Product, Retailer, MultiProduct%, LaggedReturns. Vector W includes all such 3-way interaction terms shown in Table 6, and one can verify that there are 1336 predictor variables in W.</p>	<p>3-х сторонние эффекты взаимодействия. Мы используем трехсторонние взаимодействия для достижения более высокой точности в прогнозировании доходности, а также для решения таких вопросов, как: Прогнозируется ли рост доходности какого-либо конкретного розничного продавца с течением времени? Поскольку норма возврата равна прибыли, разделенной на продажи, этот вопрос рассматривается в терминах взаимодействия Продажи · Розничный продавец · Год. Мы рассматриваем потенциально значимые трехсторонние взаимодействия между вышеупомянутыми ключевыми предикторами - Продажи, Год, Месяц, Продукт, Ритейлер, MultiProduct%, LaggedReturns. Вектор W включает все такие термины трехстороннего взаимодействия, показанные в таблице 6, и можно проверить, что в W имеется 1336 предикторных переменных.</p>
<p>Finally, we specify a new model in Eq. (7) below, which includes all the interactions discussed above.</p>	<p>Наконец, мы указываем новую модель в формуле. (7) ниже, который включает в себя все взаимодействия, обсужденные выше.</p>

$$\begin{aligned}
\text{Returns}_{itj} = & \beta_0 + \beta_1 \text{Sales}_{itj} + \beta_2 \text{Year}_t + \beta_3 \text{Month}_t \\
& + \beta_4^i \text{Product}_t + \beta_5^j \text{Retailer}_j + \beta_6 \bar{U} \\
& + \beta_7 \text{Sales}_{itj}^2 + \beta_8 \bar{V} + \beta_9 \bar{W} + \epsilon_{itj}
\end{aligned}
\tag{7}$$

This model has a total of $p = 1,717$ predictors, which comprise of 43 main effects, 1 quadratic term, 337 2-way interactions, and 1,336 3-way interactions. Recall that in the main effects model (6) we manually dropped 3 binary dummy variables December, Product3, Retailer13 to remove redundancy, however we no longer do this for model (7) and will instead let LASSO select predictor variables for us. For these reasons, we have 43 main effects now, increased from 40 in model (6). Recall that we have a training set data size $N = 1,140$ and thus we have $p > N$, and we are effectively in a high-dimensional setting. For this reason, some type of penalty is needed to reduce the number of variables used in the model, and most importantly to perform a proper model selection and address problems such as spurious correlation, noise accumulation, and non-uniqueness of the linear model solution. To do these, in Sections 5.2 and 5.3 we employ LASSO, which enjoys an appealing theoretical property—the oracle inequalities (Bickel, Ritov, & Tsybakov, 2009), meaning that under some conditions LASSO can recover the true model with high likelihood, given that the true model is sparse. Since most of the predictor variables in (7) are interaction terms, and not all of them are likely to play an important role in predicting returns, we expect sparse predictive models to fit our data well. To simplify the exposition, henceforth we use $X_l = (x_{l1}, \dots, x_{lp})^T$ to denote all the predictor variables, and y_l to denote the response variable in (7) for $l = 1, \dots, N$.

Эта модель имеет в общей сложности $p = 1717$ предикторов, которые включают в себя 43 основных эффекта, 1 квадратичный член, 337 двусторонних взаимодействий и 1336 трехсторонних взаимодействий. Напомним, что в модели основных эффектов (6) мы вручную отбросили 3 двоичные фиктивные переменные декабрь, Product3, Retailer13, чтобы удалить избыточность, однако мы больше не будем делать это для модели (7) и вместо этого позволим LASSO выбирать переменные-предикторы для нас. По этим причинам у нас сейчас 43 основных эффекта, увеличенных с 40 в модели (6). Напомним, что у нас есть размер данных обучающего набора $N = 1,140$, и, таким образом, мы имеем $p > N$, и мы эффективно находимся в многомерной установке. По этой причине необходим некоторый тип штрафа для уменьшения числа переменных, используемых в модели, и, что наиболее важно, для правильного выбора модели и решения таких проблем, как ложная корреляция, накопление шума и неединственность решения линейной модели. Для этого в разделах 5.2 и 5.3 мы используем LASSO, который обладает привлекательным теоретическим свойством - неравенствами оракула (Bickel, Ritov, & Tsybakov, 2009), что означает, что в некоторых условиях LASSO может восстановить истинную модель с высокой вероятностью, учитывая что истинная модель редка. Поскольку большинство переменных-предикторов в (7) являются терминами взаимодействия, и не все из них, вероятно, будут играть важную роль в прогнозировании доходности, мы ожидаем, что редкие прогностические модели хорошо подойдут нашим данным. Чтобы упростить изложение, впредь мы будем использовать $X_l = (x_{l1}, \dots, x_{lp})^T$ для обозначения всех переменных-предикторов и y_l для обозначения ответной переменной в (7) для $l = 1, \dots, N$.

Table 7			
Prediction performance of 2 LASSO models.			
		Model 7a	Model 7b
Training set	Number of predictors provided	1,717	1,717
	Data size	1,140	1,140
	Tuning parameter type	λ_{min}	λ_{1se}
	Tuning parameter value	1.809	4.378
	Number of predictors selected	19	9
	R^2	0.958	0.934
		MSE	103.754
Test set	Data size	220	220
	R^2	0.928	0.930
	MSE	177.815	171.824

Table 7 Prediction performance of 2 LASSO models.

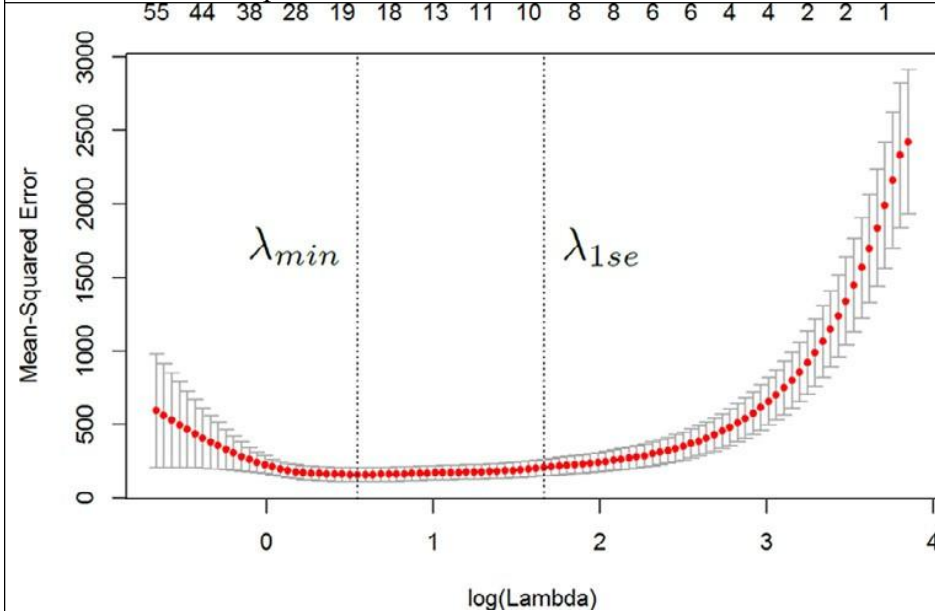


Fig. 4. Selecting λ based on 10-fold CV.

5.2. LASSO analyses

In this subsection, we focus on LASSO which solves the following convex optimization problem (8) to obtain the LASSO estimate

Таблица 7 Прогноз производительности 2 моделей LASSO.

Рис. 4. Выбор λ на основе 10-кратного CV.

5.2. LASSO анализы

В этом пункте мы сосредоточимся на LASSO, который решает следующую задачу выпуклой оптимизации (8), чтобы получить оценку LASSO

<p>$\hat{\theta}^{Lasso}$ for data (\mathbf{X}^l, y_l), $l = 1, \dots, N$.</p> $\hat{\theta}^{Lasso} = \arg \min_{(\theta_0, \theta) \in \mathbb{R} \times \mathbb{R}^p} \left[\frac{1}{2N} \sum_{l=1}^N \left(y_l - \theta_0 - \sum_{m=1}^p x_{lm} \theta_m \right)^2 + \lambda \sum_{m=1}^p \theta_m \right]. \quad (8)$	
<p>Here λ is a tuning parameter, which assigns an appropriate level of penalty, and continuously shrinks coefficients, and can force some to go to exactly zero to obtain a sparse model (see §3 in Hastie et al., 2009). Therefore, it is vital to obtain good values of λ for successful model selection, and in our study we utilize the k-fold cross-validation (CV) method. Generally, $k = 5$ or 10 works well in practice and for our study, we use the option $k = 10$ which is widely used in recent statistics literature. The reason we use CV instead of other measures such as AIC for model selection is to estimate the prediction error in the test set and compare different models (see Section 5 in James et al., 2017)</p>	<p>Здесь λ - это параметр настройки, который назначает соответствующий уровень штрафа и непрерывно сжимает коэффициенты, и может заставить некоторых стремиться к точному нулю, чтобы получить разреженную модель (см. §3 в Hastie et al., 2009). Следовательно, жизненно важно получить хорошие значения λ для успешного выбора модели, и в нашем исследовании мы используем метод перекрестной проверки (CV) в k-кратном выражении. Как правило, $k = 5$ или 10 хорошо работает на практике, и для нашего исследования мы используем параметр $k = 10$, который широко используется в недавней статистической литературе. Причина, по которой мы используем CV вместо других показателей, таких как AIC, для выбора модели, заключается в том, чтобы оценить ошибку прогнозирования в наборе тестов и сравнить различные модели (см. Раздел 5 в James et al., 2017).</p>
<p>Notice that without the L penalty term $\lambda \sum_{m=1}^p \theta_m$, the LASSO problem (8) reduces to a standard OLS problem. With the the $L1$ penalty term, the problem (8) is still a convex optimization problem, and this optimization problem can be solved efficiently using cyclical coordinate descent algorithms. It is important that all the predictors be standarized to have mean 0 and standard deviation 1, so that the different scales of variables do not impact the optimization problem. To implement LASSO with 10-fold CV, we utilized Rpackage glmnet from Friedman, Hastie, and Tibshirani (2010).</p>	<p>Обратите внимание, что без L-штрафного члена λ задача LASSO (8) сводится к стандартной задаче OLS. С штрафным слагаемым $L1$ задача (8) остается выпуклой оптимизационной задачей, и эта оптимизационная задача может быть эффективно решена с использованием алгоритмов циклического спуска координат. Важно, чтобы все предикторы были стандартизированы, чтобы иметь среднее значение 0 и стандартное отклонение 1, чтобы различные шкалы переменных не влияли на проблему оптимизации. Для реализации LASSO с 10-кратным CV мы использовали Rpackage glmnet от Friedman, Hastie и Tibshirani (2010).</p>
<p>We show a plot of cross-validation fit with respect to different values of λ in Fig. 4, where the two dotted vertical lines correspond to model fit for two tuning parameters: λ_{min} (left) and λ_{1se} (right). Here λ_{min} denotes the tuning parameter value that achieves the minimum mean CV error, whereas λ_{1se} represents a larger tuning parameter which leads to the smallest model such that mean cross-validated error is within one standard error of the minimum.</p>	<p>Мы показываем график перекрестной проверки соответствия относительно различных значений λ на рис. 4, где две пунктирные вертикальные линии соответствуют образцу соответствия для двух параметров настройки: λ_{min} (слева) и λ_{1se} (справа). Здесь λ_{min} обозначает значение параметра настройки, которое достигает минимальной средней ошибки CV, тогда как λ_{1se} представляет больший параметр настройки, который приводит к наименьшей модели,</p>

<p>In our study, we have $\lambda_{\min} = 1.809$ and $\lambda_{1se} = 4.378$ and, using these penalty values LASSO gives reduced models 7a and 7b in Table 7, which have only 19 and 9 predictors selected respectively out of a total of 1,717 predictors in Eq. (7). We find that the LASSO models 7a and 7b outperform the two best main effects models (5), (6) with respect to MSE and R² in both training and test sets; and the “more sparse” model 7b exhibits a more robust performance with a smaller gap in MSE between training and test sets (161.908 vs. 171.824), when compared to the “less sparse” model 7a` (103.754 vs. 177.815.) Hence, we focus on the “more sparse” model 7b referenced as the LASSO model hereafter.</p>	<p>такой, что средняя перекрестно проверенная ошибка находится в пределах одной стандартной ошибки минимума. В нашем исследовании мы имеем $\lambda_{\min} = 1.809$ и $\lambda_{1se} = 4.378$, и, используя эти штрафные значения, LASSO дает уменьшенные модели 7a и 7b в таблице 7, в которых только 19 и 9 предикторов выбраны соответственно из общего количества 1717 предикторов в формуле. (7). Мы считаем, что модели LASSO 7a и 7b превосходят две лучшие модели основных эффектов (5), (6) по отношению к MSE и R² как в учебных, так и в тестовых наборах; и «более разреженная» модель 7b демонстрирует более устойчивую производительность с меньшим разрывом в MSE между обучающими и тестовыми наборами (161,908 против 171,824) по сравнению с «менее разреженной» моделью 7a` (103,754 против 177,815). Следовательно, в дальнейшем мы сосредоточимся на «более разреженной» модели 7b, на которую ссылаются как на модель LASSO.</p>
<p>In Table 8 , we show the prediction performance of model (5) , model (6) , and the LASSO model (all trained on periods 1 to 33) in each test period from 34 to 39, and also over the entire 6 test periods. One can observe that if we apply these predictive models to the entire 6 test periods (34 to 39), then the LASSO model reduces the MSE of the baseline main effects model (5) by $257.709 - 171.824 = 85.885 = 33.3\%$, and the MSE of the full main effects model $257.709 - 171.824$ (6) by $189.386 - 171.824 = 17.562 = 9.3\%$. However, if we apply such predictive 189.386 models only in the first test period 34, then LASSO model has a more pronounced advantage in prediction performance by reducing the MSE of model (5) by $206.206 - 126.230 = 80.076 = 63.4\%$, and the MSE $460.126 - 126.230$ of model (6) by $333.896 = 26.4\%$.</p>	<p>В Таблице 8 мы показываем эффективность прогнозирования модели (5), модели (6) и модели LASSO (все обучаются в периоды с 1 по 33) в каждом тестовом периоде с 34 до 39, а также в течение всех 6 тестовых периодов , Можно заметить, что если мы применим эти прогностические модели ко всем 6 тестовым периодам (от 34 до 39), то модель LASSO уменьшает MSE базовой модели основных эффектов (5) на $257.709 - 171.824 = 85.885 = 33.3\%$, и MSE полной модели основных эффектов $257.709 - 171.824$ (6) к $189.386 - 171.824 = 17.562 = 9.3\%$ Однако если применить такой прогноз 189.386 моделей только в первом тестовом периоде 34, тогда модель LASSO имеет более выраженное преимущество в эффективности прогнозирования благодаря $206.206 - 126.230 = 80.076 = 63.4\%$ с MSE модели (5) к $460.126 - 126.230 = 333.896 = 26.4\%$ модели (6) по = 55.6%</p>
<p>To understand the robustness of the above finding, we conducted further analyses by using the data from periods 1 to $p - 1$ to build the model, and then tested the model in period p for each prediction period $p \in \{35, 36, 37, 38, 39\}$. The results are provided in Table 9 . The LASSO model reduces the MSE of $-126.085 - 56.03$ model (5) by $144.358 - 126.085 = 18.273 = 14.5\%$ to $206.206 - 56.03 = 150.176 = 72.9\%$ (average $144.358 - 126.085$ age 43.0%) and reduces the MSE of model (6) by $126.424 - 126.230 = 0.194 = 0.15\%$ to $126.126 - 56.03 = 70.096 = 55.6\%$ (average 29.7%).</p>	<p>Чтобы понять надежность вышеуказанного вывода, мы провели дополнительный анализ, используя данные за периоды с 1 по $p - 1$, для построения модели, а затем проверили модель в периоде p для каждого периода прогнозирования $p \in \{35, 36, 37, 38, 39\}$. Результаты представлены в таблице 9. Модель LASSO уменьшает MSE $-126.085 - 56.03$ модель (5) на $144.358 - 126.085 = 18.273 = 14.5\%$ От 7% до $206.206 - 56.03 = 150.176 = 72.9\%$ (average $144.358 - 126.085$ возраст 43,0%) и снижает MSE модели (6) на $126.424 - 126.230 = 0.194 = 0.15\%$ От 3% до $126.126 - 56.03 = 70.096 = 55.6\%$ (в среднем 29,7%).</p>
<p>Because forecasting models are often implemented on a rolling horizon basis,</p>	<p>Поскольку модели прогнозирования часто реализуются на основе скользящего горизонта, может быть целесообразно обновлять модель</p>

<p>it may be advisable to update the LASSO model every month and use it for the first test period for a more accurate prediction. Updating the LASSO model is easier than one might expect. The <code>afrontmentioend</code> R-package <code>glmnet</code> that implements LASSO is well-documented. All we need to change are the variable names and the data set names. It took 10 seconds to train and test the LASSO model in our desktop computer (Windows 10, Intel Core i7-8770 CPU, 8Gb RAM 26 6 6MHz, 1Tb 970 Samsung Evo SSD). So, we believe model updating is not going to be onerous from a cost or time perspective.</p>	<p>LASSO каждый месяц и использовать ее в течение первого тестового периода для более точного прогнозирования. Обновление модели LASSO проще, чем можно было ожидать. Передовой пакет <code>glmnet</code>, реализующий LASSO, хорошо документирован. Все, что нам нужно изменить - это имена переменных и имена наборов данных. На обучение и тестирование модели LASSO на настольном компьютере потребовалось 10 секунд (Windows 10, процессор Intel Core i7-8770, 8 ГБ ОЗУ 26 6 6 МГц, 1 ТБ 970 Samsung Evo SSD). Таким образом, мы считаем, что обновление модели не будет обременительным с точки зрения затрат или времени.</p>
<p>5.3. Selected predictors and their performance</p>	<p>5.3. Выбранные предикторы и их эффективность</p>
<p>We first provide the 9 predictor variables and their coefficients in the LASSO model in Table 10 . These can be interpreted as the strongest effects identified by LASSO to predict returns, and have been shown to achieve robust prediction performance both in the training and test data. It is noteworthy that the LASSO model chosen is very easy to implement in practice to predict future returns, as one only needs to compute Sales , LaggedReturns –which can be easily done in a spreadsheet–and use them along with Year and two retailer dummy variables Retailer 1 , Retailer 5 .</p>	<p>Сначала мы представим 9 переменных предикторов и их коэффициенты в модели LASSO в таблице 10. Они могут быть интерпретированы как самые сильные эффекты, определенные LASSO для прогнозирования отдачи, и было показано, что они обеспечивают надежную эффективность прогнозирования как в обучающих, так и в тестовых данных. Следует отметить, что выбранную модель LASSO очень легко реализовать на практике для прогнозирования будущих доходов, так как нужно только рассчитать Sales, LaggedReturns - что можно легко сделать в электронной таблице - и использовать их вместе с Year и двумя фиктивными переменными розничного предприятия. Ритейлер 1, Ритейлер 5.</p>
<p>We first note that the coefficients of all 9 predictor variables are positive, and also that only two main effects Sales , LaggedReturns are selected by LASSO, which is quite different from main effects models (see Table 5 .) Also notice that the LASSO coefficient of Sales is 0.0313, reduced from 0.099 in model (6) shown in Table 5 . A key reason is that the effect of Sales is broken down and absorbed into quadratic term Sales 2 , as well as 2-way and 3-way interaction terms which include Sales . As the coefficients of Sales and Sales 2 are both positive, the LASSO model shows that the predicted returns are convex and increasing in sales. This suggests that the likelihood of returns increases for each incremental unit sold. One possible explanation is that production process capacity limitations are pushed as sales increase (for example, sewers are pressured to work faster) leading to more errors and thus returns.</p>	<p>Сначала отметим, что коэффициенты всех 9 переменных-предикторов положительны, а также, что LASSO выбирает только два основных эффекта Sales, LaggedReturns, что довольно сильно отличается от моделей основных эффектов (см. Таблицу 5). Также обратите внимание, что коэффициент LASSO Объем продаж составляет 0,0313, по сравнению с 0,099 в модели (6), показанной в таблице 5. Основная причина заключается в том, что эффект «Продажи» разбит на два квадратичных термина «Продажи 2», а также термины двустороннего и трехстороннего взаимодействия, включающие Продажи. Поскольку коэффициенты «Продажи» и «Продажи 2» являются положительными, модель LASSO показывает, что прогнозируемая доходность является выпуклой и увеличивается в продажах. Это говорит о том, что вероятность возврата увеличивается с каждой проданной проданной единицей. Одним из возможных объяснений является то, что ограничения производительности производственного процесса</p>

<p>returns, that is, we can expect higher return rates year over year. This may be because as competition increases, retailers become more lax in accepting returns to please customers and keep them from going to competitors (e.g., see Montaldo, 2019). Recall that only Retailer 1 (auto specialty store with a strong bargaining power) and Retailer 5 (warehouse club) carry liberal return policies, and the two interaction terms Sales \times Retailer 1 and Sales \times Retailer 5 show that for the same sales, these two retailers contribute more to returns than other retailers. Sales \times LaggedReturns indicates that when historical returns are higher, sales generates higher returns, i.e. higher return rate. Also we see from LaggedReturns \times Year that the impact of historical returns increases over time.</p>	<p>есть мы можем ожидать более высокие показатели доходности из года в год. Это может быть связано с тем, что по мере роста конкуренции ритейлеры становятся все более слабыми в получении прибыли, чтобы угодить покупателям и не дать им идти к конкурентам (например, см. Montaldo, 2019). Напомним, что только розничный торговец 1 (специализированный магазин с сильной переговорной силой) и розничный торговец 5 (складской клуб) несут либеральную политику возврата, а два условия взаимодействия Sales \times Retailer 1 и Sales \times Retailer 5 показывают, что при одинаковых продажах эти два Ритейлеры вносят больший вклад в доход, чем другие. Sales \times LaggedReturns указывает на то, что когда историческая доходность выше, продажи генерируют более высокую доходность, то есть более высокую норму доходности. Также мы видим из LaggedReturns \times Year, что влияние исторических возвратов со временем увеличивается.</p>
<p>We now pay attention to the only 3-way interaction terms selected by the LASSO model: Sales \times Year \times Retailer 1 . This shows that return rate of retailer 1, which carries a liberal return policy, has increased over the years more than for others. In a way, sales through this retailer are becoming “more problematic”.</p>	<p>Теперь обратим внимание на единственные условия трехстороннего взаимодействия, выбранные моделью LASSO: Продажи \times Год \times Розничный продавец 1. Это показывает, что коэффициент возврата розничного продавца 1, который проводит либеральную политику возврата, увеличился за эти годы больше, чем для других. В некотором смысле, продажи через этого ритейлера становятся «более проблемными».</p>
<p>We end this subsection by comparing the prediction performance for the LASSO model vs. models (5), (6). We do this for the retailer and product pairs that have higher observed returns, because that is where all three models show the best prediction performance. Fig. 5 shows how well the three models predict returns for the future 6 periods for top 6 retailer-product pairs which account for 68.3% of total return volume. When compared to models (5), (6) , we find that the LASSO model predicts returns closer to the observed returns when they are high enough. It is also interesting to observe that for Retailer 6 , Product 1 the LASSO model predicts returns substantially better than models (5), (6) for the first test period, as is consistent with our discussion in the last paragraph of Section 5.2 . For low volume returns, the three models overall do not provide nearly as good a prediction performance. This is expected because such low volume returns have higher coefficient of variation. For completeness, we provide</p>	<p>Мы заканчиваем этот подраздел, сравнивая эффективность прогнозирования для модели LASSO с моделями (5), (6). Мы делаем это для пар розничной торговли и продуктов, которые имеют более высокую наблюдаемую доходность, потому что именно здесь все три модели демонстрируют наилучшую эффективность прогнозирования. На рис. 5 показано, насколько хорошо три модели предсказывают доходность в течение следующих 6 периодов для 6 ведущих пар «продавец-продукт», на которые приходится 68,3% от общего объема возврата. По сравнению с моделями (5), (6), мы находим, что модель LASSO предсказывает доход ближе к наблюдаемому доходу, когда он достаточно высок. Интересно также отметить, что для Продавца 6, Продукта 1, модель LASSO предсказывает возвраты значительно лучше, чем модели (5), (6) для первого испытательного периода, что согласуется с нашим обсуждением в последнем абзаце Раздела 5.2. Для малых объемов возврата три модели в целом не обеспечивают столь же хороших результатов прогнозирования. Это ожидается, потому что</p>

prediction results for all retailer product pairs for future 6 months in Fig. A1 in Online Appendix A5.	такие низкие объемные доходы имеют более высокий коэффициент вариации. Для полноты мы предоставляем результаты прогнозирования для всех пар продуктов розничной торговли на будущие 6 месяцев на рис. A1 в онлайн-приложении A5.
5.4. Alternative methods to LASSO	5.4. Альтернативные методы LASSO
Our goal in this subsection is to fit three alternative highdimensional statistical models, and two tree-based statistical machine learning methods to our data and compare their prediction performance against LASSO. The results are provided in Table 11 .	Наша цель в этом подразделе состоит в том, чтобы подогнать к нашим данным три альтернативных высокомерных статистических модели и два метода статистического машинного обучения на основе дерева и сравнить их результаты прогнозирования с LASSO. Результаты представлены в таблице 11.
5.4.1. High-dimensional machine learning methods	5.4.1. Методы многомерного машинного обучения
In this subsection, we discuss three alternative highdimensional statistical models. In our setting, they all select 9 predictors which we discuss in detail below.	В этом подразделе мы обсуждаем три альтернативных высокомерных статистических моделей. В нашем случае все они выбирают 9 предикторов, которые мы подробно обсудим ниже.
The first model we consider is called LARS-OLS hybrid (Efron et al., 2004) or OLS post-LASSO (Belloni & Chernozhukov, 2013). This model involves two stages: in the first stage, it utilizes LASSO to select predictor variables only; in the second stage, it uses OLS to estimate the coefficients of the selected predictor variables—also known as debiasing. By construction, this method uses exactly the same 9 predictor variables as LASSO (see Table 10). The rationale for this approach is to reduce bias in the coefficients estimated by LASSO, because LASSO tends to shrink the nonzero coefficients towards zero compared to OLS. Similar to Efron et al. (2004) , in our study we find that the LARS-OLS hybrid method achieves a smaller MSE in the training data set, compared to the LASSO model; however, it results in a large MSE in the test set.	Первая рассматриваемая нами модель называется гибридом LARS-OLS (Efron et al., 2004) или OLS post-LASSO (Belloni & Chernozhukov, 2013). Эта модель состоит из двух этапов: на первом этапе она использует LASSO для выбора только предикторных переменных; на втором этапе он использует OLS для оценки коэффициентов выбранных предикторных переменных, также известных как сглаживание. По построению этот метод использует те же 9 переменных предиктора, что и LASSO (см. Таблицу 10). Обоснование этого подхода заключается в уменьшении смещения в коэффициентах, оцениваемых с помощью LASSO, поскольку LASSO стремится уменьшить ненулевые коэффициенты до нуля по сравнению с OLS. Подобно Efron et al. (2004), в нашем исследовании мы находим, что гибридный метод LARS-OLS обеспечивает меньшую MSE в наборе обучающих данных по сравнению с моделью LASSO; однако это приводит к большому MSE в тестовом наборе.
Recall that LASSO is a convex regularization method with L 1 penalty. Next, we consider a concave regularization method— Smoothly Clipped Absolute Deviation (SCAD), which yields nearly unbiased estimators (Fan & Li, 2001). Both LASSO and SCAD solve the following penalized least-squares problems to achieve sparse solutions:	Напомним, что LASSO - это метод выпуклой регуляризации со штрафом L 1. Далее мы рассмотрим вогнутый метод регуляризации - Абсолютное отклонение с гладкой подрезкой (SCAD), которое дает почти несмещенные оценки (Fan & Li, 2001). И LASSO, и SCAD решают следующие штрафные задачи наименьших квадратов для достижения разреженных решений:

$\min_{(\theta_0, \boldsymbol{\theta}) \in \mathbb{R} \times \mathbb{R}^p} \left[\frac{1}{2N} \ \mathbf{y} - \theta_0 \mathbf{1} - \mathbf{X}\boldsymbol{\theta}\ _2^2 + \ p_\lambda(\boldsymbol{\theta})\ _1 \right] \quad (9)$	
<p>where we use compact notations for data (X, y), where $X \in \mathbb{R}^{N \times p}$ is a design matrix, $y \in \mathbb{R}^N$ is a response vector, and $p_\lambda(\boldsymbol{\theta}) = p_\lambda(\boldsymbol{\theta}) = (p_\lambda(\theta_1), p_\lambda(\theta_2), \dots, p_\lambda(\theta_p))^\top$ is a penalty for a vector of regression coefficients. Here, the penalty function $p_\lambda(t)$ is defined on $t \in [0, \infty)$ indexed by possibly more than one tuning parameter in λ. It can be observed that LASSO is a special case of (9) with one tuning parameter $\lambda \geq 0$ and a penalty function $p_\lambda(t) := p_\lambda(t) = \lambda t$. On the other hand, SCAD utilizes two tuning parameters $\lambda \geq 0, \gamma > 2$ and a penalty function defined as $p_{\lambda, \gamma}(t) = \lambda I(t \leq \lambda) + (\gamma - 1)\lambda + I(t > \lambda)$, which leads to</p>	<p>где мы используем компактные обозначения для данных (X, y), где $X \in \mathbb{R}^{N \times p}$ - расчетная матрица, $y \in \mathbb{R}^N$ - вектор отклика, а $p_\lambda(\boldsymbol{\theta}) = p_\lambda(\boldsymbol{\theta}) = (p_\lambda(\theta_1), p_\lambda(\theta_2), \dots, p_\lambda(\theta_p))^\top$ - штраф за вектор коэффициентов регрессии. Здесь штрафная функция $p_\lambda(t)$ определена для $t \in [0, \infty)$, проиндексированной возможно более чем одним параметром настройки в λ. Можно заметить, что LASSO является частным случаем (9) с одним параметром настройки $\lambda \geq 0$ и штрафной функцией $p_\lambda(t) := p_\lambda(t) = \lambda t$. С другой стороны, SCAD использует два параметра настройки $\lambda \geq 0, \gamma > 2$ и штрафную функцию, определяемую как $p_{\lambda, \gamma}(t) = \lambda I(t \leq \lambda) + (\gamma - 1)\lambda + I(t > \lambda)$, что приводит к</p>
$p_\lambda(t) := p_{\lambda, \gamma}(t) = \left[\lambda t - \frac{(t - \lambda)^2}{2(\gamma - 1)} I(t > \lambda) \right] I(t \leq \lambda \gamma) + \frac{1}{2} \lambda^2 (\gamma + 1) I(t > \lambda \gamma).$	
<p>Fig. 5. Observed vs. predicted returns for some retailer product pairs (see colors online).</p>	<p>Рис. 5. Наблюдаемая и прогнозируемая доходность для некоторых пар продуктов розничной торговли (см. Цвета в Интернете).</p>

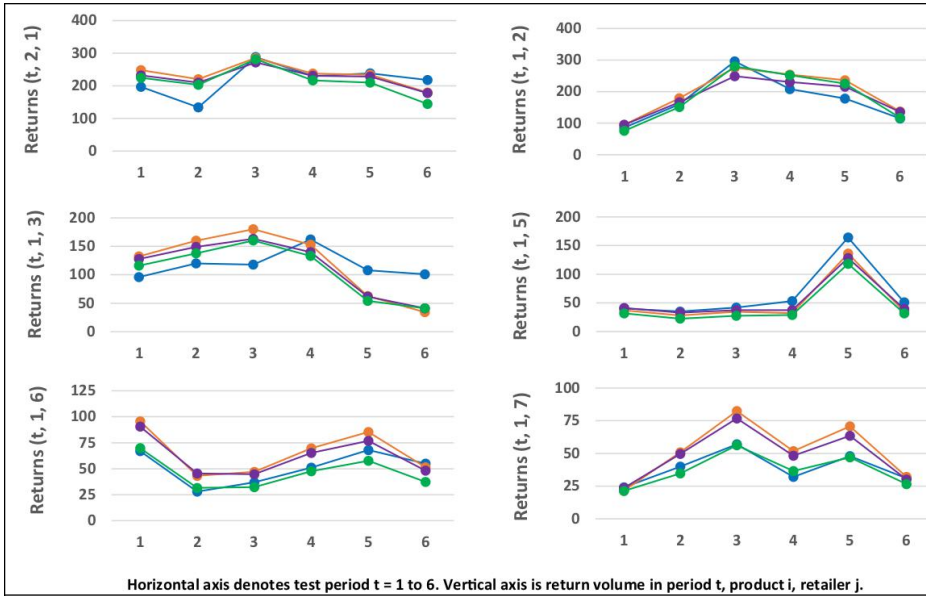


Table 11 Prediction performance of alternative methods against LASSO.

		LASSO	LARS-OLS hybrid	SCAD	Elastic Net ^a	Random Forest	Gradient Boosting
Training set	Data size	1140	1140	1140	1140	1140	1140
	R^2	0.934	0.961	0.947	0.934	0.990	0.996
	MSE	161.908	94.888	129.674	161.908	25.036	10.606
Test set	Data size	220	220	220	220	220	220
	R^2	0.930	0.912	0.874	0.930	0.878	0.882
	MSE	171.824	216.610	308.920	171.824	298.208	290.564

^a Note: Optimal Elastic Net is reduced to LASSO.

Notice that $p_{\lambda, \gamma}(t)$ is concave in t , and as $\gamma \rightarrow \infty$, the penalty functions of SCAD and LASSO coincide for $t > 0$. To solve (9) for SCAD, one may search for the optimal values for λ, γ over twodimensional space, however it is shown that $\gamma \approx 3.7$ is shown to be a robust choice (see §2.1 in Fan & Li, 2001). In our study, we considered $\gamma \in [3.5, 4.0]$ in R-package `ncvreg` (Breheny & Huang, 2011) with 10-fold CV, and find that the choice of γ has little impact on MSE in the training and test sets, and we choose $\gamma = 3.9$ which leads to 9 predictor variables, which is not surprising (see a simulated example comparing SCAD against LASSO in Table 3 in Fan, Feng, & Wu, 2009). Like LARSOLS hybrid, SCAD outperforms LASSO in the training data; how-ever, LASSO still achieves smaller MSE in the test set in our study.

A notable generalization of LASSO is the Elastic Net (Zou & Hastie, 2005)

Таблица 11 Прогноз производительности альтернативных методов против LASSO.

Обратите внимание, что $p_{\lambda, \gamma}(t)$ вогнута по t , и при $\gamma \rightarrow \infty$ штрафные функции SCAD и LASSO совпадают при $t > 0$. Чтобы решить (9) для SCAD, можно искать оптимальные значения для λ, γ над двумерным пространством, однако показано, что $\gamma \approx 3.7$ - надежный выбор (см. § 2.1 в Fan & Li, 2001). В нашем исследовании мы рассмотрели $\gamma \in [3.5, 4.0]$ в R-пакете `ncvreg` (Brey & Huang, 2011) с 10-кратным CV и обнаружили, что выбор γ мало влияет на MSE при обучении и тестировании. множества, и мы выбираем $\gamma = 3.9$, что приводит к 9 переменным предикторам, что неудивительно (см. Смоделированный пример сравнения SCAD с LASSO в Таблице 3 в Fan, Feng & Wu, 2009). Как и гибрид LARSOLS, SCAD превосходит LASSO в данных обучения; тем не менее, LASSO все еще достигает меньшего MSE в тестовом наборе в нашем исследовании.

Заметным обобщением LASSO является Elastic Net (Zou and Hastie,

which makes a compromise between the L 1 (LASSO) and L 2 (Ridge) penalties, and solves the following optimization problem:	2005), который делает компромисс между штрафами L 1 (LASSO) и L 2 (Ridge) и решает следующую задачу оптимизации:
$\min_{(\theta_0, \boldsymbol{\theta}) \in \mathbb{R} \times \mathbb{R}^p} \left[\frac{1}{2N} \ \mathbf{y} - \theta_0 \mathbf{1} - \mathbf{X}\boldsymbol{\theta}\ _2^2 + \lambda \left(\alpha \ \boldsymbol{\theta}\ _1 + \frac{(1 - \alpha)}{2} \ \boldsymbol{\theta}\ _2^2 \right) \right] \quad (10)$	
<p>Notice that (10) involves two tuning parameters $\alpha \in [0, 1]$ and $\lambda \geq 0$, and it is reduced to LASSO when $\alpha = 1$, and to Ridge regression when $\alpha = 0$. In some simulation studies (e.g., Table 1, Table 2 in Zou & Hastie, 2005), Elastic Net is shown to outperform both LASSO and Ridge. To implement Elastic Net for our study, for a range of α values between 0 and 1, we utilize the R-package glmnet with 10-fold CV to derive predictive models with respect to λ 1 se as we do for LASSO. We find that $\alpha = 1$ leads to a model with the minimum MSE in the training set, therefore conclude that the optimal Elastic Net is reduced to LASSO, thus also selects 9 predictors.</p>	<p>Обратите внимание, что (10) включает в себя два параметра настройки $\alpha \in [0, 1]$ и $\lambda \geq 0$ и сводится к LASSO, когда $\alpha = 1$, и к регрессии Риджа, когда $\alpha = 0$. В некоторых исследованиях по моделированию (например, таблица 1, таблица 2 в Zou & Hastie, 2005) показано, что Elastic Net превосходит как LASSO, так и Ridge. Чтобы реализовать Elastic Net для нашего исследования, для диапазона значений α от 0 до 1 мы используем glmnet R-пакета с 10-кратным CV для получения прогнозных моделей относительно λ 1 se, как мы делаем для LASSO. Мы находим, что $\alpha = 1$ приводит к модели с минимальным MSE в обучающем наборе, поэтому делаем вывод, что оптимальная эластичная сеть сводится к LASSO, таким образом, также выбирается 9 предикторов.</p>
5.4.2. Tree-based machine learning methods	5.4.2. Древовидные методы машинного обучения
<p>We next consider two statistical machine learning methods based on trees, namely Random Forest (Breiman, 2001) and Gradient Boosting (Friedman, 2001), which may further improve prediction performance when complex, nonlinear structures are present in the data. Both of these methods are rooted in the ensemble idea, that is, producing multiple trees which are then combined to create a single model to improve the prediction accuracy. Unlike the aforementioned high-dimensional statistical methods that yield sparse predictive models, the ensemble models based on Random Forest and Gradient Boosting can be difficult to interpret despite potential improvement in prediction accuracy.</p>	<p>Далее мы рассмотрим два статистических метода машинного обучения, основанных на деревьях, а именно: «Случайный лес» (Breiman, 2001) и «Повышение градиента» (Friedman, 2001), которые могут еще больше улучшить прогнозирование, когда в данных присутствуют сложные нелинейные структуры. Оба эти метода основаны на идее ансамбля, то есть на создании множества деревьев, которые затем объединяются, чтобы создать единую модель для повышения точности прогнозирования. В отличие от вышеупомянутых многомерных статистических методов, которые дают разреженные прогностические модели, ансамблевые модели, основанные на случайном лесном и градиентном ускорении, могут быть трудны для интерпретации, несмотря на потенциальное повышение точности прогнозирования.</p>
<p>Random forest. To produce multiple trees, this method employs bootstrap (Efron, 1979) to create B training samples, and then grows a random-forest tree from each bootstrapped sample until a certain minimum node size (e.g., 5) is reached. Often the prediction performance improves sharply in the beginning as the number of trees increases, but the performance stabilizes</p>	<p>Случайный лес. Для создания нескольких деревьев этот метод использует начальную загрузку (Efron, 1979) для создания обучающих выборок B, а затем вырастает дерево случайных лесов из каждой загруженной выборки до достижения определенного минимального размера узла (например, 5). Часто эффективность прогнозирования</p>

when we have enough trees.	резко увеличивается в начале, когда увеличивается количество деревьев, но производительность стабилизируется, когда у нас достаточно деревьев.
We begin our analysis by using 500 trees denoted by $B = 500$ as recommended in Hastie et al. (2009). Growing the tree is done as follows: in each and every step of split in the tree, randomly choose m predictors out of all p predictors as split candidates, and then use only the best one out of these m predictors. A key idea behind the Random Forest method is the random sampling of split candidates, which results in a fresh sample of m predictors chosen at each split. Thus, the choice of m impacts the prediction performance and m becomes a tuning parameter and this parameter is optimized by cross-validation. We follow Bertsimas et al. (2016) and consider $2 \cdot 10^{-15}$ m values among $\{10^{-15}, 10^{-14}, \dots, 10^{-1}\}$. We use the minimum node size 5 which ensures no split when growing the tree if the node size is less than 5 (see an example in Breiman, 2001).	Мы начинаем наш анализ с использования 500 деревьев, обозначенных $B = 500$, как рекомендовано в Hastie et al. (2009). Выращивание дерева выполняется следующим образом: на каждом шаге разбиения в дереве случайным образом выбирают m предикторов из всех p предикторов в качестве кандидатов на расщепление, а затем используют только лучший из этих m предикторов. Ключевая идея, лежащая в основе метода Случайные Лес, заключается в случайной выборке кандидатов-расщепителей, что приводит к новой выборке из m предикторов, выбранных при каждом разделении. Таким образом, выбор m влияет на эффективность прогнозирования, а m становится параметром настройки, и этот параметр оптимизируется путем перекрестной проверки. Мы следуем Берцимасу и соавт. (2016) и рассмотрим значения $2 \cdot 10^{-15}$ m среди $\{10^{-15}, 10^{-14}, \dots, 10^{-1}\}$. Мы используем минимальный размер узла 5, который гарантирует отсутствие разделения при выращивании дерева, если размер узла меньше 5 (см. Пример в Breiman, 2001).
To avoid a potential overfitting issue, we conducted further numerical experiments with larger values of minimum node size from 5 to 50 similar to Fig. 15.8 in Hastie et al. (2009) but did not see an improvement in prediction performance in the training set (see Table A4 in Online Appendix A3). We also experimented with B from 500 to 2,500 but the prediction performance in the training set did not improve and remained quite similar (see Table 5 in Online Appendix A3). We did not consider tuning another potential parameter, tree depth, because it is controlled by the minimum node size—the larger the minimum node size, the shallower the trees.	Чтобы избежать потенциальной проблемы переоснащения, мы провели дополнительные численные эксперименты с большими значениями минимального размера узла от 5 до 50, аналогично рис. 15.8 в Hastie et al. (2009), но не увидели улучшения в эффективности прогнозирования в обучающем наборе (см. Таблицу A4 в онлайн-приложении A3). Мы также экспериментировали с B от 500 до 2500, но эффективность прогнозирования в тренировочном наборе не улучшилась и оставалась довольно схожей (см. Таблицу 5 в онлайн-приложении A3). Мы не рассматривали настройку другого потенциального параметра, глубины дерева, поскольку он контролируется минимальным размером узла — чем больше минимальный размер узла, тем мельче деревья.
We implemented Random Forest in the R-package randomForest (Liaw & Wiener, 2002) and found the optimal $m = 859$ by 10-fold CV. This method fits the training data set well, in particular, it achieves a very high $R^2 = 0.990$. In the test set, however, it obtains lower prediction accuracy as compared to LASSO.	Мы реализовали Случайные Лес в R-пакете randomForest (Liaw & Wiener, 2002) и нашли оптимальное значение $m = 859$ при 10-кратном CV. Этот метод хорошо подходит для набора обучающих данных, в частности он достигает очень высокого значения $R^2 = 0.990$ в тестовом наборе, однако, он получает более низкую точность прогнозирования по сравнению с LASSO.
Gradient boosting. This alternative tree-based machine learning method, in	Повышение градиента. Этот альтернативный метод машинного

<p>contrast to Random Forest, grows trees in a sequential way to reduce bias as described next. We first build up to B trees each with $d + 1$ terminal nodes to the training data set, then repeatedly update each tree by adding a shrunk version of the new tree by a factor, $v \in (0, 1)$. Note that the new tree is fit on the current residuals, not on the response variables. At each iteration, a fraction ζ of the training data is sampled without replacement which is further used to grow the next tree. Because this is done in an adaptive way, the new tree depends on the previous tree. Finally, we combine all the trees to obtain an ensemble model as we did for Random Forest.</p>	<p>обучения, основанный на деревьях, в отличие от Случайного леса, последовательно выращивает деревья, чтобы уменьшить смещение, как описано далее. Сначала мы строим до B деревьев, каждое с $d + 1$ терминальными узлами, к набору обучающих данных, а затем многократно обновляем каждое дерево, добавляя сокращенную версию нового дерева с коэффициентом $v \in (0, 1)$. Обратите внимание, что новое дерево соответствует текущим остаткам, а не переменным ответа. На каждой итерации отбирается часть data обучающих данных без замены, которая затем используется для выращивания следующего дерева. Поскольку это делается адаптивным способом, новое дерево зависит от предыдущего дерева. Наконец, мы объединяем все деревья, чтобы получить модель ансамбля, как мы это делали для Случайного Леса.</p>
<p>As discussed above, we consider four key parameters to fit a Gradient Boosting model: d, v, ζ, B (see more details in §10.12 in Hastie et al., 2009) and implemented Gradient Boosting in the R-package gbm (Ridgeway, 2007). The parameter d is often called depth of the tree; v is referred to as shrinkage parameter or learning rate; ζ is the subsampling rate; the optimal number of iterations B^* is determined by cross validation, which is also denoted as early stopping in some literature (e.g., see Yao, Lorenzo Rosasco, & Andrea Caponnetto, 2007; Zhang & Yu, 2005).</p>	<p>Как обсуждалось выше, мы рассматриваем четыре ключевых параметра, подходящих для модели усиления градиента: d, v, ζ, B (см. Более подробную информацию в §10.12 в Hastie et al., 2009) и реализуем повышение градиента в R-пакете gbm (Ridgeway), 2007). Параметр d часто называют глубиной дерева; v упоминается как параметр усадки или скорость обучения; ζ - частота подвыборки; оптимальное число итераций B^* определяется перекрестной проверкой, которая в некоторых литературах также называется ранней остановкой (например, см. Yao, Lorenzo Rosasco, & Andrea Caponnetto, 2007; Zhang & Yu, 2005).</p>
<p>We use $d = 1$ because our set of predictor variables already includes higher-order interaction terms. We begin model fitting by exploring a range of small shrinkage parameters $v \in [0.0005, 0.1]$ as suggested by Hastie et al. (2009) with default subsampling rate, $\zeta = 0.5$ which tends to work well though perhaps is not optimal. For each v we allow the algorithm to grow up to 5000 trees and for this v value we derive the corresponding optimal B^* by 10-fold CV. For example, for $v = 0.1$ we find that the best cross-validation iteration is obtained in 1234 iterations, therefore the corresponding $B^* = 1234$. We find that in our data set $v < 0.01$ leads to a poorer model fit in the training data sets. Therefore, we focus on v values between 0.01 and 0.1 and further calibrate the model by tuning both parameters v, ζ. We provide details on the model calibration in Tables A6 and A7 in Online Appendix A4.</p>	<p>Мы используем $d = 1$, потому что наш набор переменных предикторов уже включает члены взаимодействия более высокого порядка. Мы начинаем подгонку модели, исследуя диапазон малых параметров усадки $v \in [0.0005, 0.1]$, как предложено Hastie et al. (2009) с частотой субдискретизации по умолчанию, $\zeta = 0.5$, которая имеет тенденцию работать хорошо, хотя, возможно, не является оптимальной. Для каждого v мы позволяем алгоритму расти до 5000 деревьев, а для этого значения v мы получаем соответствующий оптимальный B^* по 10-кратному CV. Например, для $v = 0.1$ мы находим, что лучшая итерация перекрестной проверки получается в 1234 итерациях, поэтому соответствующий $B^* = 1234$. Мы находим, что в нашем наборе данных $v < 0.01$ приводит к худшему соответствию модели в наборах обучающих данных. Поэтому мы сосредоточимся на значениях v в диапазоне от 0,01 до 0,1 и дополнительно откалибруем модель,</p>

	настроив оба параметра ν , ζ . Мы предоставляем подробную информацию о калибровке модели в таблицах A6 и A7 в онлайн-приложении A4.
The best calibrated model with respect to the training data set is obtained by $\nu = 0.1$, $\zeta = 0.4$, $B^* = 1140$ which achieves $MSE = 10.606$ and $R^2 = 0.996$; however, in the test set it obtains $MSE = 290.564$ and $R^2 = 0.882$. This gradient boosting model performs slightly better than Random Forest in our study but does not outperform LASSO in the test set.	Наилучшая откалиброванная модель по набору обучающих данных получается при $\nu = 0.1$, $\zeta = 0.4$, $B^* = 1140$, который достигает $MSE = 10.606$ и $R^2 = 0.996$; однако в тестовом наборе он получает $MSE = 290,564$ и $R^2 = 0.882$. Эта модель повышения градиента работает немного лучше, чем Случайный Лес в нашем исследовании, но не превосходит LASSO в тестовом наборе.
To summarize, we find that the models with unbiased estimators (LARS-OLS hybrid, SCAD) fit the training set better than LASSO, but yield lower prediction accuracy in the test set, which suggests bias-variance trade-off as discussed above. The reduction of Elastic Net to LASSO indicates that the true underlying model is much more likely a sparse model as we expected at the end of Section 5.1. It is not surprising that the two tree-based methods (Random Forest, Gradient Boosting) fit the training set very well, however, their prediction performance in the test set suggests likely overfitting issues of these complex models despite the optimal choice of tuning parameters. In our setting, the LASSO model shows up as the best because it obtains the smallest MSE in the test set among all the models considered, and also because its MSE in the training and test sets are very close, which suggests that the LASSO model is a robust choice.	Подводя итог, мы находим, что модели с несмещенными оценщиками (гибрид LARS-OLS, SCAD) соответствуют обучающему набору лучше, чем LASSO, но дают более низкую точность прогнозирования в тестовом наборе, что предполагает компромисс между отклонениями, как обсуждалось выше. Сокращение Elastic Net до LASSO указывает на то, что истинная базовая модель, скорее всего, является разреженной моделью, как мы и ожидали в конце Раздела 5.1. Неудивительно, что два древовидных метода (Random Forest, Gradient Boosting) очень хорошо соответствуют обучающему набору, однако их эффективность прогнозирования в тестовом наборе позволяет предположить вероятные проблемы наложения этих сложных моделей, несмотря на оптимальный выбор параметров настройки. В наших настройках модель LASSO показывается как лучшая, потому что она получает наименьшее MSE в тестовом наборе среди всех рассмотренных моделей, а также потому, что ее MSE в обучающем и тестовом наборах очень близки, что говорит о том, что модель LASSO надежный выбор.
6. Robustness checks	6. Проверка надежности
In our main analysis, we focused on training our predictive models on 33 periods $\{1, \dots, 33\}$ and then testing the prediction performance for the 6 future periods $\{34, \dots, 39\}$. In this section, we evaluate the robustness of three predictive models: models (5), (6) and the LASSO model on different test sets. In our study, we focus on empirically investigating whether the LASSO model consistently selects the same 9 predictors as in the main analysis shown in Table 10, and also check the sensitivity of prediction performance on such test sets. In Section 6.1, we construct different test sets that contain 6 randomly chosen periods, for example, $\{3, 13, 25, 34, 35, 37\}$. Next, in sub section 6.2, we check the robustness of the length of test period	В нашем основном анализе мы сосредоточились на обучении наших прогностических моделей на 33 периодах $\{1, \dots, 33\}$, а затем тестирование прогнозирования для 6 будущих периодов $\{34, \dots, 39\}$. В этом разделе мы оцениваем надежность трех прогнозирующих моделей: моделей (5), (6) и модели LASSO на разных тестовых наборах. В нашем исследовании мы фокусируемся на эмпирическом исследовании того, последовательно ли модель LASSO выбирает те же 9 предикторов, что и в основном анализе, показанном в таблице 10, а также проверяем чувствительность эффективности прогнозирования на таких наборах тестов. В разделе 6.1 мы создаем различные тестовые наборы, которые содержат 6 случайно выбранных периодов, например, $\{3, 13, 25, 34, 35,$

using alternative test sets with 5 and 7 future periods. For both of these scenarios, we compare the results with our finding in the main analysis.	37}. Далее, в подразделе 6.2, мы проверяем надежность продолжительности тестового периода, используя альтернативные тестовые наборы с 5 и 7 будущими периодами. Для обоих этих сценариев мы сравниваем результаты с нашими результатами в основном анализе.																																																																																										
6.1. Test on 6 random periods	6.1. Тест на 6 случайных периодов																																																																																										
In this subsection, we construct a test set by randomly choosing 6 periods over the set of periods $\{1, \dots, 39\}$, and assign the remaining 33 periods to the training set. We repeat this process 4 times to create 4 such test sets to check if different random test sets lead to substantially different results. We obtain test periods $\{3, 13, 25, 34, 35, 37\}$ for case 1, $\{10, 15, 19, 27, 28, 34\}$ for case 2, $\{4, 8, 15, 25, 29, 38\}$ for case 3, $\{2, 5, 17, 27, 30, 33\}$ for case 4. For each of the 4 cases, we build three predictive models using model (5), (6) and the LASSO model on the corresponding training sets, then test their prediction performance in each test set. The results are given in Table 12 for models (5) and (6), and Table 13 for the LASSO model.	В этом подразделе мы строим тестовый набор путем случайного выбора 6 периодов из набора периодов $\{1, \dots, 39\}$ и назначаем оставшиеся 33 периода тренировочному набору. Мы повторяем этот процесс 4 раза, чтобы создать 4 таких набора тестов, чтобы проверить, приводят ли разные случайные наборы тестов к существенно различным результатам. Мы получаем тестовые периоды $\{3, 13, 25, 34, 35, 37\}$ для случая 1, $\{10, 15, 19, 27, 28, 34\}$ для случая 2, $\{4, 8, 15, 25, 29, 38\}$ для случая 3, $\{2, 5, 17, 27, 30, 33\}$ для случая 4. Для каждого из 4 случаев мы строим три прогнозные модели, используя модель (5), (6) и модель LASSO на соответствующем обучающие наборы, а затем проверить их эффективность прогнозирования в каждом тестовом наборе. Результаты приведены в таблице 12 для моделей (5) и (6) и в таблице 13 для модели LASSO.																																																																																										
Table 12 Comparison of 4 cases of random test periods against the main analysis for models 5 and 6.	Таблица 12 Сравнение 4 наборов случайных периодов тестирования с основным анализом для моделей 5 и 6.																																																																																										
<table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Main analysis</th><th colspan="2">Case 1</th><th colspan="2">Case 2</th><th colspan="2">Case 3</th><th colspan="2">Case 4</th></tr><tr><th>model 5</th><th>model 6</th><th>model 5</th><th>model 6</th><th>model 5</th><th>model 6</th><th>model 5</th><th>model 6</th><th>model 5</th><th>model 6</th></tr><tr><td rowspan="3">Training set</td><td>Data size</td><td>1,140</td><td>1,140</td><td>1,145</td><td>1,145</td><td>1,146</td><td>1,146</td><td>1,152</td><td>1,152</td><td>1,154</td><td>1,154</td></tr><tr><td>R^2</td><td>0.912</td><td>0.921</td><td>0.910</td><td>0.919</td><td>0.911</td><td>0.922</td><td>0.914</td><td>0.924</td><td>0.913</td><td>0.924</td></tr><tr><td>MSE</td><td>214.050</td><td>192.395</td><td>224.295</td><td>202.027</td><td>226.203</td><td>196.866</td><td>212.314</td><td>186.034</td><td>219.111</td><td>190.382</td></tr><tr><td rowspan="3">Test set</td><td>Data size</td><td>220</td><td>220</td><td>183</td><td>183</td><td>256</td><td>256</td><td>256</td><td>256</td><td>256</td><td>256</td></tr><tr><td>R^2</td><td>0.895</td><td>0.923</td><td>0.917</td><td>0.941</td><td>0.912</td><td>0.922</td><td>0.897</td><td>0.912</td><td>0.901</td><td>0.912</td></tr><tr><td>MSE</td><td>257.709</td><td>189.386</td><td>188.735</td><td>134.561</td><td>175.970</td><td>156.695</td><td>248.364</td><td>213.197</td><td>209.560</td><td>187.039</td></tr></table>			Main analysis		Case 1		Case 2		Case 3		Case 4		model 5	model 6	model 5	model 6	model 5	model 6	model 5	model 6	model 5	model 6	Training set	Data size	1,140	1,140	1,145	1,145	1,146	1,146	1,152	1,152	1,154	1,154	R^2	0.912	0.921	0.910	0.919	0.911	0.922	0.914	0.924	0.913	0.924	MSE	214.050	192.395	224.295	202.027	226.203	196.866	212.314	186.034	219.111	190.382	Test set	Data size	220	220	183	183	256	256	256	256	256	256	R^2	0.895	0.923	0.917	0.941	0.912	0.922	0.897	0.912	0.901	0.912	MSE	257.709	189.386	188.735	134.561	175.970	156.695	248.364	213.197	209.560	187.039	
			Main analysis		Case 1		Case 2		Case 3		Case 4																																																																																
		model 5	model 6	model 5	model 6	model 5	model 6	model 5	model 6	model 5	model 6																																																																																
Training set	Data size	1,140	1,140	1,145	1,145	1,146	1,146	1,152	1,152	1,154	1,154																																																																																
	R^2	0.912	0.921	0.910	0.919	0.911	0.922	0.914	0.924	0.913	0.924																																																																																
	MSE	214.050	192.395	224.295	202.027	226.203	196.866	212.314	186.034	219.111	190.382																																																																																
Test set	Data size	220	220	183	183	256	256	256	256	256	256																																																																																
	R^2	0.895	0.923	0.917	0.941	0.912	0.922	0.897	0.912	0.901	0.912																																																																																
	MSE	257.709	189.386	188.735	134.561	175.970	156.695	248.364	213.197	209.560	187.039																																																																																
Table 13 Comparison of 4 cases of random test periods against the main analysis for the LASSO model.	Таблица 13 Сравнение 4 наборов случайных периодов тестирования с основным анализом для модели LASSO.																																																																																										
<table><tr><th colspan="2"></th><th>Main analysis</th><th>Case 1</th><th>Case 2</th><th>Case 3</th><th>Case 4</th></tr><tr><td rowspan="4">Training set</td><td>Data size</td><td>1,140</td><td>1,145</td><td>1,146</td><td>1,152</td><td>1,154</td></tr><tr><td>λ_{1se}</td><td>4.378</td><td>5.273</td><td>4.619</td><td>5.245</td><td>4.824</td></tr><tr><td>R^2</td><td>0.934</td><td>0.924</td><td>0.932</td><td>0.925</td><td>0.928</td></tr><tr><td>MSE</td><td>161.908</td><td>189.366</td><td>172.608</td><td>183.696</td><td>180.537</td></tr><tr><td rowspan="3">Test set</td><td>Data size</td><td>220</td><td>215</td><td>214</td><td>208</td><td>206</td></tr><tr><td>R^2</td><td>0.930</td><td>0.944</td><td>0.936</td><td>0.916</td><td>0.924</td></tr><tr><td>MSE</td><td>171.824</td><td>128.335</td><td>128.143</td><td>204.272</td><td>160.598</td></tr></table>			Main analysis	Case 1	Case 2	Case 3	Case 4	Training set	Data size	1,140	1,145	1,146	1,152	1,154	λ_{1se}	4.378	5.273	4.619	5.245	4.824	R^2	0.934	0.924	0.932	0.925	0.928	MSE	161.908	189.366	172.608	183.696	180.537	Test set	Data size	220	215	214	208	206	R^2	0.930	0.944	0.936	0.916	0.924	MSE	171.824	128.335	128.143	204.272	160.598																																								
		Main analysis	Case 1	Case 2	Case 3	Case 4																																																																																					
Training set	Data size	1,140	1,145	1,146	1,152	1,154																																																																																					
	λ_{1se}	4.378	5.273	4.619	5.245	4.824																																																																																					
	R^2	0.934	0.924	0.932	0.925	0.928																																																																																					
	MSE	161.908	189.366	172.608	183.696	180.537																																																																																					
Test set	Data size	220	215	214	208	206																																																																																					
	R^2	0.930	0.944	0.936	0.916	0.924																																																																																					
	MSE	171.824	128.335	128.143	204.272	160.598																																																																																					

<p>First, we find that the LASSO model selects the same 9 predictors in cases 1 to 4 as the ones found in the main analysis. This is consistent with the theory of LASSO's consistent variable selection with high probability that was established under the irrepresentable conditions (Meinshausen & Buhlmann, 2006; Zhao & Yu, 2006). Second, we observe that the prediction performance in the 4 cases shows some fluctuation for different test periods. Third, we find that the LASSO model outperforms both main effects models in the training and test sets in all 4 cases as in the main analysis. The advantage of the LASSO model over model (6) is not as substantial, when compared to model (5) . For instance, the LASSO model reduces the test MSE in model (5) by 33.3% in the main analysis, and between 17.8% and 32.0% among 4 cases. By comparison, the LASSO model reduces the test MSE in model (6) by 9.3% in the main analysis, and between 4.2% and 18.2% among the 4 cases.</p>	<p>Во-первых, мы обнаруживаем, что модель LASSO выбирает те же 9 предикторов в случаях с 1 по 4, что и в основном анализе. Это согласуется с теорией последовательного выбора переменных LASSO с высокой вероятностью, которая была установлена в непредставимых условиях (Meinshausen & Buhlmann, 2006; Zhao & Yu, 2006). Во-вторых, мы наблюдаем, что эффективность прогнозирования в 4 случаях показывает некоторые колебания для разных тестовых периодов. В-третьих, мы обнаруживаем, что модель LASSO превосходит обе модели основных эффектов в учебных и тестовых наборах во всех 4 случаях, как и в основном анализе. Преимущество модели LASSO над моделью (6) не так существенно, по сравнению с моделью (5). Например, модель LASSO снижает тест MSE в модели (5) на 33,3% в основном анализе, и между 17,8% и 32,0% в 4 случаях. Для сравнения, модель LASSO снижает тест MSE в модели (6) на 9,3% в основном анализе, и между 4,2% и 18,2% в 4 случаях.</p>
<p>6.2. Test on 5 and 7 future periods</p>	<p>6.2. Тест на 5 и 7 будущих периодов</p>
<p>Our goal in this subsection is to understand if the choice of training set periods impacts the predictive models and their performance. To do this, we consider two alternative analyses: train all three predictive models on periods { 1 , . . . , 34 } and test their performance on periods { 35 , . . . , 39 } ; then also train on periods { 1 , . . . , 32 } and test on periods { 33 , . . . , 39 } . We show the results in Table 14 .</p>	<p>Наша цель в этом подразделе - понять, влияет ли выбор периодов обучения на прогностические модели и их эффективность. Для этого мы рассмотрим два альтернативных анализа: обучаем все три прогнозирующие модели периодам {1, . . . , 34} и проверить их работоспособность в периоды {35, . . . , 39}; затем также тренироваться в периоды {1, . . . , 32} и тестирование на периоды {33, . . . , 39}. Мы показываем результаты в таблице 14.</p>
<p>Similar to the main analysis, we find that the LASSO model continues to select the same set of 9 predictors in the two new test sets containing 5 and 7 future periods. This indicates that the “sparse model” selection by LASSO is consistent and robust with respect to the different test periods we considered.</p>	<p>Аналогично основному анализу, мы обнаруживаем, что модель LASSO продолжает выбирать один и тот же набор из 9 предикторов в двух новых наборах тестов, содержащих 5 и 7 будущих периодов. Это указывает на то, что выбор «разреженной модели» с помощью LASSO является последовательным и надежным в отношении различных рассмотренных нами периодов испытаний.</p>
<p>Now we compare prediction performance using the measure MSE. In the main analysis, the choice of LASSO over models (5) and (6) achieved a reduction in the test MSE by 33.3% and 9.3% respectively. When we train our models on 34 periods and test on 5 peri-ods, the LASSO model reduces the test MSE of model (5) by 15.8%. However, it increases the test MSE of model (6) by 10.4%. This indicates that LASSO does not outperform model (6) when the MSE was computed over the 5 test periods. Nonetheless, if we</p>	<p>Теперь мы сравним показатели прогнозирования с помощью показателя MSE. В основном анализе выбор LASSO по сравнению с моделями (5) и (6) позволил добиться снижения тестовой MSE на 33,3% и 9,3% соответственно. Когда мы обучаем наши модели в течение 34 периодов и проводим тестирование в течение 5 периодов, модель LASSO снижает тестовое MSE модели (5) на 15,8%. Тем не менее, это увеличивает тест MSE модели (6) на 10,4%. Это указывает на то, что LASSO не превосходит модель (6), когда MSE вычислялась в течение 5 тестовых</p>

<p>only consider the prediction performance in the first test period (trained on periods 1 to 34, tested on period 35 only) we can see from Table 8 in §5.2 that LASSO still beats model (6) and reduces the $-103 \cdot 170$ MSE by $204 \cdot 381 = 49 \cdot 5\%$. In summary, even though the full $204 \cdot 381$ main effects model (6) is competitive in prediction performance compared to the LASSO model based on MSE computed over the entire set of test periods, we find that the LASSO model consistently outperforms both main effects models in the first few periods. This suggests that the LASSO model is the better choice when the predictive model is implemented on a rolling horizon basis.</p>	<p>периодов. Тем не менее, если мы рассмотрим только результаты прогнозирования в первом тестовом периоде (обученные в периоды с 1 по 34, протестированные только в период 35), мы увидим из таблицы 8 в п. 5.2, что LASSO по-прежнему превосходит модель (6) и снижает $-103 \cdot 170$ мсэ к $204 \cdot 381 = 49 \cdot 5\%$. В итоге, хотя полный $204 \cdot 381$ модель основных эффектов (6) является конкурентоспособной по эффективности прогнозирования по сравнению с моделью LASSO, основанной на MSE, рассчитанной по всему набору тестовых периодов, и мы обнаруживаем, что модель LASSO последовательно превосходит обе модели основных эффектов в первые несколько периодов. Это говорит о том, что модель LASSO является лучшим выбором, когда прогнозирующая модель реализуется на основе скользящего горизонта.</p>
<p>7. Discussion</p>	<p>7. Обсуждение</p>
<p>In this paper, we developed data-driven models to predict future return volume and applied them to a firm that sells car accessories with large product variety. We explored various factors that could help predict return volume—sales, time, product features, retailer, production process and resources, multi-product order and historical returns, which are not exclusive to the particular firm in our study. We evaluated various main effects models and an interaction effect model with convex and concave regularization, to build prediction models that have good prediction accuracy and are easy to implement in practice and also provide a preliminary understanding of important factors associated with return volume. We also considered tree-based machine learning methods to improve on the prediction accuracy. We found that LASSO was effective in selecting a small number of interaction terms, which are useful in prediction, out of a large number of possible candidates. LASSO identified a parsimonious model that achieved the highest accuracy in predicting future return volume.</p>	<p>В этой статье мы разработали модели, основанные на данных, для прогнозирования будущего объема возврата и применили их к фирме, которая продает автомобильные аксессуары с большим ассортиментом продукции. Мы исследовали различные факторы, которые могли бы помочь предсказать объемы продаж - время, время, характеристики продукта, розничный продавец, производственный процесс и ресурсы, заказы из нескольких продуктов и исторические доходы, которые не являются исключительными для конкретной фирмы в нашем исследовании. Мы оценили различные модели основных эффектов и модель эффекта взаимодействия с выпуклой и вогнутой регуляризацией, чтобы построить модели прогнозирования, которые имеют хорошую точность прогнозирования и которые легко реализовать на практике, а также обеспечивают предварительное понимание важных факторов, связанных с объемом возврата. Мы также рассмотрели древовидные методы машинного обучения, чтобы улучшить точность прогнозирования. Мы обнаружили, что LASSO был эффективным в выборе небольшого числа слагаемых взаимодействия, которые полезны для прогнозирования, из большого числа возможных кандидатов. LASSO определила экономную модель, которая достигла высочайшей точности в прогнозировании будущих объемов возврата.</p>
<p>Sales show up as an important predictor across all models. In the LASSO model, we find that the predicted return volume is convex and increasing in sales, suggesting that the likelihood of returns increases for each incremental</p>	<p>Продажи отображаются в качестве важного фактора во всех моделях. В модели LASSO мы обнаруживаем, что прогнозируемый объем возврата выпуклый и увеличивается в продажах, что позволяет предположить, что вероятность возврата увеличивается с каждой проданной единицей.</p>

<p>resold in most cases, they will have to carefully weigh the cost of such a high return volume from this retailer in determining the terms of trade with this retailer. They should also explore the reasons for the high returns with the retailer.</p>	<p>июля 2012 года по сентябрь 2015 года доходность была самой высокой - 13,6%. Поскольку возвращенные продукты в Компании А в большинстве случаев не могут быть перепроданы, они будут При определении условий торговли с этим розничным продавцом необходимо тщательно взвесить стоимость такого большого объема возврата этого продавца. Они также должны изучить причины высокой доходности у продавца.</p>
<p>It was expected ex ante that product type would be useful in predicting returns. The rationale was that the three products—dash cover, seat cover and car cover—have distinct features, designs, purchase/usage characteristics and they go through different manufacturing processes that may have varying defect rates. In fact, the results from the main effects models (see Table 5) show that dash cover products lead to significantly lower return volume than vehicle cover products. The LASSO model, however, did not select product type in predicting return volume. Similarly, the LASSO model did not select any of the production process and resources related variables, which were ex ante expected to be related to defects and useful in predicting returns. This suggests that the returns due to defects in the firm in our study may not be substantial, which is consistent with the finding in a study (Lawton, 2008) that only about 5% of returns are due to true defects.</p>	<p>Ожидалось, что этот продукт будет полезен для прогнозирования прибыли. Обоснованием было то, что эти три продукта - покрытие для панели приборов, чехол для сиденья и чехол для автомобиля - имеют отличительные особенности, конструкции, характеристики покупки / использования и проходят через различные производственные процессы, которые могут иметь различную частоту дефектов. Фактически, результаты моделей основных эффектов (см. Таблицу 5) показывают, что продукты с покрытием для панели приборов приводят к значительно меньшему возвратному объему, чем продукты с автомобильными чехлами. Модель LASSO, однако, не выбрала тип продукта при прогнозировании объема возврата. Аналогичным образом, модель LASSO не выбрала ни одну из переменных, связанных с производственным процессом и ресурсами, которые, как предполагалось, были связаны с дефектами и были полезны при прогнозировании прибыли. Это говорит о том, что доходы из-за дефектов фирмы в нашем исследовании могут быть незначительными, что согласуется с выводом исследования (Lawton, 2008) о том, что только около 5% доходов связаны с истинными дефектами.</p>
<p>We noticed in our analysis that products with logos or other types of customization appeared to have lower returns. However, we do not find them to be useful in prediction in any of the models, likely because sales of such highly customized products are very small at Company A. However, the lower return rate suggests that perhaps Company A could consider promoting such products to consumers</p>	<p>В нашем анализе мы заметили, что продукты с логотипами или другими типами персонализации, по-видимому, имеют более низкую прибыль. Однако мы не находим их полезными для прогнозирования ни в одной из моделей, вероятно, потому, что продажи таких продуктов с высокой степенью индивидуальности очень малы в компании А. Однако более низкая доходность предполагает, что, возможно, компания А могла бы рассмотреть возможность продвижения таких продуктов потребителям</p>
<p>The inability to distinguish returns due to defect from other returns is one of the limitations of our study. Unfortunately, Company A does not systematically track the reasons for returns. If we are able to accurately identify returns due to defects, then we could ask interesting questions related to defects: whether defects differ substantially among different products,</p>	<p>Неспособность отличить возвращение из-за дефекта от других возвратов является одним из ограничений нашего исследования. К сожалению, компания А не систематически отслеживает причины возврата. Если мы сможем точно идентифицировать возвраты из-за дефектов, то мы могли бы задать интересные вопросы, связанные с</p>

whether production process and resources are important in predicting defects, etc.	дефектами: существенно ли различаются дефекты у разных продуктов, важны ли производственный процесс и ресурсы для прогнозирования дефектов и т. Д.
An important issue for future research is the potential costs and benefits of retailer return policies. For example, do we want retailers to strictly enforce return policy or should we allow retailers to be liberal in their return policy and in enforcing it? With stricter return policies and enforcement, there may be fewer returns, however, this may discourage consumers from making purchases. The trade-off between sales and returns is not immediately clear. A field experiment, which modifies return policies of select retailers and tracks subsequent changes in returns, may be able to address this question. When we can accurately extract defects among returns, we may be able to address some system design questions, by analyzing effects of production process and resources on defects. For example, should we increase random inspections by supervisors, if we find that such random inspections reduce defects? Another example is – do workers who speed up produce greater defects resulting in higher returns?	Важным вопросом для будущих исследований являются потенциальные издержки и выгоды политики возврата розничных продавцов. Например, хотим ли мы, чтобы розничные продавцы строго соблюдали политику возврата, или мы должны позволить розничным торговцам быть либеральными в своей политике возврата и в ее применении? При более строгих правилах возврата и правоприменении возврат может быть меньше, однако это может отговорить потребителей от совершения покупок. Компромисс между продажами и возвратами не сразу ясен. Полевой эксперимент, который изменяет политику возврата избранных розничных продавцов и отслеживает последующие изменения в возврате, может решить этот вопрос. Когда мы сможем точно выявлять дефекты среди возвратов, мы сможем решить некоторые вопросы проектирования системы, проанализировав влияние производственного процесса и ресурсов на дефекты. Например, должны ли мы увеличить выборочные проверки со стороны надзорных органов, если мы обнаружим, что такие выборочные проверки уменьшают количество дефектов? Другой пример - рабочие, которые ускоряются, производят большие дефекты, приводящие к более высокой доходности?
Acknowledgment	
We would like to express our sincere gratitude to Richard Vidgen and three anonymous referees who provided critical feedback that helped us significantly improve the manuscript.	
Supplementary material	
Supplementary material associated with this article can be found, in the online version, at doi: 10.1016/j.ejor.2019.05.046	
References	
Akturk, M. S. , Ketzenberg, M. , & Heim, G. R. (2018). Assessing impacts of introducing ship-to-store service on sales and returns in omnichannel retailing: A data analytics study. <i>Journal of Operations Management</i> , 61 , 15–45 . Alptekino ğlu, A. , & Grasas, A. (2017). When to carry eccentric products? Optimal retail assortment under consumer returns. <i>Production and</i>	

Operations Management, 23 (5), 877–892 . Altug, M. S. , & Aydinliyim, A. (2016). Counteracting strategic purchase deferrals: the impact of online retailers return policy decisions. *Manufacturing & Service Operations Management*, 18 (3), 376–392 . Anderson, E. T. , Karsten, H. , & Duncan, S. (2009). The option value of returns: Theory and empirical evidence. *Marketing Science*, 28 (3), 405–423 . Belloni, A. , & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19 (2), 521–547 . Bertsimas, D. , O’Hair, A. , Relyea, S. , & Silberholz, J. (2016). An analytics approach to designing combination chemotherapy regimens for cancer. *Management Science*, 62 (5), 1511–1531 . Bickel, P. J. , Ritov, Y. , & Tsybakov, A. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37 (4), 1705–1732 . Box, G. E. P. , & Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society, Series B.*, 26 (2), 211–252 . Breheny, P. , & Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *The Annals of Applied Statistics*, 5 (1), 232–253 . Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24 (6), 2350–2383 . Breiman, L. (2001). Random forests. *Machine Learning*, 45 (1), 5–32 . Calmon, A. P. , & Graves, S. C. (2017). Inventory management in a consumer electronics closed-loop supply chain. *Manufacturing & Service Operations Management*, 19 (4), 568–585 . Cerag, P. , Ferguson, M. , & Toktay, L. B. (2016). Extracting maximum value from consumer returns: Allocating between remarketing and refurbishing for warranty claims. *Manufacturing & Service Operations Management*, 18 (4), 475–492 . Cheng, A. (2015). Consumers return \$642.6 billion in goods each year. MarketWatch, Inc. . Accessed April 15, 2018, <http://www.marketwatch.com/story/consumersreturn6426billioningoodseachyear20150618> . DeCroix G, A. , Song, J. S. , & Zipkin, P. H. (2009). Managing an assemble-to-order system with returns. *Manufacturing & Service Operations Management*, 11 (1), 144–159 . Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7 (1), 1–26 . Efron, B. , Hastie, T. , Johnstone, I. , & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32 (2), 407–499 . Fan, J. , Feng, Y. , & Wu, Y. (2009). Network exploration via the adaptive LASSO and SCAD penalties. *The Annals of Applied Statistics*, 3 (2), 521–541 . Fan, J. , & Li, R. (2001). Variable

selection via nonconcave penalized likelihood and its oracle properties. *Journal American Statistical Association*, 96 (456), 1348–1360 . Fan, J. , Lv, J. , & Qi, L. (2011). Sparse high-dimensional models in economics. *Annual Review of Economics.*, 3 , 291–317 . Foster, D. P. , & George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22 (4), 1947–1975 . Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29 (5), 1189–1232 . Friedman, J. , Hastie, T. , & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software.*, 33 (1) . Hastie, T. , Tibshirani, R. , & Friedman, J. (2009). *The Elements of Statistical Learning* . Springer-Verlag . Hess, J. D. , & Mayhew, G. E. (1997). Modeling merchandise returns in direct marketing. *Journal of Interactive Marketing*, 11 (2), 20–35 . Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32 (1), 1–49 . Insider, B. (2017). National retail federation estimates 8–12% US e-commerce growth in 2017 . Accessed April 15, 2018, <http://www.businessinsider.com/nationalretailfederationestimates812usecommercegrowthin201720172> . James, G. , Witten, D. , Hastie, T. , & Tibshirani, R. (2017). *An Introduction to Statistical Learning* . Springer-Verlag . Janakiramana, N. , Syrdalb, H. A. , & Freling, R. (2016). The effect of return policy leniency on consumer purchase and return decisions: A meta-analytic review. *Journal of Retailing*, 92 (2), 226–235 . Lawton, C. (2008). The war on returns. *Wall Street Journal* . (May 8) D1, Accessed April 15, 2018, <https://www.wsj.com/articles/SB121020824820975641> . Li, G., Li, L., Sethi, S. P., & Guan, X. (2017). Return strategy and pricing in a dualchannel supply chain. *International Journal of Production Economics* . doi: 10.1016/j.ijpe.2017.06.031 . Liaw, A., & Wiener, M. (2002). Classification and regression by random-forest. *R News*, 2 (3), 18–22 . <http://CRAN.R-project.org/doc/Rnews/> . Ma, S. , Fildes, R. , & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intraand inter-category promotional information. *European Journal of Operational Research*, 249 (1), 245–257 . Martin, C. (2018). Mobile millennials: 63% shop on smartphones every, day, 53% buy in stores . Accessed April 15 <https://www.mediapost.com/publications/article/282639/mobilemillennials63shoponsmartphonesevery.html> . Martinez, A., Schmuck, C., Pereverzyev, S., Pirker, C., & Haltmeier, M. (2018). A

machine learning framework for customer purchase prediction in the noncontractual setting. *European Journal of Operational Research* . doi: 10.1016/j.ejor. 2018.04.034 . Meinshausen, N. , & Buhlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34 (3), 1436–1462 . Montaldo, D. L. (2019). Target bucks the trend with a new super lenient return policy. *Manufacturing and Technology News* . Accessed April 15, 2018, <https://www.thebalance.com/targetsnewreturnpolicy939859> . Mortenson, M. J. , Doherty, N. F. , & Robinson, S. (2015). Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241 (3), 583–595 . Oztekin, A. , Al-Ebbini, L. , Sevkli, Z. , & Delen, D. (2018). A decision analytic approach to predicting quality of life for lung transplant recipients: a hybrid genetic algorithms-based methodology. *European Journal of Operational Research*, 266 (2), 639–651 . Petersen, J. , & Kumar, V. (2009). Are product returns a necessary evil? Antecedents and consequences. *Journal of Marketing*, 73 (3), 35–51 . Ridgeway, G. (2007). Generalized boosted models: A guide to the GBM package . Accessed April 15, 2018, <http://www.saedsayad.com/docs/gbm2.pdf> . Rudolph, S. (2016). E-commerce product return statistics and trends. *Business 2 Community* . Accessed April 15, 2018, <http://www.business2community.com/infographics/ecommerceproductreturnstatisticstrendsinfographic01505394> . Ryzhov, I. O. , Han, B. , & Bradic, J. (2016). Cultivating disaster donors using data analytics. *Management Science*, 62 (3), 849–866 . Sevim, C. , Oztekin, A. , Bali, O. , Gumus, S. , & Guresen, E. (2014). Developing an early warning system to predict currency crises. *European Journal of Operational Research*, 237 (3), 1095–1104 . Shang, G. , Ghosh, B. P. , & Galbreth, M. (2017b). Optimal retail return policies with wardrobing. *Production and Operations Management*, 26 (7), 1315–1332 . Shang, G. , Pekgün, P. , Ferguson, M. , & Galbreth, M. (2017). How much do online consumers really value free product returns? Evidence from eBay. *Journal of Operations Management*, 53–56 , 45–62 . Shulman, J. D. , Coughlan, A. T. , & Savaskan, R. C. (2011). Managing consumer returns in a competitive environment. *Management Science*, 57 (2), 347–362 . Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B.*, 36 (2), 111–147 . Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation

and Akaike's criterion. *Journal of the Royal Statistical Society, Series B.*, 39 (1), 44–47 . Su, X. (2009). Consumer returns policies and supply chain performance. *Manufacturing & Service Operations Management*, 11 (4), 595–612 . The Economist (2014). The business of reselling returned shop items: What happens to all the goods shoppers don't want . Accessed April 15, 2018, <https://www.economist.com/news/business/21710855whathappensallgoodsshoppersdontwantbusinessresellingreturnedshopitems> . Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B.*, 58 (1), 267–288 . Toktay, L. B. (2003). Forecasting product returns. business aspects of closed-loop supply chains. In V. D. R. Guide, & L. N. Van Wassenhove (Eds.), *International management series. vol 2* . Carnegie Bosch Institute . Tsiliyannis, C. A. (2018). Markov chain modeling and forecasting of product returns in remanufacturing based on stock mean-age. *European Journal of Operational Research*, 271 (2), 474–489 . Ülkü, M. A. , & Gürler, U. (2017). The impact of abusing return policies: A newsvendor model with opportunistic consumers. *International Journal of Production Economics*, 203 , 124–133 . Urbanke, P. , Kranz, J. , & Kolbe, L. M. (2015). Predicting product returns in e-commerce: The contribution of mahalanobis feature extraction. *Proceedings of the thirty sixth international conference on information systems* . Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28 (2), 3–28 . Wilms, I., Basu, S., Bien, J., & Matteson, D. (2017). Interpretable vector autoregressions with exogenous time series. *NIPS symposium proceedings* . <http://arxiv.org/abs/1711.03623> . Yao, Y. , Lorenzo Rosasco, L. , & Andrea Caponnetto, A. (2007). On early stopping in gradient descent learning. *Constructive Approximation*, 26 (2), 289–315 . Zhang, T. , & Yu, B. (2005). Boosting with early stopping: Convergence and consistency. *The Annals of Statistics*, 33 (4), 1538–1579 . Zhao, P. , & Yu, B. (2006). On model selection consistency of lasso. *Journal of Machine Learning Research*, 77 , 2541–2563 . Zou, H. , & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67 (2), 301–320 .