Julius-Maximilians-Universität Würzburg

Wirtschaftswissenschaftliche Fakultät

Inaugural Dissertation

To obtain the academic degree

Doctor rerum politicarum (Dr. rer. pol.)

Three Essays on the Procurement of Essential Medicines in Developing Countries

Felix Lauton

December 5, 2020

Author:
Felix Lauton, M. Sc.

Skofja-Loka-Str. 1
85354 Freising

1. Supervisor:
Prof. Dr. R. Pibernik

2. Supervisor:
Prof. Dr. R. Bogaschewsky

Acknowledgments

All of this would have never been possible without the support, help and influence of others.

I would like to thank my supervisor Richard Pibernik. During the last couple of years I have learned a lot from him, including, but not limited to, research and academia. I am thankful for the amount of freedom I was given to be able explore the topics I was truly interested in, and I truly enjoyed the discussions we had.

During my time at the University of Würzburg I was closely working together with Alexander Rothkopf, and I am quite certain that I would have not come to this point in time if it was not for him. Although we were not in the same place (in a geographical sense) most of the time, our collaboration could not have been closer and more successful for me. I hope that this will continue in the future, both on a professional and a personal level.

Furthermore, there were many colleagues that became friends that I would like to thank for the great discussions (on- and off-topic) that have made my time as a doctoral student worthwhile: special thanks to Konstantin Kloos, Jan Meller and Fabian Taigel, with whom I shared this journey for the longest time, and also to Christoph Flath, Florian Imgrund, Julian Kurz, Jana Niemeyer, Pascal Notz, Nikolai Stein and Peter Wolf.

Last, but definitely not least, I want to thank my family who always support me. Kathrin, I am deeply thankful for having you in my life. You provide me with the self-esteem and the love that help me through all instances of life. You underestimate the impact you had on this thesis. Furthermore I want to thank my mother, for providing me with the parenting and education that got me this far, my brother, my whole family and all the friends that made me the person that I am today.

Contents

List of Tables

List of Figures

List of Abbreviations

DMPA Depot Medroxyprogesterone Acetate

DP Dynamic Programming

EPO Ex-post Optimal

JRP Joint Replenishment Problem

MDG Millennium Development Goal

PO Purchasing Organization

RT Renewal Theory

SAA Sample Average Approximation

SDG Sustainability Development Goal

UN United Nations

UNFPA United Nations Population Fund

USAID United States Agency for International Development

USFDA US Food and Drug Administration

WDI William-Davidson-Institute

WHO World Health Organization

Deutschsprachige Zusammenfassung

Im Jahr 2000, während des *United Nations Millennium Summit*, wurden die *Millenium Development Goals* beschlossen, eine Agenda von acht Zielen bis 2015, von denen drei einen direkten Fokus auf Gesundheit haben (Sachs and McArthur, 2005; WHO, 2015). Bis 2015 wurden die Ziele nicht vollständig erreicht, trotzdem war der Fortschritt beachtlich: Weltweit konnten HIV, Tuberkulose und Malaria zurückgedrängt und Kinder- und Muttersterblichkeit stark gesenkt werden (WHO, 2015). Im Anschluss wurden 2015 die *Sustainablity Development Goals* beschlossen, welche bis 2030 Gültigkeit haben und nun aus 17 Zielen, immer noch mit starkem Fokus auf Gesundheit, bestehen.

Einen großen Einfluss auf das Erreichen dieser Ziele hat dabei der Einkauf von Medikamenten und medizinischen Hilfsgütern, um bei den meist restringierten Budgets mehr Menschen zu versorgen und durch eine bessere Auswahl der Lieferanten und Hersteller Lieferungen pünktlich und in angemessener Qualität an ihr Ziel zu bringen. Organisationen wie die WHO, UNICEF, USAID oder der Global Fund veröffentlichen und aktualisieren deshalb auch regelmäßig ihre Einkaufsrichtlinien und -prinzipien. Abhängig von Nachfrage, Budget, Produkt und Herstellern sind Einkäufer mit unterschiedlichen Problemstellungen konfrontiert, welche jeweils individuelle Lösungen erfordern. Im Unterschied zu Einkaufsentscheidungen in der Industrie, zu denen bereits eine breite Palette an Forschung existiert, sind die Zielsetzungen für humanitäre Organisationen oft anders (zum Beispiel anstelle von Gewinnmaximierung die Maximierung der Anzahl geimpfter Menschen). Deshalb untersuche ich in dieser Arbeit drei unterschiedliche, dem Einkauf zuzuordnende, Problemstellungen aus dem Bereich "Global Health".

Der erste Teil, welcher eine gemeinsame Arbeit mit Dr. Alexander Rothkopf und Prof. Dr. Richard Pibernik entstanden ist, ist durch eine Studie motiviert die Entscheidungsträgern zweier Einkaufsorganisationen beim Einkauf von Depot Medroxyprogesterone Acetate (DMPA), einem länger wirkenden Verhütungsmittel, unterstützen sollte.

Zum Zeitpunkt der Studie stand den Organisationen nur eine qualifizierter Lieferanten zur Verfügung. Ein zweiter Zulieferer stand kurz vor der Zulassung. Ziel der Arbeit war es den Mehrwert des neuen Lieferanten zu quantifizieren und die optimale Aufteilung der Bestellmengen zwischen beiden Lieferanten zu ermitteln. Hierbei spielt die richtige Abwägung von Preisen (getrieben durch den Wettbewerb zwischen beiden Lieferanten) und Risiko (getrieben durch unsichere Lieferzeiten und Ausfallwahrscheinlichkeiten) eine entscheidende Rolle. In unserer Arbeit zeigen wir wie sich die optimale Aufteilung anhand diverser Parameter, wie Lieferzuverlässigkeit, Kosten und Kapazität, verändert, und untersuchen die Abwägungsentscheidung zwischen Wettbewerb und Risiken.

Im zweiten Teil, der ebenfalls eine gemeinsame Arbeit mit Dr. Alexander Rothkopf und Prof. Dr. Richard Pibernik ist, untersuchen wir einen innovativen Einkaufsmechanismus den wir "Postponement Tender" nennen. Das zugrundeliegende Problem ist das eines Einkäufers, welcher mit der unsicheren Qualität eines neuen Lieferanten konfrontiert ist, und der daraus resultierenden Allokationsentscheidung zwischen bestehendem und neuen Lieferanten. Anstatt alles auf einmal zu vergeben, kann der Einkäufer auch zuerst einen Teil des Volumens an beide Lieferanten vergeben um die unsichere Qualität des neuen Lieferanten besser einzuschätzen. Nachdem die Lieferanten die ersten Volumina geliefert haben kann der Einkäufer die Qualität der Lieferanten besser beurteilen und kauft dann die nachgelagerte Menge vom besseren Lieferanten. Da die Lieferanten bereits zu Beginn Preise festlegen müssen, kann der Einkäufer durch diesen Mechanismus sowohl durch verbesserten Wettbewerb profitieren als auch von einem niedrigeren Qualitätsrisiko, da er dabei etwas über die Qualität der Lieferanten lernt. Wir zeigen in einer detaillierten Analyse wie, abhängig von Einkaufs- und Wettbewerbssituation, die Aufteilung zwischen dem ersten und dem zweiten Teil erfolgen sollte und unter welchen Bedingungen der "Postponement Tender" besser als eine klassische Einzelquellenbeschaffung ist.

Der dritte Teil ist durch den Business Case Kenianischer Apotheken motiviert: diese können durch die Koordination von Bestellungen niedrigere Einkaufspreise aufgrund von Mengenrabatten bei Lieferanten erzielen sowie fixe Bestellkosten wie Logistikkosten teilen. Aufgrund einer Vielzahl von Produkten ist diese Koordination allerdings sehr komplex und mit einem hohen Aufwand sowie Kosten verbunden. Um diese Hürde zu überwinden entwickle ich eine neue, datengetriebene Bestellpolitik für mehrperiodische Bestandsmanagementprobleme mit mehreren Produkten und Skaleneffekten in fixen sowie variablen Bestellkosten. Die entwickelte Politik beruht auf den Prinzipien von Erneuerungstheorie und Sample Average Approximation. Desweiteren analysiere ich die Performance dieser Politik anhand realer Daten aus dem zugrundeliegenden Business Case. In einer ersten Auswertung zeigt sich, dass die resultierenden Kosten nah an denen der theoretisch optimalen Bestellpolitik liegen. Weiter wird gezeigt, dass sich das Verhältnis zu ex-post optimalen Kosten in Szenarien in denen es keine theoretisch optimale Bestellpoltik gibt (mehrere Produkte und Mengenrabatte) im selben Rahmen befindet wie in Szenarien mit optimaler Bestellpolitik. Insgesamt zeigt der entwickelte Ansatz viel Potential für die Lösung komplexer Bestandsplanungsprobleme.

Chapter 1

Introduction

In 2000, during the United Nations (UN) Millennium Summit, 147 heads of state adopted the Millennium Development Goals (MDGs) (Sachs and McArthur, 2005), an agenda of eight goals to achieve by 2015, of which three focused directly on health (WHO, 2015). By 2015, "progress on the three health goals and targets [was] considerable. Globally, the HIV, tuberculosis (TB) and malaria epidemics were 'turned around', [and] child mortality and maternal mortality decreased greatly (53% and 44%, respectively, since 1990), despite falling short of the MDG targets" (WHO, 2015, p. 3). In 2015 the UN adopted the Sustainability Development Goals (SDGs), now consisting of seventeen goals, which have a large focus on health-related topics.

To achieve these goals and supply a large number of people with the right medicine and medical equipment, procurement costs must be kept low because of constrained budgets. According to Barraclough and Clark, "given the limited budgets of virtually all health programs, pharmaceutical procurement costs must be a concern of all policy makers, senior officials, essential medicines program managers, and procurement staff" (Management Sciences for Health, 2013, p. 18.5). Initiatives must be sustainable, as "health care is increasingly expensive [because of] growing population, increasing health care standards, new medicines offering better therapeutic perspectives, [and] modern therapeutic techniques requiring more expensive equipment and staff with special training" (Chisale, M., 2017). Procurement costs can be lowered by, for example, using economies of scale (e.g., buying large quantities, negotiating long-term procurement contracts) or introducing more competition between suppliers (e.g., incenting new suppliers to build capacity or using advanced procurement mechanisms). Global health organizations like the WHO, UNICEF, USAID, and the Global Fund regularly publish and update their procurement principles. Their efforts have led to the implementation of advanced procurement mechanisms like pooled procurement, where individual entities share information (e.g. about suppliers or products), coordinate their procurement activities, and procure their supplies together (WHO, 2019). Depending on the demand, budget, product, and supplier market, each procurement situation can be different, and buyers face difficulties deciding on the adequate volume allocation, procurement mechanism, and order timing. Focusing only on single sourcing and low prices can drive suppliers out of the market, which can create supply shortages, as UNICEF experienced in the measles vaccine market in early 2000 (USAID, 2014, p. 49), when suppliers left the market because of high price pressure. Therefore, procurement organizations want to incent new suppliers to join the market and build capacity to increase competitive pressure for existing suppliers, but they must also ensure that these markets stay healthy and sustainable in terms of supply.

However, a focus on low procurement costs can introduce the potential for uncertain quality, increased volatility in lead times, and higher default risk, because "the limited funds available are frequently spent on ineffective, unnecessary, or even dangerous medications." (Management Sciences for Health, 2013, p. 16.2). For example, in 2008, after awarding a contract for supply of an anti-malarial to the lowest bidder, Kenya's suppliers often saw stock-outs (Tren et al., 2009). Therefore, work is needed to identify and set up the right mechanisms, such as the WHO's Prequalification of Medicines Programme (FM't Hoen et al., 2014). Finding the right balance of decreasing procurement costs (budget utilization) while maintaining a high level of quality and availability (effective supply) is difficult, as it requires the right volume allocation among available suppliers, optimal procurement mechanisms, and optimal order timing. To provide decision-makers and stakeholders in the global health industry with the tools and insights needed to improve access to essential medicines in low-income countries, this thesis considers the three procurement problems of the right volume allocation among available suppliers, optimal procurement mechanisms and an optimal order timing using rigorous analytical and numerical methods of analysis.

The first problem, discussed in Chapter 2, is that of the optimal volume allocation in procurement. The choice of this problem was motivated by a study whose objective was to support decision-making at two procurement organizations for the procurement of Depot Medroxyprogesterone Acetate (DMPA), an injectable contraceptive.[1] At the time of this study, only one supplier that had undergone the costly and lengthy process of WHO pre-qualification was available to these organizations. However, a new entrant

[1]This chapter is based on Lauton et al. (2019) and is a joint work with Alexander Rothkopf and Richard Pibernik.

supplier was expected to receive WHO qualification within the next year, thus becoming a viable second source for DMPA procurement. When deciding how to allocate the procurement volume between the two suppliers, the buyers had to consider the impact on price as well as risk. Higher allocations to one supplier yield lower prices but expose a buyer to higher supply risks, while an even allocation will result in lower supply risk but also reduce competitive pressure, resulting in higher prices. Our research investigates this single- versus dual-sourcing problem and quantifies in one model the impact of the procurement volume on competition and risk. To support decision-makers, we develop a mathematical framework that accounts for the characteristics of donor-funded global health markets and models the effects of an entrant on purchasing costs and supply risks. Our in-depth analysis provides insights into how the optimal allocation decision is affected by various parameters and explores the trade-off between competition and supply risk. For example, we find that, even if the entrant supplier introduces longer leads times and a higher default risk, the buyer still benefits from dual sourcing. However, these risk-diversification benefits depend heavily on the entrant's in-country registration: If the buyer can ship the entrant's product to only a selected number of countries, the buyer does not benefit from dual sourcing as much as it would if entrant's product could be shipped to all supplied countries. We show that the buyer should be interested in qualifying the entrant's product in countries with high demand first.

In the second problem, presented in Chapter 3, we[2] explore a new tendering mechanism called the postponement tender, which can be useful when buyers in the global health industry want to contract new generics suppliers with uncertain product quality. The mechanism allows a buyer to postpone part of the procurement volume's allocation so the buyer can learn about the unknown quality before allocating the remaining volume to the best supplier in terms of both price and quality. We develop a mathematical model to capture the decision-maker's trade-offs in setting the right split between the initial volume and the postponed volume. Our analysis shows that a buyer can benefit from this mechanism more than it can from a single-sourcing format, as it can decrease the risk of receiving poor quality (in terms of product quality and logistics performance) and even increase competitive pressure between the suppliers, thereby lowering the purchasing costs. By considering market parameters like the buyer's size, the suppliers' value (difference between quality and cost), quality uncertainty, and minimum order volumes, we derive optimal sourcing strategies for various market structures and explore how competition is affected by the buyer's learning about the suppliers' quality through the initial volume.

[2]This chapter is a joint work with Alexander Rothkopf and Richard Pibernik.

Chapter 4 considers the repeated procurement problem of pharmacies in Kenya that have multi-product inventories. Coordinating orders allows pharmacies to achieve lower procurement prices by using the quantity discounts manufacturers offer and sharing fixed ordering costs, such as logistics costs. However, coordinating and optimizing orders for multiple products is complex and costly. To solve the coordinated procurement problem, also known as the Joint Replenishment Problem (JRP) with quantity discounts, a novel, data-driven inventory policy using sample-average approximation is proposed. The inventory policy is developed based on renewal theory and is evaluated using real-world sales data from Kenyan pharmacies. Multiple benchmarks are used to evaluate the performance of the approach. First, it is compared to the theoretically optimal policy — that is, a dynamic-programming policy — in the single-product setting without quantity discounts to show that the proposed policy results in comparable inventory costs. Second, the policy is evaluated for the original multi-product setting with quantity discounts and compared to ex-post optimal costs. The evaluation shows that the policy's performance in the multi-product setting is similar to its performance in the single-product setting (with respect to ex-post optimal costs), suggesting that the proposed policy offers a promising, data-driven solution to these types of multi-product inventory problems.

The remainder of this thesis is structured as follows. Chapters 2, 3 and 4 analyze the three aforementioned problems. Chapter 5 draws conclusions from the findings and provides suggestions for future research. The appendices for Chapters 2, 3, and 4 can be found in Chapters 6, 7, and 8, respectively.

Chapter 2

The Value of Entrant Manufacturers: A Study of Competition and Risk for Donor-Funded Procurement of Essential Medicines

Global-health purchasing organizations (POs) want to increase access to essential medicines in income countries. One way to purchase more medicines with limited funds is to contract with generics manufacturers, thereby increasing competition and lowering prices. However, many POs fear that these entrants are less reliable than others and increase supply risks: failure to adhere to lead times and supplier defaults may cause disruptions that result in unsuccessful medical treatments. The problem can be remedied or at least reduced if POs have a sound basis for assessing manufacturers. To this end, we develop a mathematical framework that supports decision-makers in an integrated evaluation of an entrant's effect on purchasing costs and supply risks. Our approach accounts for the characteristics of donor-funded global-health markets and the particular tasks and specific challenges of POs in these markets. More specifically, our approach enables a PO to quantify a potential entrant's value depending on important characteristics of the incumbent and the entrant manufacturer. We use data from a project for donor-funded procurement of Depot Medroxyprogesterone Acetate (DMPA) of two large POs. Our results show the feasibility of our approach for POs, manufacturers, and philanthropic investors in the global-health domain, and we explore the trade-off between competition and supply risks and provide insights into how the entrant's value is affected by parameters like production costs, capacity, lead time and default risk, and in-country registration.

2.1 Motivation

Researchers and practitioners have discussed extensively the pros and cons of single-sourcing and multiple-sourcing with respect to such issues as risk exposure, competition, production costs and overall purchasing costs. However, most of the academic research has focused on individual aspects of the single-versus-multiple-sourcing problem. For instance, researchers have addressed managing a company's risk exposure (e.g. Tomlin and Wang, 2005; Tomlin, 2006) and the impact of competition on prices and purchasing costs (e.g. Perry and Sákovics, 2003; Gong et al., 2012). Addressing these individual parts of the problem has led to rich and complicated formal analyses that have often yielded very interesting results from an academic viewpoint, but decision-makers in practice usually require more comprehensive answers: They want to know how many manufacturers they should contract, how they should split the procurement volume among these manufacturers, and what procurement mechanism they should employ in order to strike an optimal balance between purchasing costs and supply risks.

One domain in which these issues are particularly pertinent is the donor-funded global-health market, where a significant portion of the procurement of essential medicines (e.g., medicines to treat malaria, HIV/AIDS, tuberculosis; reproductive health products; and a variety of vital vaccines) is carried out by global purchasing organizations (POs), such as the United Nations Children's Emergency Fund (UNICEF) and the Global Fund to Fight HIV/AIDS, Tuberculosis, and Malaria. These POs consolidate the demands of low-income countries, negotiate favorable terms with pharmaceutical manufacturers, and take an active role in ensuring that the medicines they purchase reach the population. With few exceptions, these POs are allowed to procure only from manufacturers that undergo for each drug a strict quality assurance process known as "WHO-Prequalification" (WHO, 2016) and/or that are accredited by large stringent regulatory authorities like the US Food and Drug Administration (USFDA). As a result, only one or two manufacturers are pre-qualified, and these are usually branded manufacturers that do not provide generics.[1] In addition, manufacturers must register their products in the low-income

[1] For example, Pfizer is the only pre-qualified manufacturer for the injectable contraceptive Depot Medroxyprogesterone Acetate (DMPA), a key reproductive health product; only Bayer and Merck cur-

countries. The in-country registration and the WHO-prequalification process can be complex and may take more than a year, depending on the manufacturer's experience.

Clearly, such monopoly or duopoly situations diminish the POs' bargaining power, so they are interested in increasing the number of manufacturers that are pre-qualified for a particular drug. However, obtaining pre-qualification is difficult, time-consuming, and costly. When manufacturers apply for pre-qualification, they do not know how much volume they will be awarded, if any, or whether the qualification process and any investment in capacity will pay off. What they do know is that the (monopoly) incumbent has a strong position and is likely to take measures to defend its position—before, during, and after the new manufacturer's entry.

The POs and other stakeholders in the global-health domain can incent potential entrants to pursue pre-qualification and to invest in manufacturing capacity by, for example, providing financial or management support and/or promising to procure certain volumes from the entrant. However, whether a PO should undertake such measures raises the question that lies at the heart of our study: How much value does a new entrant provide to the PO? Answering this question is not trivial: First of all, the PO needs to gauge how the entry of a new manufacturer will impact the purchasing prices. Clearly, prices should decline with increased competitive pressure, however, the ability to lower prices depends on the differential in costs and capacities between the incumbent and the entrant(s).

Despite the potential benefits of additional entrants, POs are often hesitant to employ generics manufacturers, even if they are already pre-qualified or will obtain it in the future. This reluctance is due primarily to supply risks, which POs often perceive as being higher than the supply risks of the incumbent (branded) manufacturers that have an established track-record for delivery performance. The perception of a greater supply risk for new generics manufacturers has been fueled by a number of incidents in which manufacturers caused supply disruptions because they did not meet lead-time expectations, had temporary production/supply outages, lost their pre-qualification, or defaulted entirely.[2]

The inability to quantify the effects of a new entrant on purchasing costs, the higher (perceived) risk, and other hard and soft factors (e.g., higher transaction costs, higher personal effort, a long-standing relationship with the incumbent) often lead POs to

rently hold pre-qualifications for contraceptive implants, another key reproductive health product; and only Alkem Laboratories and Rodael Laboratories have pre-qualifications for Zinc Sulfate, which is important for treating diarrhea (WHO, 2016).

[2]For example, WHO suspended pre-qualification for manufacturers of yellow fever vaccine (UNICEF Supply Division, 2013) and a manufacturer of HIV medicines (FM't Hoen et al., 2014).

stay with the incumbent manufacturer rather than incenting potential entrants to seek pre-qualification and entering into purchasing contracts with them. Therefore, potential entrants, especially generics manufacturers, are even less incented to make the investment required to enter the donor-funded market.

This problem can be remedied (or at least reduced) only if POs have a sound basis for assessing a new manufacturer's value—that is, only if they can evaluate and trade-off the effects of competition on purchasing costs and the effects in terms of supply risks. The objective of the research we present in this paper is to provide POs in the global-health domain with rigorous decision support for such assessments. We propose an approach that supports an integrated evaluation of an entrant's effect on purchasing costs and supply risks depending on the volume split a PO chooses between its suppliers. Our approach accounts for the characteristics of donor-funded global-health markets and the particular tasks and specific challenges of POs in these markets. More specifically, our approach enables a PO to quantify a potential entrant's value depending on important characteristics of the incumbent and the entrant manufacturer. We use a case study to demonstrate our approach's applicability and to highlight how it can provide effective decision support for a PO.

Our research contributes to the comparatively new literature stream that seeks to evaluate competition benefits *and* supply risks jointly. We present an integrated approach to evaluating competition benefits and supply risks that depends on how the PO splits the volume between an incumbent and an entrant manufacturer. Our approach enables us to determine the optimal volume split and to quantify the value of a potential new entrant based on that split. The proposed approach is also novel from a methodological point of view: We extend the concept of bilateral bargaining to account for volume splitting. This allows for an adequate representation of the negotiations between the PO, the incumbent, and the entrant, and enables us to study how alternative volume splits affect competition.

The remainder of our paper is organized as follows. Section 2.2 introduces the example case that motivates our research and develops a set of detailed research questions to guide our analyses. After a brief literature review in Section 2.3, Section 2.4 provides a formal characterization of the problem the PO faces. More specifically, Section 2.4 presents two related sub-models that provide a realistic representation of the interactions among the PO, the manufacturers (an incumbent and a new generics manufacturer), and the recipient countries. To show the applicability of our approach and to derive meaningful results for the PO, we implement our mathematical models in a simulation tool and carry out extensive analyses based on real-world data from several sources. The simulation model, our analyses, and insights and recommendations for the POs are described in Section 2.5.

2.2 Practical Setting and Research Questions

2.2.1 The DMPA Case

We use a case example to develop our approach and to highlight its applicability. In 2015 we carried out a study to support decision-making at two POs, the procurement divisions of the United Nations Population Fund (UNFPA) and the United States Agency for International Development (USAID), both of which procure the majority of donor-funded Depot Medroxyprogesterone Acetate (DMPA), an injectable contraceptive, for low-income countries. At the time of our study, only one manufacturer, Pfizer, held the necessary pre-qualification for providing the drug, although it seemed likely that a generics manufacturer would obtain pre-qualification for DMPA in the near future. To address the problems associated with an entrant manufacturer and to support the POs in their procurement decisions regarding DMPA, we analyzed the competition and the risk effects associated with the new entrant and quantified its value for the PO dependent on parameters that were not perfectly known to the PO at the time (e.g., the entrant's capacity, the cost differential between incumbent and entrant, the entrant's lead time).

POs are intermediaries that consolidate demand from multiple low-income countries and negotiate prices and delivery terms with the manufacturer(s) on behalf of the recipient countries. For example, UNFPA consolidates the demand from the sixty-nine countries that the global-health community deemed to be focus countries for reproductive health efforts in 2012 (FP2020, 2016). POs also take an active role in coordinating supply from the manufacturer(s) with demand from individual countries. Therefore, we can broadly distinguish two main tasks of a PO: negotiation and coordination. Because these tasks are central to our analysis, we provide a detailed description of the decisions and the mechanics involved in the *negotiation phase* and in the *coordination phase*.

Negotiation phase: Currently, with only one DMPA manufacturer, there is essentially no competitive element to the negotiations between the incumbent manufacturer and a PO, leaving little room for bargaining. However, this changes when a second manufacturer receives pre-qualification. We conducted interviews with experts in the field in order to determine how negotiations with two (or more) manufacturers are typically carried out and reviewed publicly available information on extant award mechanisms in similar settings. We found that negotiations are usually conducted in multiple rounds of bilateral bargaining with the manufacturers: The PO determines, a priori, feasible options for splitting the volume between the incumbent and the entrant and negotiates the price in several rounds, during which the PO and the manufacturers exchange price offers until they reach an agreement. Of course, a manufacturer's willingness to quote lower prices depends on its cost structure and the manufacturer's and its competitor's capacities. Section 2.4.1 develops a formal model that captures the dynamics of the negotiation phase and its outcomes.

Coordination phase: During the contractual period, countries send their orders to the PO, the PO forwards each order individually to the manufacturers, and the manufacturers ship the products directly to the countries. The PO's main task is to ensure that each (donor-funded) country's orders reach the manufacturer(s) and that the products arrive in the countries; that is, the PO keeps track of orders and shipments. The PO also takes a coordinating role in case of supply shortages, as it knows when shipments are delayed and when a country is likely to face a supply shortage. Short-term supply disruptions may occur, for example, because of uncertain lead times, temporary production outages, and quality problems related to individual batches. Longer-term supply shortages may be caused, for example, by WHO withdrawing a pre-qualification or manufacturer's bankruptcy.

Especially in case of short-term supply disruptions, a PO can use its knowledge about pipeline inventory (completed orders at the manufacturers' sites or orders that are en-route to other countries but are not yet due in these countries) to re-route available units and avoid shortages.[3] With more than one manufacturer, this re-routing increases in complexity but also increases flexibility, depending on the in-country registration of the entrant's product. In-country registration is a pre-requisite for a manufacturer to be allowed to distribute a pharmaceutical product in a particular country. Typically, an entrant's product is not registered in all countries, as doing so is costly and time-consuming. Section 2.4.2 provides a formal model that captures the dynamics of the coordination phase and its outcomes.

The outcomes of the negotiation phase and the coordination phase are linked by one key variable: The volume split between the incumbent and the entrant. The volume split drives how competition and prices evolve in the negotiation phase and determines the PO's exposure to supply risks—in terms of how much the PO may be exposed to disruptions caused by the manufacturers and the PO's means to mitigate disruptions' negative effects through re-routing in the coordination phase.

[3]Note that the POs do not carry buffer inventory, but only take on a coordinating role to avoid shortages.

A PO has a straightforward overall objective to maximize within the constraints of its budget the number of successful treatments in countries that are eligible for donor funding. Demand for most of the essential medicines of interest is higher than what can be procured and supplied with donor funds, so the outcome of the negotiation phase has a direct impact on the overall objective: The lower the average per-unit price is, the more patients can be supplied within a given budget. The average per-unit price depends on how competition evolves between manufacturers, which is largely determined by how a PO splits the volume and factors like the manufacturers' production costs and capacity.

In the coordination phase, the POs are responsible for preventing supply shortages in the recipient countries. For any essential medicine that patients must receive on a schedule, late deliveries and longer-term disruptions put health outcomes at risk. If a pharmaceutical product is needed but not available, patients face the adverse consequences of symptoms (e.g., in the case of malaria treatment), multi-drug resistances (e.g., tuberculosis treatment), or inadequate protection (e.g., for contraceptive pharmaceuticals like DMPA). Therefore, in the coordination phase, the PO wants to minimize the shortages in the recipient countries to achieve the overall objective of maximizing the number of successful treatments. Section 2.4 formalizes the PO's overall objective and shows how the outcomes of the negotiation and coordination phases can be aggregated into a single performance criterion.

2.2.2 Research Questions

Our research supports a welfare-oriented decision-maker in answering a key question: How much value does a new generics manufacturer provide to a PO? Loosely speaking, the entrant provides positive value if it enables the PO to increase the number of successful treatments within the PO's budget. For the time being, we define the entrant's value as

$$\text{Entrant value} = \frac{\text{\# of successful treatments if incumbent and entrant are available}}{\text{\# of successful treatments if only incumbent is available}} - 1. \tag{2.1}$$

The entrant value is therefore the percentage increase in successful treatments if the entrant is available to the PO. We provide a more rigorous formal definition of the entrant's value in Section 2.4.

Our analyses show that an entrant's value is determined by a complex interplay of different variables and decisions in the negotiation and coordination phases. To clarify the effects and structure our analysis, we develop corresponding research questions, each of which deals with specific concerns of decision-makers when a generics entrant may be available. Addressing these questions provides a comprehensive picture of the factors that determine whether an entrant can bring additional value to the PO.

From our presentation in Section 2.1 it is clear that practitioners associate certain benefits and risks with a generics manufacturer that is entering the DMPA market. On one hand, they assume that the entrant will have lower production costs and conventional wisdom suggests that the lower the entrant's costs, the more competitive pressure he induces, and the more attractive it is for the PO to allocate larger shares of the total volume to the entrant. On the other hand, POs expect the entrant to increase supply risks because of increased uncertainty about lead times. For the PO it is important to determine how much value the entrant promises and how the PO should split the procurement volume between the incumbent and the entrant. This leads us to our first and most fundamental research question. QUESTION 1: *What is the value of a lower-cost, higher-risk entrant and how does this value depend on the volume split?*

Additionally, a potential entrant will be careful about investing in production capacity and it is likely that it will not be able to fulfill a PO's entire demand. While lower entrant capacity negatively impacts competition, reducing the entrant's value, it is not clear how constrained entrant capacity impacts the optimal volume split and how much the entrant's value changes in the presence of capacity constraints. Therefore, we ask in QUESTION 2: *What is the value of an entrant with limited capacity and how does this value depend on the volume split?*

An entrant supplier must register its product in the countries that are eligible for procurement via the PO. Because this is a costly and time-consuming endeavor, the entrant's product is not likely to be registered in all of the PO's recipient countries. However, a PO's flexibility in matching manufacturers' supply with countries' demands during the coordination phase depends on the extent to which the entrant's products are registered in the recipient countries. Therefore, practitioners are concerned with the potential impact of limited in-country registration, which motivates our third research question. QUESTION 3: *What is the value of an entrant whose product is not registered in all recipient countries and how does this value depend on the volume split?*

To answer these research questions, we develop in Section 2.4 two interrelated models, a negotiation model and a coordination model, to determine the optimal volume split based on the entrant's characteristics and to derive the entrant's value. Our analysis in Section 2.5 builds on an implementation of these two models and is structured along the lines of our three research questions.

2.3 Literature

Two streams of research are particularly relevant to our work: research on competition among manufacturers and research on managing operational drivers of risk, such as uncertainty regarding lead times and manufacturer defaults. We structure our literature review according to these two research streams.

Our study considers how manufacturers' prices behave in a competitive environment. Perry and Sákovics (2003) and Klotz and Chatterjee (1995) study split-award auctions in contexts that include entry costs for manufacturers, and both show that, if the number of manufacturers is fixed, bids increase when the procurement volume is split equally.

These and other papers such as Anton and Yao (1989) and Gong et al. (2012) have in common that they identify manufacturer bids that are decreasing if the buyer is willing to allocate more volume to a single manufacturer, which implies a concave cost function for the buyer, where costs are at a maximum if the buyer splits the procurement volume evenly. Our setting differs from this extant research, as it is characterized by manufacturers with limited capacity such that a single manufacturer typically can not satisfy the buyer's entire demand. Ausubel (2004) shows that, in such a setting, manufacturers have a minimum volume that they are certain to sell, so these units are excluded from competition because the manufacturers know they can sell them at the buyer's reservation price. Therefore, the available production capacity can be viewed as a major driver of competition.

Important to our study is the theoretical work that addresses bargaining and negotiations, as we model bilateral negotiations between a buyer and two manufacturers. Binmore et al. (1986) study a two-person bargaining game in which the players make alternating offers until they reach an agreement. The authors propose models that incorporate either the bargainer's impatience or their fear of not coming to an agreement and find that this impatience or fear weakens the bargainer's position. Horn and Wolinsky (1988) study a setting that is closely related to ours, where one firm engages multiple other firms in bilateral negotiations. The authors introduce, as one of the first, a bilateral Nash bargaining model (also referred to as Nash-in-Nash (Collard-Wexler et al., forthcoming in 2018; Moellers et al., 2017) or Nash-Nash bargaining (Feng and Lu, 2013)), where each negotiation between two firms is modeled by a Nash bargaining game, and disagreement points describe individual utilities in the case of disagreement. They argue that this approach appropriately describes a negotiation's outcome with alternating offers which can oftentimes be observed in real-world settings (Horn and Wolinsky, 1988, p. 411). Feng and Lu (2013) apply the concept of bilateral Nash bargaining in a supply chain contracting context and compare it to the outcome of a Stackelberg game. They suggest that the "Stackelberg game may not be a sufficient device to predict supply chain contracting behaviors in reality where bargaining is commonly observed" (Feng and Lu, 2013, p. 661). Our model extends the bilateral bargaining models because we formulate a bargaining model that accounts for a buyer's volume allocation decision between two manufacturers. We also use a new approach to model the bargaining position: Nash bargaining models usually use an abstract parameter that indicates how negotiation power is distributed between two negotiating parties (see e.g. Heese (2015)). In contrast, we explicitly account for the effect of a supplier's production capacity on its bargaining position, which we use as an implicit measure of bargaining power. Explicitly modeling the impact of limited capacity is also motivated by Ausubel (2004)'s work, which captures the minimum shares that a manufacturer is certain to obtain if the competitor's capacity is limited.

An extensive body of literature on lead-time uncertainty has emerged. Kouvelis and Li (2008) consider the effect of lead-time uncertainty on volume-splitting between two manufacturers. The authors study a replenishment decision under constant demand in which the buyer has the option, after ordering from a manufacturer with stochastic lead time, to use a flexible back-up manufacturer if orders arrive late. Their results suggest that the benefits of dual sourcing are enhanced when lead-time uncertainty is high. Ramasesh et al. (1991) analyzes the total expected cost of a stochastic lead-time (s,Q) inventory model with the option of dual sourcing and conclude that, when uncertainty is high and ordering costs are low, dual-sourcing can be the cost-optimal option. Chiang and Benton (1994) consider how dual sourcing influences the performance of inventory models and find that "except for cases where the ordering costs are high, the lead-time variability is low, or the customer service level is low, dual sourcing performs better than single-sourcing" (Chiang and Benton, 1994, p. 609). Kelle and Miller (1990) analyze how splitting replenishment orders among vendors influences the expected shortage and suggest that a dual-sourcing policy can improve the service level or reduce the safety stock for a given service level. Kelle and Miller (2001) find that order-splitting can reduce the stock-out risk even if one manufacturer is much less reliable.

Considering manufacturer defaults, Yu et al. (2009) analyze a two-stage supply chain in which a buyer faces price-sensitive demand and can procure goods from a global and a domestic manufacturer that offer different prices and Bernoulli-distributed disruption risks. Tomlin and Wang (2005) study unreliable supply chains with multiple products with Bernoulli-distributed supplier defaults and find that the benefits of dual-sourcing increase as supply-chain reliability decreases (Tomlin and Wang, 2005, p. 51). Tomlin

(2006) focuses on a procurement setting in which a buyer has two capacity-constrained sourcing options: a reliable but expensive manufacturer and an unreliable manufacturer.

Kouvelis and Li (2008), Ramasesh et al. (1991) and Chiang and Benton (1994) state that the buyer's objective function is convex in the allocation decision. Moreover, we can deduce from the findings of all papers on supply risks, that the buyer can benefit from dual sourcing and that shortage costs are convex in the volume split.

In Section 2.4.2 we develop a coordination model to capture (short-term) supply shortages from uncertain supply lead times and manufacturer defaults. Inspired by the papers that investigate supply risks and volume-splitting, we build a model that accounts for the PO's particular coordination tasks. We model a make-to-order system that tracks the PO's (virtual) inventory, captures expected mismatches between countries' demands and manufacturers' supplies, and allows for re-routing in case of shortages.

Finally, some researchers merge the economic perspectives of competition and operational risk to analyze the potential trade-offs in a combined approach, which is similar to the objective of this paper. For example, Yang et al. (2012) analyze a setting in which a buyer, who can choose single- or dual-sourcing from two default-prone manufacturers, has incomplete information regarding the manufacturers' reliability. The authors show that, because a buyer loses competition benefits through dual sourcing, it may be inclined to diversify less, even if the manufacturers' reliability decreases. The authors argue that this effect is driven primarily by asymmetric information about risks. Babich et al. (2007b) also investigate the buyer's competition and diversification benefits when the suppliers' defaults are correlated and show that higher default correlations increase competition and lower prices and that these benefits can outweigh the reduction in diversification benefits.

Although they provide insights into the trade-off between competition and supply risks, these papers do not address the practical issue of volume splitting resulting in concave procurement costs; that is, the finding that a buyer should favor a winner-take-all allocation over splitting volume when the buyer considers only procurement costs. Research on supply risks indicates that volume-splitting typically results in a diversification benefit – that is, that costs are convex with respect to the split. The papers that consider both competition and risk propose stylized models that provide analytically tractable solutions, but in doing so they make simplifying assumptions that help them avoid the problem of an objective function that could be either convex or concave. In contrast, our research contributes to the literature by providing a close representation of the real-world dynamics of the problem, capturing both the convex nature of risk diversification and the concave nature of competition with respect to the volume allocation.

2.4 Model

In Section 2.2.1 we explained that the PO seeks to maximize the number of successful medical treatments in eligible countries and how the PO's actions in the negotiation and coordination phases impact this objective. Now we formalize the PO's objective in order to operationalize the entrant's value. It is more convenient for our analysis to use as a criterion the number of treatments that were not delivered successfully during a contractual period instead of the number of successful treatments. Accordingly, we assume that a PO wants to minimize what we term the *total expected shortage* for a contractual period (e.g., one year). Total expected shortage measures the expected number of treatments that the PO can not deliver either because of an insufficient budget or because of supply shortages in the country. We denote the total expected shortage by $\xi(\alpha)$; it depends on the entrant's share α. The incumbent's share is therefore $(1-\alpha)$.[4]

Total expected shortage $\xi(\alpha)$ has two elements: the *price-induced shortage* ($\xi^p(\alpha)$) and the *risk-induced shortage* ($\xi^r(\alpha)$). Loosely speaking price-induced shortage is the procurement volume that the PO can not buy, because prices are too high. To be more rigorous, let q be the target volume (measured in treatments) for a particular essential medicine across all countries that are eligible for donor funding from the PO. The target volume is exogenous to our model and usually stems from an extensive forecasting effort of the PO which can include demographic trends, epidemiological developments and other drivers. Let b denote the PO's budget for this product. We assume that a PO's budget is insufficient to satisfy the target volume at the current (reservation) price r. The outcome of the negotiation phase is a weighted average per-unit price that the PO pays per treatment. We denote the weighted average per-unit price by $\bar{p}(\alpha)$. The price-induced shortage can then be expressed as $\xi^p(\alpha) = q - \frac{b}{\bar{p}(\alpha)}$, which captures the fact that, with a lower weighted average per-unit price $\bar{p}(\alpha)$, the PO can buy more treatments ($\frac{b}{\bar{p}(\alpha)}$) with the same budget b, reducing some of the gap between q and $\frac{b}{\bar{p}(\alpha)}$. Adding the risk-induced shortage $\xi^r(\alpha)$, which we measure as the expected shortage caused by supply disruptions, dependent on the volume split, gives the total expected shortage $\xi(\alpha)$. The PO wants to choose α to minimize $\xi(\alpha)$. We denote the minimal total expected shortage as $\xi(\alpha^*)$.[5] The formalization of total expected shortage also

[4] Table 6.2 in the appendix provides an overview of all variables of the competition and the coordination model.

[5] Without loss of generality we assume throughout the paper that if a PO finds that more than one

allows us to provide a more rigorous measure of the entrant's value (see Eq. (2.1)):

$$w = \frac{\xi_{\text{sole-sourcing}} - \xi(\alpha^*)}{\xi_{\text{sole-sourcing}}}. \tag{2.2}$$

The entrant's value w is the difference between the total expected shortage in case of sole-sourcing from the incumbent—that is, the product can be sourced only from the incumbent—and the total expected shortage for the optimal volume split between the incumbent and the entrant, normalized to the total expected shortage in case of sole-sourcing. The ratio w captures the relative increase in the number of successful treatments a PO can expect to achieve if the entrant enters the market and is available to the PO. Clearly, even if $\alpha^* = 0$, the entrant may provide positive value to the PO if the competitive pressure of the entrant's presence reduces the incumbent's price. Note that in our presentation we differentiate sole-sourcing, which refers to when the entrant is not available, from single-sourcing, which refers to when the entrant is available and the PO decides to allocate the entire volume either to the incumbent or the entrant. To evaluate the entrant's value w we proceed as follows: we first obtain estimates of the prices of the incumbent and the entrant depending on α. In the next section we develop a negotiation model that allows us to estimate these prices for any feasible α. Note that α represents a relative volume split; the absolute volumes allocated to the entrant ($\alpha \frac{b}{\bar{p}(\alpha)}$) and the incumbent ($(1-\alpha)\frac{b}{\bar{p}(\alpha)}$) depend on the negotiated prices and the budget of the PO. They are, thus, an outcome of our negotiation model. Knowing the absolute volume allocations to the incumbent and the entrant for a specific α allows us to estimate the associated (expected) risk-induced shortage. In Section 2.4.2 we introduce a coordination model to evaluate the risk-induced shortage for a given α (and the associated volume allocations derived from the negotiation model). Hence, we solve the two interrelated models in a sequential fashion and account for the fact that the risk-induced shortage (estimated by the coordination model) depends on the outcome of the negotiation model. Clearly, one can argue that the prices determined in the negotiation model should already account for the supply risks of the incumbent and the entrant. From our coordination model, however, it will become clear that these effects are intricate and can hardly be approximated. Simultaneously solving the competition and the coordination model is analytically intractable. For this reason, we believe that it is most appropriate to first calculate the prices and absolute volume allocations without explicitly considering the supply risk and to use these results to obtain an estimate of the risk-induced shortage associated with a particular volume split.

---

split minimizes total expected shortages he chooses the split with the highest allocation to the incumbent.

2.4.1 Negotiation

In this section we propose a model to capture the PO's negotiation with the incumbent (I) and the entrant (E). This model enables us to obtain the weighted average per-unit price $\bar{p}(\alpha)$ and the price-induced shortage $\xi^p(\alpha) = q - \frac{b}{\bar{p}(\alpha)}$ for a given split α.

We model the outcome of the negotiation as a bilateral Nash bargaining model between a PO and two manufacturers. In his seminal work Nash proposed that the result of a negotiation between two individuals can be modeled as a problem in which the product of the two individual utilities are maximized. The two parties seek to increase their utility by increasing or decreasing the price and in the optimum they reach a mutually acceptable price (Nash, 1950). The Nash bargaining model can be seen "as the reduced form of an appropriate dynamic bargaining model" (Horn and Wolinsky, 1988, p. 411). A bilateral Nash bargaining model is the extension to a negotiation with multiple parties such as a PO and multiple manufacturers. Essentially, a bilateral bargaining model captures the outcome of the underlying negotiation dynamics between a buyer and multiple suppliers including the impact of competition among these manufacturers (Horn and Wolinsky, 1988; Feng and Lu, 2013). In our case the bilateral negotiation begins with the PO announcing the volume allocation he seeks to award to the incumbent and the entrant, that is the split α. Given this split, the PO conducts a multi-round negotiation with the manufacturers to minimize price-induced shortage $\xi^p(\alpha)$ (see Section 2.2.1).

To formulate the Nash bargaining model, we restate the PO's objective and refer to it as the PO's utility: $u(\alpha) := -\xi^p(\alpha) = {}^{b}/_{\bar{p}(\alpha)} - q$, with $\bar{p}(\alpha) = \alpha p_E(\alpha) + (1-\alpha)p_I(\alpha)$ as the weighted per-unit price given a split α. $p_E(\alpha)$ is the entrant's per-unit price and $p_I(\alpha)$ the incumbent's per-unit price. During the negotiation each manufacturer $j \in \{I, E\}$ seeks to maximize its individual profit, $\pi_j(\alpha) = z_j(p_j(\alpha) - c_j)$, where c_j are the per-unit production costs and z_j is the volume sold to the PO. It is important to note that because the PO's budget and the negotiated prices determine how much the PO procures, the PO's procurement volume is endogenously determined by the negotiation outcome, and it differs from the target volume q. While each manufacturer is certain to receive the announced share α (or $1-\alpha$), the absolute size of the volume allocation to each manufacturer is determined by the negotiated prices, as lower prices increase the procurement volume. Therefore, we obtain the incumbent's procurement volume as

$$z_I = (1-\alpha)\frac{b}{\bar{p}(\alpha)} \leq cap_I, \tag{2.3}$$

where cap_I is the production capacity of the incumbent, and the entrant's procurement volume as

$$z_E = \alpha \frac{b}{\bar{p}(\alpha)} \leq \min[cap_E, q^{IE}], \tag{2.4}$$

where cap_E is the entrant's production capacity. The allocation to the entrant can also be limited by the target volume $q^{IE} = q - q^I$ for countries in which the entrant's product is registered.

Before we formulate the bargaining problem, we have to specify the PO's utility and the manufacturers' profits if the negotiation fails. These disagreement points, which are also referred to as status-quo utilities, serve as reference points for the PO and the manufacturers (Horn and Wolinsky, 1988). We assume that the previous market price r, the reservation price, is restored if the negotiation fails and that the manufacturer who stopped the negotiation becomes the residual claimant of the procurement volume. For example, if the PO's negotiation with the entrant fails the PO will procure the entire volume from the incumbent at a price of r per unit. If the incumbent's capacity is not sufficient, that is, $cap_I \leq \frac{b}{r}$, the PO uses the remaining budget to buy units from the entrant at price r, that is, a volume of $\frac{b - r \cdot cap_I}{r} \leq \min[cap_E, q^{IE}]$. To avoid trivial solutions, we assume that $q > \frac{b}{r}$, so the PO is not able to fulfill the target volume at the reservation price given its budget, and $cap_I + cap_E \geq q$, so the total capacity is sufficient to fulfill the target volume. Accordingly, the PO's status-quo utility is $u^{sq} = \frac{b}{r} - q$, that is, the utility if there are no gains from negotiation. The incumbent's status-quo profit is $\pi_I^{sq} = \frac{(b - \min[r \cdot \min[cap_E, q^{IE}], b])^+}{r}(r - c_I)$, and the entrant's status-quo profit is $\pi_E^{sq} = \frac{(b - \min[r \cdot cap_I, b])^+}{r}(r - c_E)$. Therefore, the status-quo profit of manufacturer j depends on its competitor's capacity cap_{-j}. If manufacturer $-j$'s capacity is sufficient to supply the PO's entire volume at the reservation price (i.e., $r \cdot cap_{-j} > b$) manufacturer j's status-quo profit is zero; so it has a weaker bargaining position in the negotiation because manufacturer $-j$'s capacity is sufficient to exhaust the PO's budget at the reservation price. For smaller capacities of manufacturer $-j$ the status-quo profit of manufacturer j becomes positive, indicating a stronger bargaining position.

We are now ready to formulate the bilateral bargaining problem, which seeks to maximize the product of utility/profit functions.

$$\max_{p_E(\alpha)} \quad (u(\alpha) - u^{sq})(\pi_E(\alpha) - \pi_E^{sq}) \tag{2.5}$$

$$\max_{p_I(\alpha)} \quad (u(\alpha) - u^{sq})(\pi_I(\alpha) - \pi_I^{sq}) \tag{2.6}$$

$$\text{s.t.} \quad \pi_j(\alpha) > \pi_j^{sq}, \quad j \in \{I, E\} \tag{2.7}$$

$$u(\alpha) > u^{sq}, \tag{2.8}$$

$$0 < p_j(\alpha) \leq r, \quad j \in \{I, E\}. \tag{2.9}$$

Eq. (2.5) models the bilateral negotiation between the PO and the entrant regarding the entrant's price. It captures that, for a given split α, the PO and the entrant try to agree on a price $p_E(\alpha)$ that maximizes the product of the utility and profit both parties receive (compared to the status-quo utility or profit). A higher price increases the entrant's profit and a lower price increases the PO's utility. Similarly, Eq. (2.6) models the negotiation between the PO and the incumbent. Conditions (2.7), (2.8), and (2.9) ensure that, if the parties negotiate, all players have positive gains from the negotiation compared to their status quo utility or profit.

For reasons that become clear later we do not (yet) include that the absolute volume allocation to each manufacturer must obey the respective capacity constraint given in Eqs. (2.3) and (2.4).

Proposition 1.

Consider the bilateral bargaining problem given in Eqs. (2.5) to (2.9).

a) The feasible set of utility-profit allocations $A_j = \{(u(\alpha), \pi_j(\alpha)) | 0 < p_j \leq r\}$, $j \in \{I, E\}$, is convex.

b) The bilateral bargaining problem has a unique solution. $p_E^*(\alpha)$ and $p_I^*(\alpha)$ are equilibrium prices if they fulfill the following system of equations:

$$p_E^*(\alpha) = \min[r, \frac{2rb(\alpha c_E + (1-\alpha)p_I^*(\alpha))}{\alpha(b(\alpha c_E + (1-\alpha)p_I^*(\alpha)) + r(b - \pi_E^{sq}))} - \frac{1-\alpha}{\alpha} p_I^*(\alpha)], \tag{2.10}$$

$$p_I^*(\alpha) = \min[r, \frac{2rb(\alpha p_E^*(\alpha) + (1-\alpha)c_I)}{(1-\alpha)(b(\alpha p_E^*(\alpha) + (1-\alpha)c_I) + r(b - \pi_I^{sq}))} - \frac{\alpha}{1-\alpha} p_E^*(\alpha)], \tag{2.11}$$

A prerequisite for applying a Nash-bargaining model is to verify that the solution space is convex. Proposition 1a confirms that the solution space of our bilateral bargaining problem is a convex set; thus, the solution to the bargaining problem must be located on the boundary of this set.[6] Proposition 1b shows that unique equilibrium prices exist for the bilateral bargaining problem. These equilibrium prices represent a negotiation outcome for a given volume split α at which all players' utility or profits are balanced. A deviation from these prices would decrease the utility or profit of one player more than it would increase another player's utility or profit. Status-quo utility and profits directly impact these equilibrium prices because they determine each player's bargaining position. Proposition 1 also shows that the PO is not willing to pay a price above the reservation price.

Constraints (2.3) and (2.4) limit the feasible volume allocations to the interval $\alpha \in [\frac{b-cap_I p_I^*(\alpha)}{b-cap_I(p_I^*(\alpha)-p_E^*(\alpha))}, \frac{\min[cap_E,q^{IE}]p_I^*(\alpha)}{b-\min[cap_E,q^{IE}](p_E^*(\alpha)-p_I^*(\alpha))}]$. The results presented in Proposition 1 do not account for these capacity constraints. Because the PO announces a relative volume share α and the volumes of the entrant and incumbent are endogenous (see Eqs. (2.3) and (2.4)), the solutions to the bargaining problem stated in Eqs. (2.5) to (2.9) may be infeasible with respect to the capacity constraints. Explicitly incorporating the capacity constraints would, however, lead to undesirable effects: if one of the capacity constraints was binding, the model would result in higher prices and, consequently, lower procurement volumes in order to satisfy the capacity constraint. Thus, the model would adjust the prices to ensure that the solution is feasible with respect to the capacity constraints. We can show that in these instances it would never be optimal for the PO to choose the corresponding α. The marginal increase in the price-induced shortage always exceeds any possible reduction in the risk-induced shortage for $\alpha \notin [\frac{b-cap_I p_I^*(\alpha)}{b-cap_I(p_I^*(\alpha)-p_E^*(\alpha))}, \frac{\min[cap_E,q^{IE}]p_I^*(\alpha)}{b-\min[cap_E,q^{IE}](p_E^*(\alpha)-p_I^*(\alpha))}]$.[7] Thus, we deal with the capacity constraints as follows: we first solve the bargaining model Eqs. (2.5) to (2.9) without considering Eqs. (2.3) and (2.4) and then omit the solutions for which $\alpha \notin [\frac{b-cap_I p_I^*(\alpha)}{b-cap_I(p_I^*(\alpha)-p_E^*(\alpha))}, \frac{\min[cap_E,q^{IE}]p_I^*(\alpha)}{b-\min[cap_E,q^{IE}](p_E^*(\alpha)-p_I^*(\alpha))}]$.

The equilibrium prices given in Proposition 1 allow us to calculate the price-induced shortage for a given split

$$\begin{aligned}\xi^p(\alpha) &= q - \frac{b}{\alpha p_E^*(\alpha) + (1-\alpha)p_I^*(\alpha)}, \\ &\text{for } \alpha \in [\frac{b - cap_I p_I^*(\alpha)}{b - cap_I(p_I^*(\alpha) - p_E^*(\alpha))}, \frac{\min[cap_E, q^{IE}]p_I^*(\alpha)}{b - \min[cap_E, q^{IE}](p_E^*(\alpha) - p_I^*(\alpha))}].\end{aligned} \tag{2.12}$$

[6] The proof of Proposition 1 can be found in the Appendix.

[7] The price-induced shortage increases at a higher rate in α than the volume of either manufacturer. The risk-induced shortage, however, can at best decrease by the amount of units allocated to this manufacturer, even if a manufacturer exhibits no uncertainty with regard to lead times and defaults. A formal proof of this can be found in the Appendix.

In the following, we restrict price-induced shortage to be strictly positive because a PO will not buy more units than its target volume q. This assumption is motivated by our practical problem of a not-for-profit PO that has limited funding so that for reasonable outcomes of the negotiation he will never be able to satisfy all of the countries' demands. This implies that the prices will not be lower than a certain threshold. To formalize this restriction, let $\hat{c}_j$ denote a lower bound for the manufacturers' production costs. Then, c_j satisfy the following conditions

$$c_E \geq \hat{c}_E \;\&\; c_I \geq \hat{c}_I \text{ such that } \max_{\alpha}\left\{\frac{b}{(\alpha p_E^*(\alpha) + (1-\alpha)p_I^*(\alpha)))}\right\} \leq q. \tag{2.13}$$

Based on our bilateral bargaining model, we can obtain the price-induced shortage ($\xi^p(\alpha)$) dependent on the volume split (α), which constitutes the first important element for quantifying the entrant's value. The next section explains how we determine the second important element, the risk induced shortage ($\xi^r(\alpha)$).

2.4.2 Coordination

The PO not only negotiates with the manufacturers on behalf of the recipient countries but also performs certain tasks to match the manufacturers' supply with countries' demands. The following model allows us to approximate the risk-induced shortage $\xi^r(\alpha)$ dependent on the volume split α, that is, the entrant's share of the total procurement volume. Note, that our model is not intended to optimize the overall system; for example, we do not intend to propose an alternative, improved inventory policy but to reflect how the PO would act as an intermediary between the manufacturers and the recipient countries given the PO's current operations.

Because the negotiation between the PO and the manufacturers takes place before the coordination phase, the target volume q is adjusted by the price-induced shortage $\xi^p(\alpha)$; that is, the actual procurement volume $d(\alpha)$ referred to in the coordination model is defined as $d(\alpha) = q - \xi^p(\alpha)$. Therefore, we only consider the real volume procured by the PO for determining the risk-induced shortage.

Figure 2.1 illustrates the flows of demand information, orders, and supplies among the recipient countries, the PO, and the manufacturers during the coordination phase. In our model, the PO faces two demand streams: one that originates from countries in which only the incumbent's product is registered (I-demand stream), and one that originates from countries in which both the incumbent's and the entrant's products are registered (IE-demand stream). For the latter case, we assume that countries are

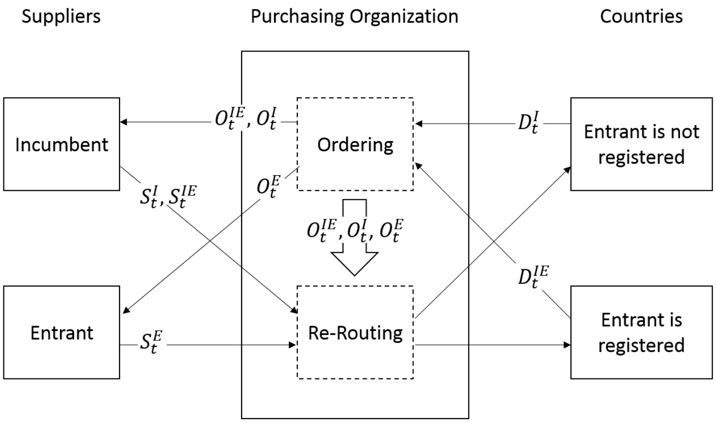


Figure 2.1: Demand, order, and supply information streams in the coordination phase

indifferent to whether they receive shipments from the incumbent or the entrant; that is, the products are perfect substitutes. Countries' order sizes and arrival rates are uncertain. D_t^I denotes the random demand on day $t \in \{0, \ldots, T\}$ from countries in which the incumbent's product is registered, and D_t^{IE} denotes the demand from countries in which the incumbent's and the entrant's products are registered. The planning horizon T is typically one year or two years, that is, the duration of a contract between the PO and the manufacturers. Each day $t \in \{0, \ldots, T\}$, the PO observes the demand realizations $D_t^I = d_t^I$ and transmits to the incumbent a corresponding order realization $O_t^I = o_t^I = d_t^I$. By definition, only the incumbent can fulfill demands d_t^I. For demand realizations $D_t^{IE}{=}d_t^{IE}$, the PO has to decide how much to order from the incumbent and how much to order from the entrant. This decision is driven by the volume split α that was determined during the negotiation phase. Therefore, the PO has to split $D_t^{IE}{=}d_t^{IE}$ ($t \in \{0, \ldots, T\}$) in such a way that α is achieved for the overall procurement volume. Our model accounts for the volume split by assuming that the PO places order realizations $O_t^E = o_t^E = \hat{\alpha} d_t^{IE}$ with the entrant, where $\hat{\alpha} = \alpha \frac{d^I + d^{IE}}{d^{IE}}$. Here, d^I denotes the procurement volume from countries in which only the incumbent's product is registered, and d^{IE} denotes the procurement volume from countries in which both products are registered, such that $d(\alpha) = d^I + d^{IE}$. The orders placed with the incumbent are $O_t^{IE} = o_t^{IE} = (1 - \hat{\alpha}) d_t^{IE}$.

The manufacturers' lead times are uncertain. L_j ($j \in \{I, E\}$) denotes the uncertain lead time for any order the PO places with manufacturer $j \in \{I, E\}$. In addition, let τ denote a lead-time threshold after which an order is considered to be short. An order placed at time t must be fulfilled by $t + \tau$, as any later fulfillment will cause a shortage in the recipient country. Because of the lead-time uncertainty, realizations l_j of the uncertain lead time can be smaller ($l_j \leq \tau$) or larger ($l_j > \tau$) than the shortage threshold. Put differently, the supply required to fulfill an order placed in $t - \tau$ (and that becomes short on day t) may materialize before or after t. S_t^I, S_t^{IE}, and S_t^E denote the uncertain supplies available at the manufacturer's site on day t, where S_t^I and S_t^{IE} represent the uncertain supply from the incumbent for the I-demand stream and the IE-demand stream, respectively, and S_t^E represents the uncertain supply from the entrant for the IE-demand stream.

It is here that the PO's coordinating role comes into play: If on day t, supply is less than the orders that are due, the PO can cover the gap by using "excess" supplies that became available before t but were not yet needed to fulfill country demands. This coordination reflects the PO's re-routing capability. Of course, if excess supplies from previous days are insufficient to cover the gap, the recipient countries will incur a shortage.

To model the PO's "supply-demand coordination" adequately, we introduce the notion of a virtual inventory (V), which captures the pipeline inventory that is available at $t+1$. For example, the virtual inventory for the I-demand stream can be expressed as: $V_{t+1}^I = \left[V_t^I - (O_{t-\tau}^I - S_t^I)\right]^+$ for $t \in \{\tau, \ldots, T-1\}$. If supply on day t is larger than what is required to fulfill orders that are due, the excess units are added to the virtual inventory. Then, if supplies from the incumbent on day t are insufficient to satisfy the orders that are due on this day, the PO can avoid shortages by compensating with supplies that are held in the virtual inventory. Consequently, the virtual inventory characterizes how many units are available for re-routing in the I-demand stream on any day t.

The PO may also consider using virtual inventory from the I-demand stream to avoid shortages in the IE-demand stream. Thus, if shortages on day t occur in the IE-demand stream and the PO can not avoid all of the shortages with virtual inventory in the IE-demand stream (i.e., $[O_{t-\tau}^{IE}(\alpha) - S_t^{IE} + O_{t-\tau}^E(\alpha) - S_t^E - V_t^{IE}] > 0$), the PO can use virtual inventory from the I-demand stream to avoid shortages. Of course, this approach lowers the virtual inventory that is available in the I-demand stream on the next day $t+1$. Therefore, we refine our formulation of V_{t+1}^I:

$$
\begin{aligned}
V_{t+1}^I = & \big[V_t^I - (O_{t-\tau}^I - S_t^I) \\
& - [O_{t-\tau}^{IE}(\alpha) - S_t^{IE} + O_{t-\tau}^E(\alpha) - S_t^E - V_t^{IE}]^+\big]^+ \quad \text{for} \quad t \in \{\tau, \ldots, T-1\}.
\end{aligned}
\tag{2.14}
$$

Eq. (2.14) captures the available virtual inventory that the PO can use to avoid shortages in the I-demand stream on any day t. It is convenient for our exposition also to express the residual virtual inventory after the PO has exhausted the virtual inventory to cover shortages in one demand stream:

$$R_t^I = \big[V_t^I - (O_{t-\tau}^I - S_t^I)\big]^+ \quad \text{for} \quad t \in \{\tau, \dots, T\}. \tag{2.15}$$

Eq. (2.15) expresses the residual virtual inventory in the I-demand stream. Suppose that orders that are due on day t exceed the supplies on this day, but the available virtual inventory is sufficient to avoid the shortages (i.e. $\big[V_t^I - (O_{t-\tau}^I - S_t^I)\big] > 0$). This residual virtual inventory R_t^I is available to the PO to avoid shortages in the IE-demand stream on day t. Note, that we can simplify Eq. (2.14) by substituting R_t^I.

Expressions for the virtual inventory V_t^{IE} and the residual virtual inventory R_t^{IE} that correspond to the IE-demand stream follow a similar structure. First, we express the residual virtual inventory of the IE-demand stream as:

$$R_t^{IE} = \big[V_t^{IE} - (O_{t-\tau}^{IE}(\alpha) - S_t^{IE} + O_{t-\tau}^{E}(\alpha) - S_t^{E})\big]^+ \quad \text{for} \quad t \in \{\tau, \dots, T\}. \tag{2.16}$$

Because of limited in-country registration, the PO can use only the incumbent's product for the IE-demand stream to avoid shortages in the I-demand stream. To separate the units the incumbent produced from those the entrant produced, we introduce the operator $\langle\cdot\rangle$, such that $\langle R_t^{IE}\rangle$ captures only the units the incumbent produced and $R_t^{IE} - \langle R_t^{IE}\rangle$ captures only the units the entrant produced.

Knowing the residual virtual inventory, we can also write the virtual inventory of the IE-demand stream as

$$\begin{aligned} V_{t+1}^{IE} &= R_t^{IE} - \langle R_t^{IE}\rangle \\ &\quad + \Big[\langle R_t^{IE}\rangle - [O_{t-\tau}^I - S_t^I - V_t^I]^+\Big]^+ \quad \text{for all} \quad t \in \{\tau, \dots, T-1\}. \end{aligned} \tag{2.17}$$

With Eqs. (2.14)–(2.17) we are now ready to express the PO's expected risk-induced shortage:

$$\begin{aligned} \xi^r(\alpha) = \mathbb{E}\Bigg\{ &\sum_{t=\tau}^{T} \Big[[[O_{t-\tau}^I - S_t^I]^+ - V_t^I]^+ - \langle R_t^{IE}\rangle \Big]^+ \\ &+ \Big[[[O_{t-\tau}^{IE}(\alpha) - S_t^{IE} + O_{t-\tau}^{E}(\alpha) - S_t^{E}]^+ - V_t^{IE}]^+ - R_t^I \Big]^+ \Bigg\}. \end{aligned} \tag{2.18}$$

Eq. (2.18) captures the PO's expected lead-time related shortages during the contractual period $t \in \{0, \ldots, T\}$. If on any day t orders that are due in a demand stream (e.g., the I-demand stream) can not be matched with supplies that arrive, a PO first avoids shortages by re-routing virtual inventory that is in that demand stream. If this virtual inventory is insufficient, the PO re-routes residual virtual inventory from the other demand stream. Any remaining open orders will result in shortages. The sum of expected (daily) shortages in the I- and the IE-demand stream in the procurement cycle constitutes the risk-induced shortage $\xi^r(\alpha)$.

The model in Eqs. (2.14) to (2.18) captures short-term mismatches between countries' demands and manufacturers' supplies, but a PO is also concerned with longer-term supply outages. To account for such longer-term interruptions, we extend Eqs. (2.14) to (2.18) of our model. Suppose that, on any given day t, a supply outage of manufacturer j may occur with probability ρ_j, resulting in the indicator variable $\theta_{j,t} = 0$. The manufacturer remains active with probability $1 - \rho_j$, so $\theta_{j,t} = 1$. We assume that a manufacturer that defaults at time t_d will not recover during the remainder of the contractual period (i.e., $\theta_{j,t} = 0$ for $t \geq t_d$), so the probability that a manufacturer defaults during the contract period is given by $\Omega_j = \sum_{t=1}^{T}(1-\rho_j)^{t-1}\rho_j$. We substitute the supplies S_t^I, S_t^{IE}, and S_t^E in Eqs. (2.14) to (2.18) with supplies $\hat{S}_t^I$, $\hat{S}_t^{IE}$, and $\hat{S}_t^E$ to account for manufacturer defaults as follows: If a manufacturer defaults ($\theta_{j,t} = 0$), supplies of this manufacturer will be disrupted for the remainder of the contractual period:

$$\hat{S}_t^I := \theta_{I,t} S_t^I \quad \text{for } t \in \{0, \ldots, T\}, \tag{2.19}$$

$$\hat{S}_t^{IE} := \theta_{I,t} S_t^{IE} \quad \text{for } t \in \{0, \ldots, T\}, \tag{2.20}$$

$$\hat{S}_t^E := \theta_{E,t} S_t^E \quad \text{for } t \in \{0, \ldots, T\}. \tag{2.21}$$

If we can obtain an estimate of the distributions of the I-demand and the IE-demand streams and an estimate of the distributions of the incumbent's and entrant's lead times L_j and default probabilities Ω_j, we can use Eq. (2.18) in conjunction with Eqs. (2.14) to (2.17) and (2.19) to (2.21) to compute the expected risk-induced shortage $\xi^r(\alpha)$, dependent on the volume split α. Section 2.5.1 describes how we obtained these estimates for the DMPA case.

2.5 Analyses and Discussion

In Section 2.4 we formalized the negotiation and coordination phases and showed how these two models can be used to determine the individual elements $\xi^p(\alpha)$ and $\xi^r(\alpha)$ of the PO's objective $\xi(\alpha) = \xi^p(\alpha) + \xi^r(\alpha)$. To compute the entrant's value (see Eq. (2.2)) we must characterize how $\xi(\alpha)$ depends on α and determine the expected overall shortage $\xi(\alpha^*)$ at the optimal volume split α^*. However, this task is challenging: While Eqs. (2.10), (2.11), and (2.12) of our negotiation model allow for straightforward computation of the equilibrium prices and, based thereon, the associated price-induced shortage depending on α, we cannot obtain a closed-form expression for $\xi^r(\alpha)$. Eq. (2.18) of our coordination model provides an expression for the risk-induced shortage $\xi^r(\alpha)$, but it is not analytically tractable because of the stochastic nature of the problem and the complex inter-temporal relationships of the parameters. Moreover, the calculation of $\xi^r(\alpha)$ in the coordination model depends on the PO's procurement volume, which is an outcome of the negotiation model. Because of these issues, we carried out a numerical study to evaluate the entrant's value and answer the research questions stated in Section 2.2.2.

2.5.1 Model Implementation and Input Data

Figure 2.2 shows how we implemented the negotiation model and the coordination model for our numerical study. Here we explain the individual calculations according to the sequence in which they are carried out.

1. *Computation of price-induced shortages $\xi^p(\alpha)$:* As stated previously, we can use Eqs. (2.10) and (2.11) to compute the incumbent's and the entrant's equilibrium prices, $p_I^*(\alpha)$ and $p_E^*(\alpha)$, respectively, for a given split α. Then we can use Eq. (2.12) to compute the corresponding price-induced shortage $\xi^p(\alpha)$. We compute $\xi^p(\alpha)$ for all feasible values of α, given the relevant input parameters summarized in Figure 2.2.

2. *Computation of the order-scaling factor $\varphi(\alpha)$:* As explained in Section 2.4.2, the PO's overall procurement volume depends on the prices the PO negotiates with the manufacturers—which, in turn, depend on the volume split α—and can be computed as $d(\alpha) = q - \xi^p(\alpha)$. To adequately reflect the PO's overall demand, we scale the orders of the recipient countries in our coordination model using an order-scaling factor, which is calculated as $\varphi(\alpha) = 1 - \frac{\xi^p(\alpha)}{q}$. Thus, for each value of α, we use the price-induced shortage (calculated in step 1) to obtain $\varphi(\alpha)$. As Figure 2.2 shows, the order-scaling factor is an input to our coordination model.

3. *Estimation of the risk-induced shortage $\xi^r(\alpha)$:* To reflect the stochastic nature of the order lead times and manufacturers' defaults, we use a simulation approach to obtain estimates of $\xi^r(\alpha)$ depending on α. We create countries' demands within a procurement cycle (we refer to this as the order book) which the PO translates into orders to the manufacturers. For each order we draw the corresponding lead-time realization from

a lead-time distribution. We also draw the realization of a manufacturer's default at a specific time within the procurement cycle from a probability distribution. Based on Eq. (2.18) we account for orders that are satisfied within the shortage threshold, for orders that avoid shortage through re-routing, and we collect the (realized) supply shortage in a procurement cycle $\xi_n^r(\alpha)$ for a simulation run n. We repeat this N-times and estimate the expected risk-induced shortage as $\xi^r(\alpha) = \frac{1}{N}\sum_n \xi_n^r$. This procedure is carried out for all feasible values of α (see Figure 2.2). We implemented this simulation model using Microsoft Excel with the add-in @Risk.

4. *Calculation of* $\xi(\alpha)$ *and the entrant value* w*:* Based on the results obtained in Steps 1 and 3, we calculate $\xi(\alpha) = \xi^p(\alpha) + \xi^r(\alpha)$ for all feasible values of α and determine the optimal split α^* and the entrant value $w = \frac{\xi_{\text{sole sourcing}} - \xi(\alpha^*)}{\xi_{\text{sole sourcing}}}$. Note that we calculate the shortage of sole sourcing as $\xi_{\text{sole sourcing}} = \xi^p_{\text{sole sourcing}} + \xi^r_{\text{sole sourcing}}$, where $\xi^p_{\text{sole sourcing}} = q - \frac{b}{r}$, which is the outcome of the negotiation if the entrant is not available to the PO, and $\xi^r_{\text{sole sourcing}} = \xi^r(\alpha = 0)$.

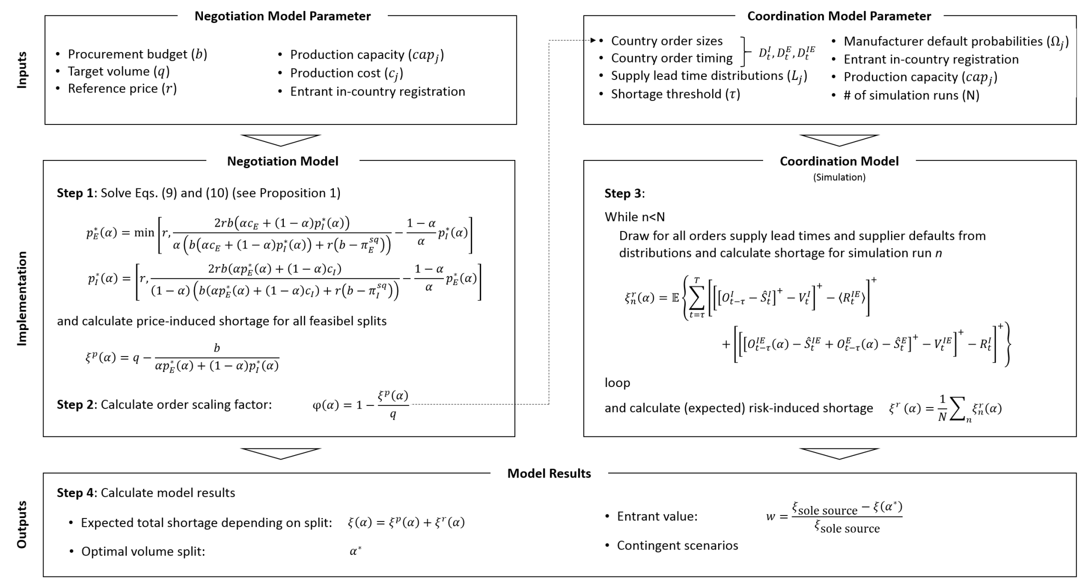


Figure 2.2: Model inputs, model implementation, and model outputs.

Next, we explain how we obtained estimates of the input parameters in order to establish a reference case for the aforementioned calculations (see Table 2.1). We rely for our data estimates on two categories of sources: On the one hand we have access to well documented data from reports and data bases like lead times for country orders, order volumes, or past procurement prices. On the other hand, some parameters of our model are inherently difficult to estimate. For example, it is difficult to obtain "reliable" estimates of the probability that a manufacturer will default, the size of production capacity and the (variable) production costs. We interviewed industry experts to collect estimates for these parameters.[8] We acknowledge that these experts' estimates tend to be subjective and not well grounded, but absent more precise and reliable information we are not able to come up with a more robust estimate. However, we can explore how the particular choice of a parameter impacts the PO's objective of minimizing expected shortage. We used the following estimates of the relevant parameters to compute the price-induced shortage $\xi^p(\alpha)$ (Step 1): We assume the *procurement budget* (b) is sufficient for buying the yearly procurement volumes for DMPA in 2012 and 2013 (approx. $48,000,000$ units per year for USAID[9]) at the current list price of \$0.8 per-unit, such that $b = \$38,400,000$. These estimates are based on the annual reports of both POs (RHSC, 2016). Based on recent forecasts (WDI/CF, 2015; CSP/RHSC, 2015), we assume that the *target volume* (q) that the PO faces in the reference case increases by approximately 20 percent (compared to 2012/13) yielding a target volume of $q = 58,000,000$ units. We set the current list price (RHSC, 2016) as the *reservation price* $r = \$0.8$. For our reference case we assume that both manufacturers have sufficient *production capacity* (cap_j) to produce the PO's target volume ($cap_I = cap_E = 60,000,000$ units). Our estimate of the incumbent's *production costs* is based on information we obtained from discussions with manufacturing experts in the field of injectable pharmaceuticals. We estimate the incumbent's production costs as $c_I = \$0.66$ and assume that the entrant in the reference case has the same costs as the incumbent.

We used the following estimates of the relevant parameters to compute the risk-induced shortage $\xi^r(\alpha)$ (Step 2): We build an *order book* for the coordination model that contains the order sizes and order dates for each country, d_t^I, d_t^E, d_t^{IE}, which the PO translates into orders to the manufacturers, o_t^I, o_t^E, o_t^{IE}. We use data from the RHInterchange database (RHInterchange, 2016), which contains 280 entries of order-placement dates and order volumes on a country-by-country basis for USAID and UNFPA in the 2012/2013 procurement cycle, to create this order book. That is, we use the past orders that, for example, USAID placed at one manufacturer on behalf of the countries and assume that each order is now split between the incumbent and the entrant according to the volume split α. We also scale each individual order by the scaling factor

[8] We interviewed experts in essential medicines procurement, and pharmaceutical supply chain and production management experts from the following organizations (in alphabetical order): the Reproductive Health Supplies Coalition, the United States Agency for International Development, the United Nations Population Fund, and the William-Davidson-Institute.

[9] We present our results for data of USAID, but the intuition and findings apply to UNFPA as well.

$\varphi(\alpha) = 1 - \frac{\xi^p(\alpha)}{q}$ to account for the volume the PO can procure given the outcome of the negotiation phase.[10] Our model offers the possibility to manipulate on a country-by-country basis in which countries the entrant's product is registered (*in-country registration*). For the reference case we assume that the entrant's product is registered in all countries. Table 6.1 in the appendix presents the countries that USAID currently supplies. Using the RHInterchange database, we also estimate the *supply lead time* distribution (L_j) for the incumbent manufacturer via distribution fitting. We found the Gamma-distribution $\Gamma[2.0995, 24.712]$ to be a good fit for the lead-time distribution[11] and use this distribution to simulate the manufacturers' lead times, that is, the arrivals of supplies S_t^I, S_t^E, S_t^{IE}. To explore the impact of varying levels of lead-time uncertainty, we implement two parameters that vary the mean and the standard deviation of our lead-time distribution. We assume that the entrant has the same type of lead-time distribution as the incumbent and implement a parameter for mean and standard deviation to model other lead-time distributions for the entrant. For the reference case we use the same lead-time distribution estimators for the incumbent and the entrant.[12] We also use the lead-time distribution to estimate the *shortage threshold* (τ) as the mean lead time ($\mu_j \simeq 52$) plus one standard deviation ($\sigma_j \simeq 36$). We asked pharmaceutical experts to assess the probability of longer-term defaults in factories of similar size, technology, and complexity. Our industry experts estimated this probability as a lower single-digit percentage. Based on this information, we use an *annual default probability* (Ω_j) of 0.03 (i.e. 3 percent) for our analyses which gives $\rho_j = 0.000081$. Table 2.1 summarizes the parameter values for the negotiation and the coordination model in the reference case.

We sample $N = 1,000$ instances[13] from the lead-time and the default distribution to calculate the risk-induced shortage for a given split $\alpha \in [0, 1]$. To quantify the risk-induced shortage depending on the split, we vary the split in increments of 0.01.

The two outputs, price-induced and risk-induced shortage, for every (feasible) value of $\alpha \in [0, 1]$, allow us to determine which split minimizes the total shortage and how the two components individually drive the optimal split. Based on these outputs, we can compute the entrant's value w using Eq. (2.2) and show how parameter variations affect

[10] We could also manipulate the order book country-by-country if a decision-maker were interested in the effects of specific changes to individual country volumes or the impact of a different order policy.

[11] Using the Cramér-von Mises goodness-of-fit test for the distribution yields $K = 0.0471998$, which is smaller than the critical value of 0.33 for our sample at a 0.01 significance level.

[12] We tested to determine whether order size and lead times or the lead times over time are correlated and found no evidence that suggests such correlations.

[13] We tested different sample sizes and found that a sample size of $N = 1,000$ leads to a sufficient accuracy. We performed 30 runs with sample size $N = 1,000$ and found that risk-induced shortage varies by less than 1% across different simulation runs.

	Parameter	PO	Incumbent	Entrant
Annual procurement budget [$]	b	38,400,000	-	-
Annual target volume [units]	q	58,000,000	-	-
Annual production capacity [units]	cap_j	-	60,000,000	60,000,000
List price [$]	r	.8	-	-
Unit production costs [$]	c_j	-	.66	.66
Order sizes & times	o^I, o^E, o^{IE}	-	order book	order book
Supply lead-time distribution	L_j	-	Γ[2.0995;24.712]	Γ[2.0995;24.712]
Shortage threshold [days]	τ	-	88	88
Entrant in-country registration	-	-	full reg.	full reg.
Annual default probability	Ω_j (ρ_j)	-	0.03 (0.000081)	0.03 (0.000081)

Table 2.1: Parameter values in the reference case

the optimal split and the entrant's value.

2.5.2 Analysis of the Entrant's Value

We are now ready to address the key question of our study: how much value does an entrant provide to a PO? The following analyses are structured along the research questions introduced in Section 2.2.2, which were motivated by the needs of practitioners in our DMPA case. As highlighted in the introduction, a PO wants to contract new generics manufacturers to access lower prices through increased competitive pressure. This is oftentimes fueled by the assumption that generics manufacturers are able to produce the same product at lower costs. However, POs are also concerned about the potential for increased supply risks and are often not sure whether the benefits of competition outweigh the negative effects of increased supply risks. This trade-off motivated our first research question (Q.1): *What is the value of a lower-cost, higher-risk entrant and how does this value depend on the volume split?* (see Section 2.2.1).

To answer this question Section 2.5.2 compares the entrant's value in the reference case—in which all parameters are the same for both manufacturers—to the entrant's value in three different scenarios: In Scenario I the entrant has moderately lower unit costs ($c_I = .66 > .62 = c_E$), which is a realistic assumption when new generics manufacturers enter the market. In Scenario II the entrant has not only lower unit costs but also higher supply risk. In this scenario we assume the same entrant costs as in Scenario I but a (moderate) 15 percent increase in expected lead time ($\mu_E = 1.15\mu_I$) and lead-time variability ($\sigma_E = 1.15\sigma_I$). In Scenario III we consider an entrant that offers lower unit costs, but has substantially higher supply risk (i.e., $\mu_E = 1.3\mu_I$ and $\sigma_E = 1.3\sigma_I$). Table 2.2 summarizes the most important parameters for the reference case and the three scenarios and the main results of our analyses in Sections 2.5.2, 2.5.2, and 2.5.2.

Section 2.5.2 addresses our second research question (Q.2): *What is the value of an entrant with limited capacity and how does this value depend on the volume split?* To answer this question, we reduce the entrant's production capacity ($cap_E = 25,000,000 < cap_I = 60,000,000$) and investigate how the entrant's value changes across the three scenarios that we outlined previously. We calculate the entrant's value when all parameters except the capacity are equal and compare it with the entrant's value in Scenario I (lower entrant costs), Scenario II (lower entrant costs and moderately higher entrant risk), and Scenario III (lower entrant costs and substantially higher entrant risk) (see Table 2.2). In addition we investigate a moderately capacitated entrant ($cap_E = 42,500,000$) and compare the results to our previous analyses.

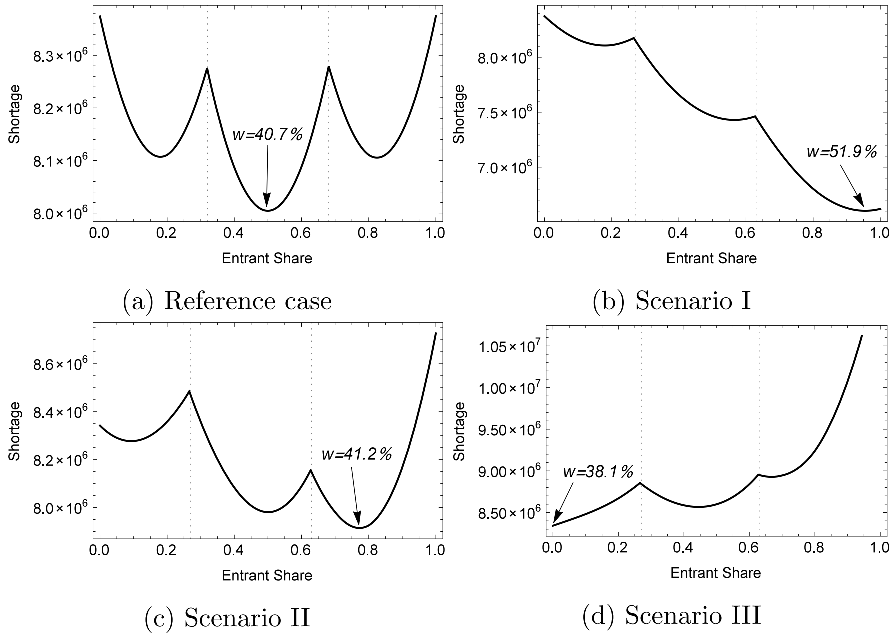
			Reference Case	Scenario I	Scenario II	Scenario III
			$c_I = c_E = .66$, $\mu_I = \mu_E = 52$, $\sigma_I = \sigma_E = 36$	$c_E = .62$	$c_E = .62$, $\mu_E = 1.15\mu_I$, $\sigma_E = 1.15\sigma_I$	$c_E = .62$, $\mu_E = 1.3\mu_I$, $\sigma_E = 1.3\sigma_I$
Q.1	lower costs & higher risk	α^*	.50	.95	.77	0
		w	40.7 %	51.9 %	41.2 %	38.1 %
Q.2	$cap_E = 25,000,000$	α^*	.48	.48	.48	.48
		w	33.3 %	38.5 %	34.8 %	31.1 %
	$cap_E = 42,500,000$	α^*	.76	.76	.76	.66
		w	39.9 %	48.5 %	41.3 %	33.9 %
Q.3	limited registration	α^*	0	.48	0	0
		w	21.7 %	22.3 %	21.7 %	21.7 %

Table 2.2: Comparison of optimal volume split (α^*) and corresponding entrant values (w) in case the entrant has sufficient capacity (Q.1), in case the entrant has limited capacity (Q.2), and in case the the entrant's product is not registered in all countries (Q.3) for the reference case and Scenarios I, II, and III.

Our introduction (Section 2.1) explained that a new manufacturer must undertake a significant effort to register its product in different countries. It is unlikely that the entrant's product will be registered in all of the PO's recipient countries and these missing registrations will likely have an impact on the entrant's value to the PO. Section 2.5.2 addresses this question (Q.3): *What is the value of an entrant whose product is not registered in all recipient countries and how does this value depend on the volume split?* Our analysis follows the same path as before: We calculate the entrant's value in the reference case but with limited in-country registrations and compare the result to the entrant's value in Scenarios I, II, and III.

The Value of a Risky Low Cost Entrant

To determine the value of a risky, low-cost entrant to a PO (Q.1), we first calculate the total expected shortage ($\xi(\alpha) = \xi^p(\alpha) + \xi^r(\alpha)$) and the entrant's value for the reference case, in which all parameters are equal for both manufactures. Figure 2.3a shows that, in this case, the total shortage exhibits a symmetric triple-v-shape in α with its (global) minimum at equal shares for both manufacturers ($\alpha^* = 0.5$).

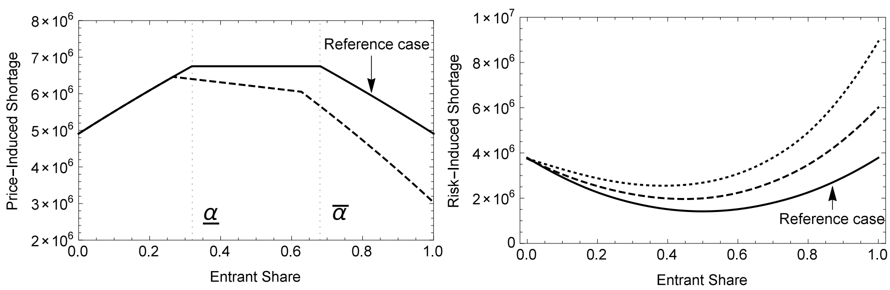


(a) Reference case (b) Scenario I

(c) Scenario II (d) Scenario III

Figure 2.3: Total expected shortages ($\xi^p(\alpha) + \xi^r(\alpha)$) depending on the entrant's share featuring an entrant with (a) lower unit costs $c_E = 0.62 < c_I = 0.66$; (b) lower unit costs and a 15 percent increase in expected lead time and lead-time standard deviation $\mu_E = 1.15\mu_I, \sigma_E = 1.15\sigma_I$; (c) lower unit costs and a 30 percent increase in expected lead time and in lead-time standard deviation $\mu_E = 1.3\mu_I, \sigma_E = 1.3\sigma_I$.

The triple-v-shaped curve is the result of the aggregation of price-induced shortage (see Figure 2.4a) and risk-induced shortage (see Figure 2.4b). In the reference case the price-induced shortage is symmetric and concave in the entrant's share and reaches its minimum if the PO chooses to single-source from either the incumbent or the entrant. In line with literature on split-awards (for example, Anton and Yao (1989) and Gong et al. (2012)) any deviation from single-sourcing yields higher average prices and, consequently, some level of *dual-sourcing inefficiency.* Note that price-induced shortage is partitioned into three regions separating different manufacturer's incentives to compete. For low allocations to the entrant ($0 \leq \alpha \leq \underline{\alpha}$), the entrant provides the reservation price and the PO benefits from the incumbent providing lower prices (and vice versa for $\overline{\alpha} \leq \alpha \leq 1$). For $\underline{\alpha} \leq \alpha \leq \overline{\alpha}$ the price-induced shortage is almost constant because both suppliers provide competitive bids and increasing or decreasing the split does not benefit the PO (see also Figure 6.1 in the appendix). Figure 2.4a also shows that if a manufacturer (e.g. the entrant) has lower costs it is able to introduce stronger competition for higher volume splits such that price-induced shortage is lower and a PO prefers to single-source from the less expensive supplier. Risk-induced shortage is convex in the entrant's share because the PO benefits from a *diversification effect.* In the reference case, in which both suppliers expose the PO to the same supply risks, a PO should allocate an equal split between incumbent and entrant. If, for example, the entrant has a higher expected lead time (μ_E) and lead-time variability (σ_E) risk-induced shortage increases and makes it more attractive to award a higher share to the incumbent.[14] However, even an entrant with very high risk provides the PO with diversification benefits (see dotted line).[15]



(a) Price-induced shortages ($\xi^p(\alpha)$). (b) Risk-induced shortages ($\xi^r(\alpha)$).

Figure 2.4: (a) Price-induced shortages ($\xi^p(\alpha)$) depending on the entrant's share for $c_E = c_I = .66$ (solid) and $c_E = .62 < c_I = .66$ (dashed), and (b) Risk-induced shortages ($\xi^r(\alpha)$) depending on the entrant's share for $\mu_E = \mu_I$ and $\sigma_E = \sigma_I$ (solid), $\mu_E = 1.15\mu_I$ and $\sigma_E = 1.15\sigma_I$ (dashed) and $\mu_E = 1.3\mu_I$ and $\sigma_E = 1.3\sigma_I$ (dotted).

Depending on which individual effect is more powerful, the total shortage can be

[14]We opt to increase both the mean as well as the standard deviation of the lead time because in our conversations with practitioners we learned that this is a likely scenario. Of course, we could only modify one of the parameters. This does not change the directionality of the insights we present here.

[15]In our analyses we only present the mean risk-induced shortage of N simulation runs. In Figure 6.2 in the appendix we report the standard deviation of risk-induced shortage in the reference case. The plot shows that the standard deviation drops if the PO dual-sources. This finding is consistent throughout our analyses. To keep our exposition lean and comprehensive, we refrain from reporting the standard deviations for our subsequent analyses.

at a minimum if the PO single-sources (competition effects dominate) or awards an equal split (risk effects dominate). In the reference case benefits from risk diversification outperform the benefits from competition. Thus, an equal split leads to the lowest overall expected shortage and the highest entrant's value of 40.7 percent in the reference case. Notably, in the reference case there is always a clear trade-off between price- and risk-induced shortage for any split; when price-induced shortage increases (decreases) the risk-induced shortage decreases (increases) (see Figure 2.4 for the individual effects). We will highlight situations in which this common managerial intuition does not hold and where there is no longer a strict trade-off.

It is interesting to observe that we can directly link the shape of the total expected shortage curve to results obtained in Figure 2.4a. The incentive to quote lower prices in the negotiation phase, which resulted in the three partitions, causes the triple-v-shaped structure of total expected shortage. Decision-makers should be aware that local maxima occur for splits $\underline{\alpha}$ and $\overline{\alpha}$. As a result, a PO's objective function is not robust against changes in the allocation, and expected shortage can increase significantly if the PO chooses splits close to $\underline{\alpha}$ and $\overline{\alpha}$. This has an important managerial implication: the value of the entrant depends heavily on how the PO splits the volume between the incumbent and the entrant. The PO will experience the maximum entrant value (40.7 percent in our reference case) only if it sets α correctly.

Next, we compare the results of the reference case to Scenario I, where the entrant has smoderately lower production costs ($c_E = .62 < c_I = .66$) but still features the same risks as those of the incumbent.[16] Such a cost advantage will introduce stronger competition, and a PO should allocate the entire volume to the entrant to minimize price-induced shortage (see Figure 2.4a). Figure 2.3b shows the total expected shortage for Scenario I and the entrant's value in Scenario I. The entrant provides the highest overall value if the PO allocates almost the entire volume to the entrant ($\alpha^* = 0.95$); that is, the price decrease that results from increased competition offsets almost any increase in risk-induced shortages that go along with the entrant receiving a higher share. As a result, even a relatively small cost advantage of approximately 6 percent leads to substantial increases of the entrant's value (from $w = 40.7\%$ in the reference case to $w = 51.9\%$; see Table 2.2). However, to reap these benefits of a lower-cost entrant a PO must choose the correct volume split, which is, in this case, sourcing almost the entire volume from the entrant.

[16] Practitioners indicated that the generics supplier will have a production cost advantage over the incumbent. From our analysis the reader can easily deduce the directional changes to the results if the incumbent has the production cost advantage.

Scenario II takes into account decision-maker's concerns that an entrant that provides a cost advantage can also be more risky. Figure 2.3c shows that, with moderately higher entrant risk ($\sigma_E = 1.15\sigma_I$ and $\mu_E = 1.15\mu_I$), the PO should choose to allocate less to the entrant (i.e., $\alpha^* = 0.77$). The higher risk has the entrant's overall value decline to $w = 41.2\%$ (see Table 2.2), but the benefits of risk diversification still compensate for disadvantages of dual-sourcing inefficiency that a PO incurs from choosing to dual-source. If the entrant exposes the PO to even higher risks, the entrant's value declines farther, and it becomes optimal for the PO to source the entire volume from the incumbent (see Figure 2.3d). It is interesting to note, that, in this case, the entrant still provides a substantial positive value to a PO ($w = 38.1\%$), even though it is optimal for the PO to single-source from the incumbent. The mere existence of the entrant introduces competitive pressure and lowers the price-induced shortage compared to that incurred from sole-sourcing. This result suggests that an entrant always provides value to a PO, even if the entrant-related risk is substantially higher. Because of this risk the PO will not allocate volume to the entrant, but the PO benefits from the added competitive pressure in the negotiation with the incumbent and will experience lower price-induced shortages. However, it is clear that this benefit may not survive long-term because it is unlikely that an entrant will maintain the production capacity over a longer period to compete with the incumbent if the PO does not allocate any volume to the entrant. Therefore, in its decision to allocate volume, the PO may want to sacrifice some current benefits (by allocating some volume to the entrant) in order to ensure longer-term competition and to reduce shortages in future procurement cycles. Although out of the scope of our analysis, we observe that our approach allows the PO to evaluate the immediate consequences of deviating from the optimal split α^* in order to incent the entrant to compete in future procurement cycles.

The results established in this section answer our first research question (Q.1) by revealing that a lower-cost entrant provides substantial value to a PO. This holds true for relatively low cost advantages (in our case we only considered a cost advantage of approximately 6 percent) and even if the entrant exposes the PO to substantially higher supply risks. In the latter case the PO should not allocate volume to the entrant, but the entrant still provides value because it induces the competitive pressure that leads to a reduction in price-induced shortage. While this effect may not be long-term, it highlights that, even if the PO does not contract with the entrant, incenting a new entrant may still be attractive. Another important insight can be derived from our results: the entrant's value heavily depends on the PO's choice of the volume split, as the overall shortage and the entrant's value are sensitive to the volume split α and the PO must determine and implement the optimal volume split (α^*) in order to gain the most benefit from the entrant. Our discussions with practitioners in the field revealed that they often use pre-defined volume splits (e.g., 30 percent for the entrant, 70 percent for the incumbent), but our results suggest that pre-defined volume splits can have unfavorable outcomes for the PO. For example, if the PO chooses the 30%/70% instead of the optimal split in the reference case ($\alpha = .5$), the entrant's value drops from $w = 40.7\%$ to $w = 38.6\%$.

The Value of a Capacity Constrained Entrant

In the previous section, we assumed that the entrant has sufficient capacity to fulfill the PO's entire demand. While this assumption may hold true in other settings, such as when POs procure a comparatively small volume relative to the overall market volume, it is unlikely to hold in the domain we address. UNFPA and USAID procure a considerable share of the global market volume, and we cannot assume that an entrant will immediately install adequate production capacity to fulfill this demand.

Therefore, we seek to answer the question concerning how much value an entrant with limited capacity provides to a PO (Q.2). We structure our analysis using the same scenarios with the same parameters as in the previous section. However, we reduce the production capacity of the entrant to $cap_E = 25,000,000$ ($< cap_I = 60,000,000$), which is only 43 percent of the PO's target volume q. We obtained this estimate for one of the potential generics manufacturers that intended to enter the DMPA market in 2015.

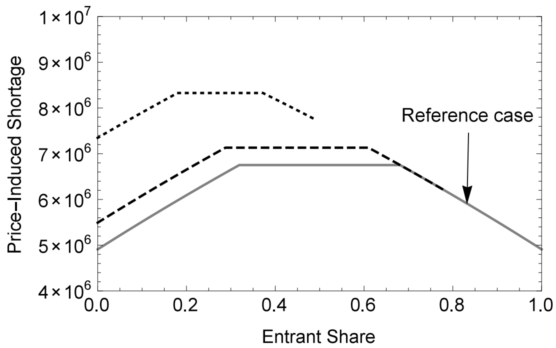
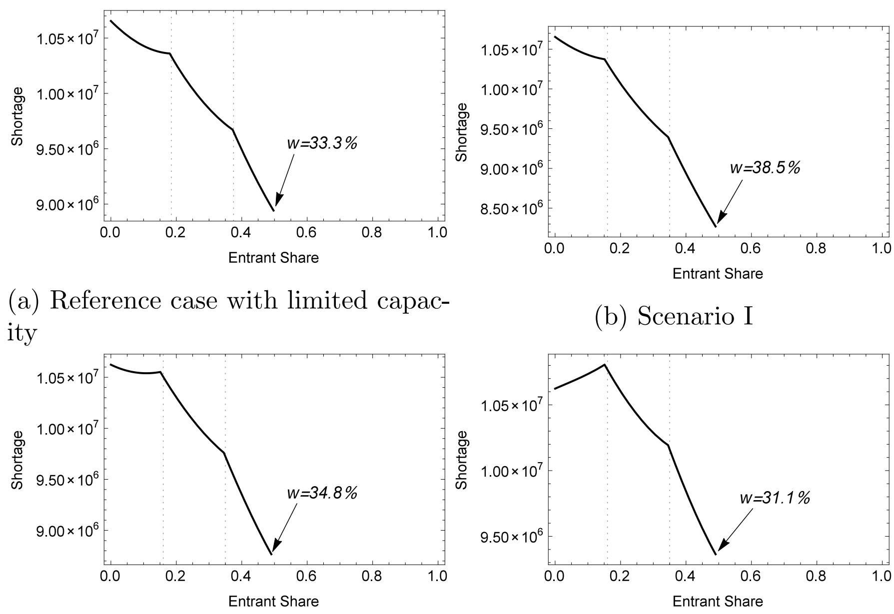


Figure 2.5: Price-induced shortages $\xi^p(\alpha)$ depending on the entrant's share for varying entrant's capacity ($cap_E = 60,000,000$ (solid), $cap_E = 42,500,000$ (dashed), and $cap_E = 25,000,000$ (dotted)).

Figure 2.5 shows that limited capacity has a negative impact on competition because it lowers the status-quo utility of the capacitated manufacturer. As a result a PO sees higher price-induced shortages when large shares are allocated to the entrant. A PO should therefore single-source from the incumbent to minimize price-induced shortage.

We first calculate the total expected shortage using the reference case's parameter values but with limited entrant capacity. Figure 2.6a shows that even if the entrant has limited capacity, a PO should allocate as much procurement volume to the entrant as possible. However, the reduced competitive pressure decreases the relative importance of price-induced shortage and increases the relative importance of risk diversification, the latter of which has a particularly strong effect when awarding a relatively high share to the entrant (see Figure 2.6a). Hence, allocating as much as possible to the entrant is optimal in this setting. Compared to the reference case in Section 2.5.2 the entrant's value drops from $w = 40.7\%$ to $w = 33.3\%$ (see Table 2.2) because of the increase in price-induced shortage, but the (capacity-constrained) entrant still provides substantial value to the PO.



(a) Reference case with limited capacity

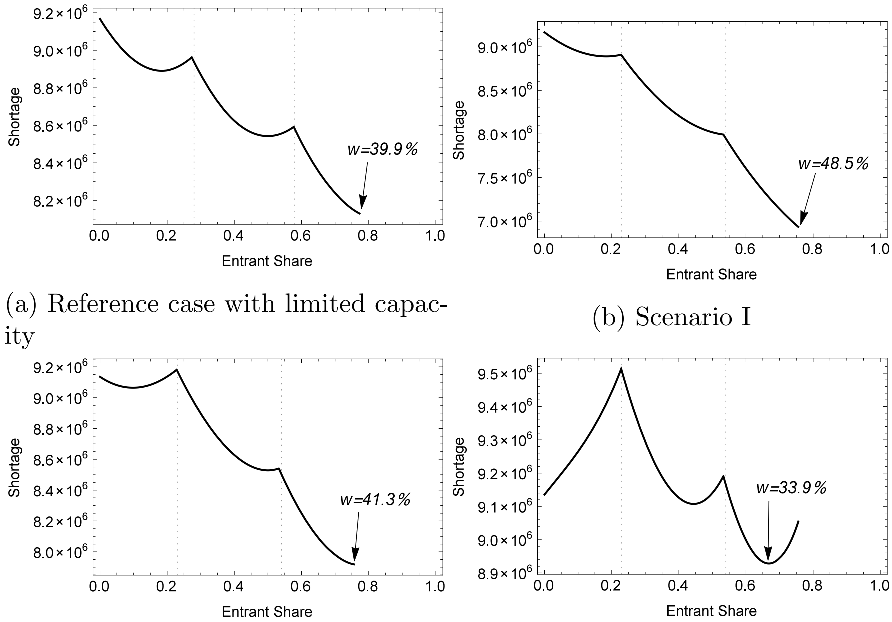
(b) Scenario I

(c) Scenario II

(d) Scenario III

Figure 2.6: Total expected shortages ($\xi^p(\alpha) + \xi^r(\alpha)$) depending on the entrant's share for $cap_E = 25,000,000$ for Scenarios I, II, and III.

The entrant value increases further when the entrant has a cost advantage over the incumbent (Scenario I). Even if the entrant's risk increases (Scenarios II and III), the buyer minimizes total shortages when it allocates the largest possible share to the entrant (see Figure 2.6d). However, we again observe that the total expected shortage at the thresholds $\underline{\alpha}$ and $\overline{\alpha}$ is highly sensitive to the choice of α. Regarding research question 2



(a) Reference case with limited capacity

(b) Scenario I

(c) Scenario II

(d) Scenario III

Figure 2.7: Total expected shortages ($\xi^p(\alpha) + \xi^r(\alpha)$) depending on the entrant's share for $cap_E = 42,500,000$ for Scenarios I, II, and III.

(Q.2), our results point to substantial differences between the scenarios for an entrant with sufficient capacity and a strongly capacity-constrained entrant (compare Figure 2.3 and 2.6). Our results confirm that the entrant's value decreases if its capacity is limited (see Table 2.2). However, the intricate interplay of risk and costs can result in drastic changes to the structure of total expected shortage and a PO's optimal allocation decision. Decision-makers are intuitively hesitant to allocate volumes to a risky entrant with lower capacity, but our analysis reveals that POs should consider—employing a simple allocation rule—to allocate high shares to a highly capacity constrained entrant because risk-diversification and competition benefits can both work in the PO's favor. It is rather intuitive that this also holds in situations in which the entrant's capacity is even more constrained. However, when the entrant has higher capacity, the results gradually converge to those presented in Section 2.5.2, where the entrant has sufficient capacity. To illustrate this effect, Figure 2.7 shows the total expected shortage for an entrant with less constrained capacity (i.e., $cap_E = 42,500,000$ which is 73 percent of the target volume q). We see that in Scenario III, where the entrant exposes the PO to substantial supply risks, it is optimal for the PO to allocate less volume to the entrant. In this setting, there is again a strict trade-off between price-induced and risk-induced shortages: while the former decrease (see Figure 2.5), the latter increase (at a higher rate) in the entrant's share (see Figure 2.4b).

The Value of an Entrant with Limited In-country Registration

Section 2.5.2 explained that both competition and supply risks can be decisive for the PO's optimal split, so they are both important drivers of the entrant's value. Our results demonstrated that, depending on the incumbent's and entrant's characteristics, risk-induced (price-induced) shortages may have a stronger effect on the total expected shortage than price-induced (risk-induced) shortages (see Figure 2.3a). Section 2.5.2 showed how the effect of an entrant's limited capacity on price-induced shortage can substantially change the entrant's value and the optimal volume split. Our discussions with practitioners revealed that in-country registration can also be a limiting factor, as in-country registration impacts the PO's ability to match manufacturers' supply with countries' demands. To explain the value of an entrant with limited in-country registration, we compare the total shortage in the reference case (where the products of both manufacturers are registered in all countries) to a case of limited in-country registration that is based on a set of countries in which the product of the generics manufacturer who was seeking to enter the market in 2015 were registered (Table 6.1 in the appendix shows these countries). In this case, the entrant registered its product in ten (of twenty-six) countries that account for approximately 45 percent of the target volume (dashed line in Figure 2.8). Figure 2.8 shows that compared to full registration limited in-country registration increases risk-induced shortage for a given volume split because the PO has less opportunities to compensate for late shortages. Limited in-country registration also reduces the feasible allocations a PO can allocate to the entrant. Because limited in-country registration limits the potential allocations to the entrant it also affects competition and price-induced shortages.

As before, we calculate total shortage in the reference case, but with limited registration, and compare the results to those obtained for Scenario I, II, and III. Figure 2.9a displays the total shortage depending on the entrant's share for the reference case but with limited in-country registration. The results indicate that a PO should source the entire volume from the incumbent, as the entrant's value amounts to $w = 21.7\%$, which is substantially lower than in the reference case with unconstrained capacity (where $w = 40.7\%$) and the reference case with constrained capacity (where $w = 33.3\%$)—see Table 2.2. The sharp decrease in the entrant's value is the result of two effects. First, limited registration has essentially the same impact on competition as constrained

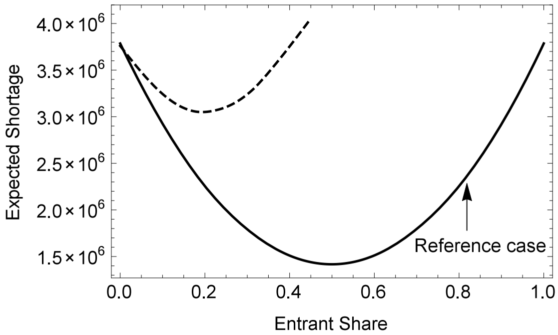
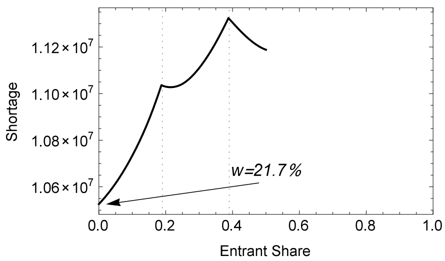
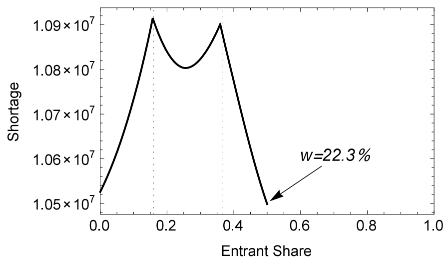


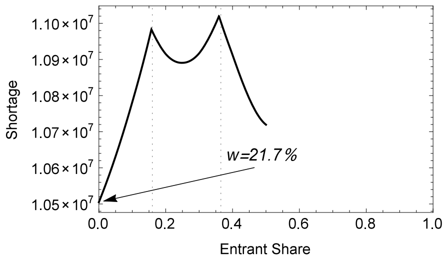
Figure 2.8: Risk-induced shortage depending on the entrant's share for full entrant registration (solid) and limited entrant registration (dashed).



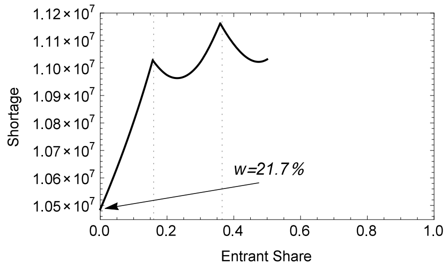
(a) Reference case with limited reg.



(b) Scenario I with limited reg.



(c) Scenario II with limited reg.



(d) Scenario III with limited reg.

Figure 2.9: Expected shortages ($\xi^p(\alpha) + \xi^r(\alpha)$) depending on the entrant's volume split for the case of limited in-country registration.

entrant capacity: It restricts the volume the PO can allocate to the entrant, so the incumbent's bargaining position improves, which increases the price-induced shortage and creates a preference for sourcing the entire volume from the incumbent (see Figure 2.5 for the effect of limited capacity on price-induced shortage). Second, limited registration increases risk-induced shortage for larger allocations to the entrant (Figure 2.8). The resulting total shortage curve (in Figure 2.9a) indicates that the benefits of competition from larger shares allocated to the entrant cannot outweigh the increase in the risk-induced shortages, so the PO should single-source from the incumbent. In Scenario I the entrant's lower costs provides sufficient competitive benefits at high allocations to the entrant to compensate for the higher risk-induced shortage. In this case, the PO should source as much volume as possible from the entrant. However, we recognize that the difference between expected shortage for $\alpha = 0$ and $\alpha = .48$ is comparatively small, and the entrant value is only marginally higher at $w = 22.3\%$. However, it is optimal for the PO to single-source from the incumbent when the entrant also introduces higher risks (Scenarios II and III). The entrant's value then falls back to $w = 21.7\%$. Based on these results, the answer to our third research question (Q.3) is that an entrant with limited in-country registration still provides value to the PO, but because limited in-country registration has a negative effect on both price-induced shortage and risk-induced shortage, it has negative consequences and leads to a significant decline in the entrant's value.

The entrant's in-country registration is an important factor for the PO because it has a significant effect on the entrant's value *and* the entrant itself as it affects the entrant's competitive position. Therefore, both benefit from higher levels of in-country registration. However, registering a product in a country is costly and time-consuming, so knowing how much additional value registration provides on a country-by-country basis would be useful. It would allow the entrant to prioritize its registration efforts and to determine whether it is worthwhile from an economic perspective to pursue registration in a particular country. Knowing the value of registration in individual countries can also help the PO in incenting the entrant by, for example, providing financial, political, or management support for registration in a particular country. Our approach allows us to provide this information to both the entrant and the PO. Table 2.3 presents the marginal increase in the entrant's value ($\Delta w = (w^{lim+c} - w^{lim})$) that is associated with registration in an additional country (c). Figure 2.10 illustrates the decrease in the price-induced and risk-induced shortages for registration in individual countries.

Country (c)	Mali	Afghanistan	Haiti	Bangladesh	Mozambique	Pakistan	Ethiopia
Δw	0.6%	1.5%	1.8%	2.0%	2.4%	6.9%	7.3%

Table 2.3: Increase in the entrant's value for selected, additionally registered countries compared to the limited registrations case.

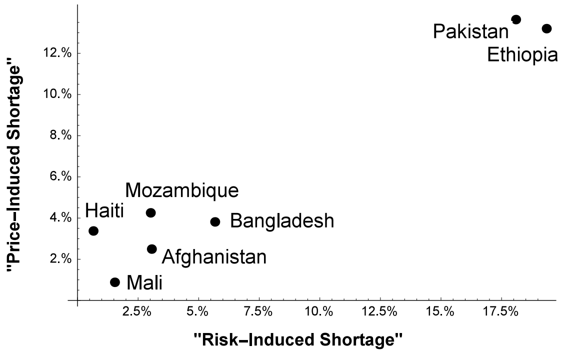


Figure 2.10: Reduction in risk-induced and price-induced shortages of selected countries in which the entrant has not registered its product in the limited registration case.

2.6 Conclusion

Purchasing organizations (POs) in donor-funded global-health markets seek to provide people in low-income countries with essential medicines, but they struggle with high demand and insufficient budgets. Generics manufacturer's entry into these markets promises lower purchasing prices through increased competition, thereby allowing POs to supply more people with pharmaceutical products and increasing the number of treatments delivered. However, these entrants may also increase the supply risk because of, for example, longer lead times or higher default probability, resulting in shortages that reduce the number of treatments delivered. Therefore, POs are confronted with the difficult question concerning when a new generics manufacturer's value is sufficient to justify receiving a volume allocation, and how much volume should be allocated to it.

To answer these questions, we developed two interrelated models that capture the specific dynamics of the negotiation phase (in which a PO negotiates with manufacturers) and the coordination phase (in which a PO seeks to avoid supply-demand mismatches), and we carried out analyses to clarify the underlying trade-offs between competition and risk and to determine the entrant's value under scenarios that are particularly relevant to practitioners. Our analysis was inspired by a project we conducted for donor-funded procurement of DMPA through two large POs with the purpose of showing how the POs should react if a new generics manufacturer becomes available. To capture the entrant's value, we defined a welfare-oriented global-health PO's objective as total expected shortage, a unit-based measure that is the sum of price-induced and risk-induced shortage. Price-induced shortage captures the fact that a donor-funded PO does not focus as much on minimizing procurement costs as it does on treating as many patients as possible with a given budget. Risk-induced shortage measures the consequences of supply disruptions without explicitly assuming shortage costs. This formulation of the PO's objective distinguishes our approach from standard commercial approaches that focus on profit maximization or cost minimization.

In jointly considering price- and risk-effects, we show that a PO's objective function can have multiple maxima and minima (for varying splits between incumbent and entrant). We identify that a root cause of the multi-modal objective function are changes in the incentives during the negotiation and show that manufacturers' costs, capacities, and in-country registration will impact the position of the local minima and maxima. These results provide new insights into the intricacies involved in determining the optimal volume allocation and maximizing the entrant's value. They show that the PO's outcome is highly sensitive to the volume split, so they suggest that a PO must be careful when choosing its allocations.

We also contribute to the literature on bargaining models by being the first to incorporate the splitting decision into a bilateral bargaining model and to model the impact of capacity constraints explicitly.

We used various practically relevant settings to investigate how a PO should split its procurement volume among the manufacturers in order to reap the highest entrant value. We derived from this analysis several important insights for practitioners. Most important, we observed that an entrant always provides at least a short-term value because he introduces competitive pressure, regardless of the entrant's supply risks. However, the value depends heavily on the entrant's characteristics (i.e., costs, capacity, supply risks and in-country registration), and our analyses shed light on how these characteristics impact the entrant's value. In particular, we found that the entrant's in-country registrations have a very strong, perhaps the strongest, effect on the entrant's value because they reduce competitive pressure and inhibit risk diversification at the same time. Our results also highlight that the PO will benefit from the entrant only if it can determine the right volume split. Because the entrant's value is highly sensitive to the volume split, the simple rules of thumb (e.g. a 70%/30% split) that we regularly encounter in practice are often not the most beneficial for the PO.

Our results can serve global-health decision-makers in several ways. First, our findings help to refute simplified assumptions about how competition and risk impact outcomes. Second, they help to evaluate how changes to important exogenous parameters (e.g., costs, capacity) that are uncertain or difficult to quantify ex-ante influence competition and risk and, ultimately, the entrant's value. Our findings and recommendations are useful not only for decision-makers in global-health POs but also for other stakeholders (e.g., philanthropic investors, technical advisors) who must determine the value of generics manufacturers in donor-funded global-health markets. Finally, generics manufacturers that consider joining new markets can use our findings to tailor their offerings to POs and increase their chances of receiving attractive volume allocations.

Our study has some limitations, the most important of which is that we opted for an approach that is close to the reality of global-health POs. This choice came with the necessity to employ a simulation in order to derive meaningful insights and practically useful recommendations. In addition, our model considers only one incumbent and one entrant. Although we believe that the general intuition of our results would still hold, we cannot provide detailed recommendations on, for example, how to split the volume among an incumbent and two new entrants. Lastly, we do not explicitly account for quality risks in our study because in most instances global-health procurers buy from pre-qualified suppliers. Similarly, perishability was not a concern in the settings that we considered. This may, however not hold in every setting and provides an interesting avenue for future research.

Disclaimer

This research was partially funded by the William-Davidson-Institute (WDI) at the University of Michigan. Support and feedback came from Prashant Yadav and Leslie Arney at WDI, and our work also benefited from discussions with individuals at the Gates Foundation, i+Solutions, the Reproductive Health Supplies Coalition, USAID, and UNFPA (in alphabetical order) and from feedback from participants at the 2015 Reproductive Health Supplies Coalition Annual Meeting in Oslo.

Chapter 3

The Postponement Tender and its Effect on Competition and Learning

Motivated by procurement challenges faced by global-health buyers like the Global Fund and the United States Agency for International Development (USAID), this paper addresses a new tendering mechanism, the *postponement tender*, that seems to be particularly suitable when a buyer wants to incentivize new suppliers (e.g., generics suppliers) to bid on a large procurement contract despite the considerable supply risk that results from not having reliable information about the quality of the new suppliers. The postponement tender can be considered as a procurement auction in which eligible suppliers bid for an overall volume, receive some initial, partial volume for sure, and may also win the remaining (postponed) volume, depending on their bids and the quality of the initial volume they deliver. Our formal and numerical analyses show under which conditions a postponement tender has greater benefits for a buyer than a traditional single-sourcing, winner-take-all, format. We find that a postponement tender can lower purchasing prices by increasing competitive pressure among suppliers while reducing the buyer's quality risk. Our analyses allow us to derive optimal postponement policies and optimal sourcing strategies for market structures that differ in terms of the size of the buyer, suppliers' value, and quality uncertainty.

3.1 Introduction

Inducing supply-side competition is a key concern of most buyers in the private and public sectors because intense competition among suppliers is likely to drive down purchasing prices. Many buyers of specialized goods rely on competitive tendering to stimulate competition among potential suppliers, but the outcome of a tendering process depends heavily on the number of potential suppliers that participate in the tender. In the most general terms, we can assume that the intensity of competition increases with the number of competitors,[1] so buyers are typically interested in creating incentives for potential suppliers to enter the tendering process.

However, in many instances, lowering purchasing prices is not a buyers' only objective. Especially when it comes to the procurement of specialized (i.e., non-commoditized) goods, supplier quality in terms of product quality and delivery service is also of concern because these factors may vary substantially among suppliers. From the buyer's perspective, the prices (bids) and quality of the competing suppliers are uncertain at the beginning of the tendering process. While price uncertainty is resolved by the end of the process, when competitors' bids are known to the buyer, uncertainty about different dimensions of quality persists. Eventually, the buyer has to evaluate the trade-off between known supplier prices and uncertain supplier quality. Clearly, it would be beneficial for the buyer to employ a tendering mechanism that maintains competitive pressure while reducing the buyer's uncertainty about suppliers' quality. In this paper we study such a mechanism that we term "postponement tender".

Our study is motivated by the challenges faced by global-health buyers like the Global Fund (GF), GAVI, and the United States Agency for International Development (USAID), which consolidate demand for life-saving essential medicines (e.g., treatments for malaria, HIV/AIDS, or tuberculosis, vaccines, or reproductive health products) in low-income countries, issue tenders and manage the tendering process, negotiate contractual terms with pharmaceutical manufacturers, and take an active role in ensuring that the purchased medicines reach the population. These global-health buyers encourage new generics suppliers to bid on their tenders to increase supply-side competition, bring down purchasing prices, and ultimately to supply more people in low-income countries with pharmaceutical products (e.g. USAID, 2014; Gavi, 2016). Despite the potential benefits of these new generics suppliers, global-health buyers are often hesitant to contract with them because they fear lower quality in terms of, for example, product quality or de-

[1]For example, in a first-price auction, the equilibrium bid approaches a bidder's real valuation (or cost) as the number of bidders increases (Krishna, 2009, p. 16).

livery performance. This obstacle to contracting generics suppliers is reinforced by the lack of reliable information about their quality, as the buyer may have little or no prior experience with these new suppliers, making it difficult to anticipate their quality (e.g. Global Drug Facility, 2013, 2014). Consider the GF, who consolidates low- and middle-income country demand for antimalarial drugs, HIV medications, and tuberculosis drugs (Global Fund, 2014b, 2017c, 2018). In the past, for example, Artemether combination therapies for treating Malaria were bought almost exclusively from Novartis who had undergone WHO pre-qualification and registered its product in the countries procuring through the GF. In 2009 the GF procured Artemether combination therapies for $29.7 Mil.; approximately 97% of the volume was purchased from Novartis and most of the remaining volume from one generics manufacturer (Global Fund, 2017b). The GF wanted to contract more generics suppliers to spur competition and develop a more diversified and reliable supply base and established a new tendering procedure. In multiple RFPs for anti-malarial drugs the GF announced that "For certain [anti-malaria] products, the Global Fund may reserve a portion of available volumes for subsequent negotiation with newly eligible suppliers and with existing [...] Suppliers that can offer products that become compliant with the Global Fund Quality Assurance Policy [...] after the close of this RFP" (Global Fund, 2017c, p. 22). Depending on the supply market situation of a specific treatment regime, the GF would reserve 5% to 30% of the procurement volume to motivate new suppliers to join. The initial intent of the withheld volume was to provide an incentive for new suppliers to enter the market and to ensure that the market is not closed during the entire period of the contract (typically two to three years). Interestingly, the first time they introduced this mechanism in 2014, they already had attracted more suppliers and a broader supply base for many of their treatments. For example, in 2014 the GF procured Artemether combination therapies for $31.8 Mil. from Novartis and five generics suppliers with shares between 7% and 33% (Global Fund, 2017b). As such, it was important to also establish how the withheld volume would be allocated if no new suppliers join. "If no new entrants emerge, this volume will be released to existing [...] Suppliers on a six-monthly basis according to performance"(Global Fund, 2017c, p. 22). The GF considered lead time performance (on time in full, OTIF), and responsiveness to new order requests as critical performance criteria—that is, the GF explicitly analyzed supplier performance during the contract phase and determined which of the panel suppliers would receive the withheld volumes based on their performance. The GF found that postponing a certain share of the procurement volume is not only an effective measure for motivating new suppliers to join their tenders, but also an effective measure to deal with quality uncertainty, because they were able to assess the quality of new (generics) suppliers and to allocate more volume to suppliers with better performance.

Clearly, dealing with quality uncertainty becomes the more important issue as soon as additional suppliers partake in the tenders, but it is unclear how much of the overall volume should be reserved for later allocation. In this study we focus on the aforementioned trade-off between suppliers' prices and uncertain quality and address the question concerning if and how much volume should be reserved and allocated later ("postponed") when there are new suppliers with uncertain quality. Motivated by the example of the GF, we propose a new tendering mechanism that we term a "postponement tender". With a postponement tender buyers seek to strike a balance between the benefits of competition and the quality risks induced by new entrants by withholding a share of the overall procurement volume and allocating it after they have received additional information about suppliers' quality by means of testing (USAID, 2010; UNFPA, 2012, 2015) and evaluating logistics performance (RHInterchange, 2016; European Parliament Research Service, 2020).

Suppliers bidding in the initial award will consider how likely it is that they can receive the postponed volume. Since the Global Fund evaluates both performance and cost, the suppliers will consider the interplay of their own and their competitors' performance and the respective bids. The postponed volume gives the buyer flexibility in volume allocation and an additional lever with which to increase competition. Of course, a buyer may also consider splitting a single tender into multiple consecutive tenders, learning about suppliers' quality in the first tender, and using this information in the subsequent tenders. However, the procurement cycle length is often insufficient to justify multiple consecutive tenders (e.g., with a duration of one year), especially in light of the time and money that the buyer and the suppliers must spend conducting and evaluating the tenders. This paper sheds light on the mechanics and outcomes of a postponement tender and compares it to a typical single-sourcing arrangement under which the "winner" receives the entire volume. We abstract from the many other incentives that may be introduced through withholding volume and focus on the trade-off between the buyer's quality and costs.

The main benefit of this mechanism is the *learning effect*, that is, the buyer's opportunity to learn about the suppliers' quality from the initial volume and then to use this information in making an informed decision about to which supplier to award the postponed volume. Even without considering competition, the effect of learning is interesting: the buyer has a larger opportunity to inspect the delivered products and learn about the suppliers when awarding larger initial volumes. A large initial volume, however, implies a small postponed volume, which impacts how much the buyer benefits from learning. That is, there is a trade-off between the amount of learning and the benefits of learning.

The postponed volume also has a pronounced effect on competition. Intuitively, a larger postponed volume should increase competition and lower prices. After the buyer publishes the relevant tender documents, the suppliers know that the buyer has to award some of the initial volume to each of the suppliers under consideration, or else she cannot learn about quality. This implies that suppliers will compete primarily for the larger postponed volume, incenting them to submit lower bids. However, as we will show, this intuition is not universally true, as the interaction between learning effects and competition effects is complex.

We study how the learning and competition effects jointly impact the outcome of a postponement tender for a buyer that faces an incumbent supplier with known quality and an entrant with uncertain quality, and contrast that outcome to the outcome of a single-sourcing arrangement. In so doing, we address two interrelated questions that are of particular importance to buyers: (1) Under what conditions should a postponement tender be preferred to single-sourcing? (2) If it is optimal to employ a postponement tender, how much volume should be postponed?

We identify relevant conditions for the design of a postponement tender and consider four distinct market scenarios that are characterized by different types of buyers (large vs. small) and different types of supply market structures (homogeneous vs. heterogeneous suppliers). Large buyers (e.g., the GF, GAVI, or UNICEF's supply division) consolidate the demand of many recipient countries and therefore have large procurement volumes, as evidenced by the example of Artemether combination therapies bought by the GF. Smaller buyers are, for example, governmental procurers of low-income countries with smaller populations and social marketing and relief organizations that operate decentralized procurement departments in certain countries or regions (e.g., Marie Stopes International). We use the suppliers' values to characterize the supply market structure; loosely speaking, the supplier value is the difference between expected quality and cost of a supplier (see Section 3.3. for a formal definition) that the buyer considers when awarding the postponed volume. The buyer faces homogeneous suppliers when both the incumbent and the entrant have very similar values, which is likely the case when both are either branded manufacturers or generics manufacturers. In contrast, the buyer faces heterogeneous suppliers when the incumbent's and the entrant's values differ, which may, for example, be the case when the incumbent is a branded manufacturer and the entrant is a generics manufacturer.

To answer our two research questions, we first formalize the learning and competition effects of the postponement tender for the case of two suppliers, an incumbent and an entrant, in individual models and study their structural properties (Section 3.3.1 and Section 3.3.2). In particular, we characterize separately the postponement quantities that maximize the buyer's expected a priori quality and the buyer's expected costs. In Section 3.3.3 we address the overall problem of determining the postponement quantity that maximizes the buyer's expected utility—that is, the difference between the expected a priori quality and the expected procurement costs. Based on analytical and numerical analyses we find two surprising results: (1) a postponement tender should always be preferred over a single-sourcing auction format when suppliers are heterogeneous in terms of their values, (2) the buyer should exclusively pursue either a "learning strategy" by choosing the postponed volume that maximizes her expected a priori quality, or a "competition strategy" by choosing the postponed volume that minimizes her expected procurement costs (i.e., one of the two solutions established in Sections 3.3.1 and 3.3.2). The reason for why a postponement tender should be preferred in most cases is that it induces an additional element of competition. While competition is usually seen only as the pressure that emerges from another supplier lowering his price, we can show that under a postponement tender a supplier may lower his bid even if the competitor does not, as the mere threat of losing the postponed volume after the buyer learns about the entrant's quality introduces competitive pressure. We are also able to identify the conditions under which it is optimal to pursue a learning or a competition strategy, and we find that the choice of the optimal strategy depends on the overall procurement volume and suppliers' values. We use our formal and numerical results and insights to derive managerial implications, especially for buyers in the global-health domain, and tie the postponement strategies to the types of buyers and the supply market structure (Section 3.4). For example, our results show that a large buyer should pursue a competition strategy if she faces heterogeneous suppliers and should only opt for a learning strategy when suppliers are homogeneous and quality uncertainty is high.

While our study of the postponement tender was motivated by our work with global-health procurement organizations, we believe that the postponement tender can be employed successfully in various other settings. The postponement tender appears to be particularly relevant in cases where the supply base is weak and the quality of some of the suppliers is uncertain. Clearly, this is the case for new product introductions in which new suppliers are following the innovator. However, products that have been in the market for a long time might also be good candidates for the postponement tender. For example, during the COVID-19 pandemic, the United States and the European Union faced shortages of critical medicines, medical devices, and personal protective equipment. In some of theses cases the reason for the shortage was that supply is heavily concentrated and only produced by one or two manufacturers in Asia. As a result, countries are considering, among other things, to contract new suppliers to diversify the supply base (European Parliament Research Service, 2020), which may expose procurers to uncertain quality and making a postponement tender a relevant option.

3.2 Literature

The postponement tender strikes a balance between purchasing costs that are driven largely by competition between suppliers and the quality risks faced by the buyer by postponing the decision to allocate a portion of the procurement volume. The idea to postpone a decision and acquire more information has been used in other contexts, for example, to delay product differentiation and observe more accurate demand signals (Lee and Tang, 1997; Swaminathan and Tayur, 1998; Swaminathan and Lee, 2003), or to postpone pricing decisions (Van Mieghem and Dada, 1999). The way the postponement tender is structured makes it a specific form of an auction in which the suppliers bid for a known overall volume, but are aware of the fact, that they may receive only an initial volume, and may or may not—depending on their bids but also on the quality of their initial volume—win the postponed volume. Therefore, our work is related to procurement auctions, more specifically: procurement auctions with uncertainty about suppliers' quality. However, the postponement tender can also be viewed in light of the single-sourcing vs. dual/multiple-sourcing problem that has received considerable attention in research and among practitioners (see Minner (2003) and Snyder et al. (2016) for reviews). In fact, the postponement tender can be considered a hybrid mechanism that combines elements of both single- and dual-sourcing. If volume postponement is feasible, the buyer has a dual-sourcing option, as described by Yang et al. (2012). For an initial volume that is greater than zero but lower than the total volume, the buyer dual-sources the initial volume, learns about the suppliers' quality, and single-sources the postponed volume. If the buyer chooses to postpone the entire volume (i.e., the initial volume is zero), the postponement tender "collapses" into a single-sourcing, winner-take-all mechanism in which there is no learning and the supplier with the highest expected value receives the entire procurement volume. The buyer could also choose an initial volume that equals the overall procurement volume—that is, the postponed volume is zero. Then the postponement tender would collapse into a pure dual-sourcing arrangement. Although we allow for this outcome in our analysis, it is an inferior solution, at least in our setting: From Anton and Yao (1989) we know that, in general terms, a “split award” (i.e., a dual-sourcing arrangement) leads to lower competition among suppliers and higher prices than a winner-take-all auction in which one supplier receives the entire volume. Under pure dual-sourcing, the buyer does not benefit in the tender from any learning effects—–she can learn about the quality of the suppliers, but does not benefit from this information in the current tender because she has already allocated the entire procurement volume among the two suppliers. In practice, other reasons may justify dual-sourcing: Gong et al. (2012) show that dual-sourcing can lead to lower bids of the suppliers when suppliers can invest into cost reductions. Chaturvedi et al. (2014) show that splitting the procurement volume may also be beneficial in a multi-period setting with qualification costs, as a split award helps to maintain a certain supply base. The strongest arguments for choosing a “pure” dual-sourcing arrangement instead of a hybrid postponement tender come from the literature on the management of supply risks and disruptions. A large number of papers have addressed the benefits of multiple-sourcing and derived optimal procurement policies in the presence of a multiple or dual-sourcing option and supply disruption risks (e.g., Tomlin and Wang (2005); Tomlin (2006); Dada et al. (2007); Federgruen and Yang (2008); Babich et al. (2007a)). Broadly speaking, these papers showed that the benefit of dual-sourcing comes from risk diversification and/or the possibility that one supplier will compensate for the shortfalls of another. However, as Yang et al. (2012) and Qi et al. (2015) pointed out, these papers did not account for competition among suppliers but typically assumed “that suppliers’ prices and reliability are exogenous” (Qi et al., 2015, p. 91). Yang et al. (2012) consider a setting in which the buyer has a dual-sourcing option in the presence of supply risks and competition among suppliers. The buyer may choose to order from only one supplier, using a winner-take-all strategy, or from both suppliers, using a diversification strategy. The authors find a trade-off similar to that we assumed with regard to our postponement tender, that the more the buyer favors a winner-take-all strategy, the greater the competition benefits and the lower the diversification benefits, and vice versa. Apart from the fact that Yang et al. (2012) consider the disruption risk, modelled as a random yield with either zero or one, and our study’s addressing quality risks in more general terms, the paper differs in their objective of solving for the buyer’s optimal procurement contract, consisting of an order quantity and fixed and variable payments. In contrast, we model the postponement tender as a specific auction format.

Other papers also address dual-sourcing in the presence of supply uncertainty and competition. Qi et al. (2015) study a buyer’s dual-sourcing decision under uncertain demand and unreliable supply of two suppliers. Kumar et al. (2018) employ Bertrand competition ”between two price-setting retailers and analyze how pricing can be used as an important lever under supply disruptions" (Kumar et al., 2018, p. 536). They also provide a good overview of the relevant literature. Our work is particularly relevant to the humanitarian sector where sourcing is a key challenge because of long and uncertain lead times (Komrska et al., 2013), quality issues, uncertain financing (Natarajan and Swaminathan, 2014), or constrained budgets (Taylor and Xiao, 2014). Recently, Lauton et al. (2019) proposed a framework with which to evaluate the trade-off between competition and supply risks and show under which conditions a new entrant supplier can provide value to a buyer in the global-health domain. To capture competition between a buyer and two suppliers, the authors propose a Nash-in-Nash bargaining model that allows them to derive equilibrium prices. The supply risk in their model pertains to lead-time uncertainty. Lauton et al. (2019) show that splitting the volume between an incumbent and an entrant to balance competition and risk is difficult because purchasing costs are concave in the volume split, that is, dual-sourcing reduces competitive pressure, and risk-related costs are convex, as dual-sourcing reduces risk exposure. As a result, overall procurement cost can have multiple optima, and identifying the overall cost minimum is difficult. Iakovou et al. (2014) study different sourcing strategies in humanitarian supply chains and find that dual sourcing can significantly reduce disruption costs in exchange for some premium that needs to be paid.

The research presented in this paper differs from the contributions above in a number of ways. First, we do not consider the "classical" questions concerning whether to single-source or dual-source and how much to allocate to the suppliers. Our postponement tender is structurally different in that it splits the procurement volume across time, allows for learning depending on how the volume is split, and incorporates elements of both single- and dual-sourcing. Simply speaking, we do not consider how to split the entire procurement volume across two or more suppliers so much as how to split the volume into an initial volume and a postponed volume. To the best of our knowledge, research has not addressed this type of mechanism, which appears to be attractive to practitioners. Second, how we consider the concept of (supply) risk differs from the aforementioned research. The literature on dual-sourcing focuses on specific supply-disruption risks, but we assume that the buyer faces multiple risks in terms of, for example, product quality and delivery performance and assume that we can consolidate these uncertain non-price attributes into a single quality measure. This approach is in line with previous literature, for example, Engelbrecht-Wiggans and Katok (2007) and Fugger et al. (2015). In our model, the buyer has beliefs about suppliers' quality at the beginning of the tender and can learn (i.e., update her beliefs) after having received and inspected the initial volume. The way in which we model the buyer's learning is similar to the "Bayesian model of supply learning" proposed by Tomlin (2009).

We draw from the literature on procurement auctions to incorporate the competition element into our model. The postponement tender features a setting in which a buyer wants suppliers to compete and, after the buyer learns about the suppliers' quality, decides who will receive the postponed volume. Therefore, the supplier with the lowest price does not necessarily win the postponed volume because the buyer considers price and quality, and the winner of the postponed volume will only be announced later in the process after the buyer updated her beliefs about the suppliers' quality. Therefore, the postponement tender is structurally similar to a buyer-determined auction, which Jap (2002) defines as an auction in which "the buyer commits to select one winner among the bidders, but selection may occur on any basis other than price" (Jap, 2002, p. 508). Engelbrecht-Wiggans and Katok (2007) compare the buyer-determined auction format to a solely price-based format, where the buyer ignores uncertain non-price attributes in the final evaluation. Katok and Wambach (2008) analyze the effects of supplier collusion in buyer-determined auctions. Wan et al. (2012) consider a setting in which an entrant's qualification is unknown, and the buyer must decide whether the entrant should be screened prior to the auction. Santamaria (2015) compares the buyer-determined auction to the well-established scoring auction and identifies settings in which one or the other format may yield more benefit for the buyer. Because our setting is similar to that considered by Katok and Wambach (2008) and Santamaria (2015), we adopt their dynamic open-bid auction format to analyze a buyer-determined auction setting and to model the competition between the two suppliers.

We integrate the Bayesian updating model into this auction format to model the effect of the buyer's learning on suppliers' competition. Therefore, a key methodological contribution of our paper is our development of an integrated model of learning and competition that allows us to study how learning impacts competition between suppliers and the outcome for the buyer in the setting of a postponement tender. Next, we first introduce and analyze the learning model and the competition model independently. We then develop and study the integrated model.

3.3 The model

Our model addresses the procurement problem of a buyer (she) who faces a demand of M units for a fully divisible good that can be sourced from an incumbent supplier (i) and a new generics supplier, to which we refer as the entrant (e). The buyer bases her procurement decision not only on the procurement costs, but also takes into account non-price attributes like product quality and delivery performance. In line with previous literature (e.g., Engelbrecht-Wiggans and Katok (2007) and Fugger et al. (2015)), we assume that the buyer can consolidate these non-price attributes into a single quantifiable measure, to which we refer as the supplier's quality. We also assume that the buyer has a longstanding relationship with the incumbent and knows the incumbent's quality. We denote the incumbent's (deterministic) quality by q_i. In contrast, the entrant is new to the market, and we assume that the buyer has incomplete information about its quality. We model the entrant's uncertain quality as a random variable Q_e with probability density function f and distribution function F.

The buyer uses a procurement mechanism, the postponement tender, with which the buyer splits the procurement volume M into an initial volume and a postponed volume. After the buyer announces the allocation between initial and postponed volume, the suppliers bid prices at which they are willing to sell their products. The buyer dual-sources the initial volume from both suppliers, so both receive some share of the initial volume. After the buyer has received and inspected the initial volumes, she updates her beliefs about the entrant's quality and single-sources the postponed volume from the supplier that maximizes the buyer's expected utility (the difference between the supplier's expected quality and his bid).

We denote the postponed share of the entire volume M by $d_p \in [0, 1]$. Thus, the volume of the initial award is $(1 - d_p)M$, and the postponed volume is d_pM. Essentially, for $d_p < 1$, the buyer employs a postponement tender, and for $d_p = 1$, she employs single-sourcing. We assume that the initial volume $(1 - d_p)M$ is split equally between the two suppliers into $d_i = d_e = \frac{1-d_p}{2}M$. In practice, a buyer may consider splitting the initial volume unequally between the suppliers to increase learning effects, for example. Section 7.1.3 in the Appendix investigates the impact of an unequal initial volume split and shows that our results do not change structurally.

Figure 3.1 summarizes the timing of events in our model. First, at time t_0, the buyer announces the overall procurement volume M and the postponed volume d_pM. The suppliers then bid for the entire contract, so their price bids apply to both the initial volume and the postponed volume, although they cannot be sure that they will be awarded the postponed volume. We denote the outcome of this bidding process, that is, the terminal bids of the incumbent and the entrant, by $b_j^T(d_p)$ ($j \in \{i, e\}$). The terminal bids depend on d_p. When suppliers bid for the overall contract, they assess the "value" of the initial volume $\frac{1-d_p}{2}M$ and the postponed volume d_pM and their probability of winning the latter. Section 3.3.2 details the bidding process and specifies how the terminal bids depend on the postponed volume.

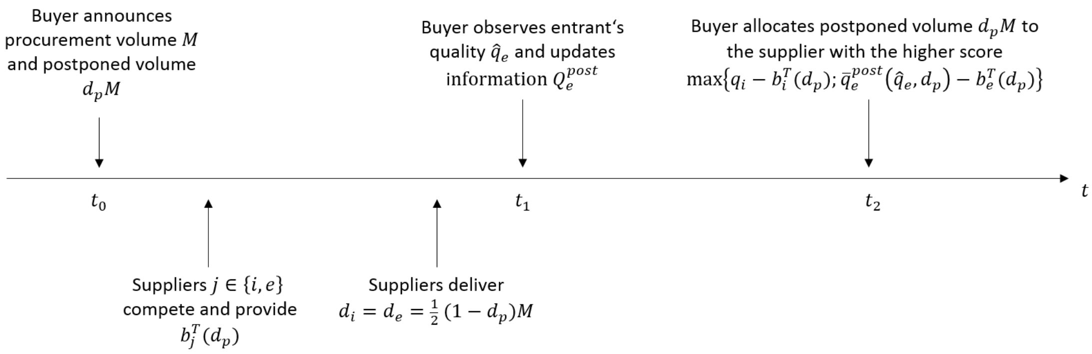


Figure 3.1: Timing of events

After the terminal bids are placed, the suppliers deliver their initial volumes $\frac{1-d_p}{2}M$. At time t_1, the buyer observes the entrant's quality. We denote by $\hat{q}_e$ the fraction of the entrant's share $\frac{1-d_p}{2}M$ that is of acceptable quality. Based on this observation, the buyer updates her prior information Q_e about the entrant's quality. We denote by $Q_e^{post}(\hat{q}_e, d_p)$ the buyer's updated quality information. The number of units the buyer observes determines her ability to learn about the entrant's quality such that, the larger the initial volume, the more units the buyer can inspect and the more accurate the quality signal will become (Berry, 2006). Thus, $Q_e^{post}(\hat{q}_e, d_p)$ depends on $\hat{q}_e$ and d_p. Section 3.3.1 specifies how we model the buyer's updating/learning.

Our model allows for two interpretations of the procurement volume M. As suggested in our previous description, M can simply represent the total number of units the buyer intends to procure. Under this interpretation each unit's quality can be observed individually by the buyer and provides unique information about a supplier's quality. More generally, M can also be interpreted as the number of learning opportunities a buyer has, for example, the number of orders or individual lots a buyer receives.[2] It is reasonable to assume that, in this case, the buyer inspects each individual order she receives—she samples a number of units from each order, evaluates their quality and infers the supplier's quality for the order.[3]

At time t_2, after updating the quality information about the entrant, the buyer determines the incumbent's score $q_i - b_i^T$ and the entrant's (expected) score $\mathbb{E}[Q_e^{post}(\hat{q}_e, d_p)] - b_e^T$, and awards the postponed volume to the supplier with the higher score. In line with the literature on scoring auctions we assume that costs are measured on the same scale

[2]In the *Procurement Performance Indicators Guide* the *USAID* describes "supplier performance" as either the "Percentage of orders in compliance with the contract criteria" or the "Percentage of orders delivered on time". USAID (2013)

[3]For this interpretation we have to assume that each order provides the same amount of information.

as quality and that costs never exceed quality (Engelbrecht-Wiggans and Katok, 2007; Santamaria, 2015; Haruvy and Katok, 2013). We formalize this later in Section 3.3.2. The overall process, as depicted in Figure 3.1, makes clear that the buyer's key decision pertains to setting the postponed volume d_p at time t_0. Both the terminal bids of the suppliers and the expected (overall) quality depend on d_p, which the buyer announces before the suppliers place their bids. We assume a risk-neutral buyer who wants to choose the postponed volume d_p that will maximize her expected utility $\mathbb{E}[U(d_p)]$, that is, the difference between the expected quality of the M units she purchases and the corresponding purchasing costs. Therefore, the buyer's problem can be stated as:

$$\begin{aligned}\max_{0\leq d_p\leq 1}\mathbb{E}[U(d_p)] = &\underbrace{(1-d_p)M\cdot\left(\tfrac{q_i}{2}+\tfrac{\mathbb{E}[Q_e]}{2}\right)+d_pM\cdot\int_0^1 f(\hat{q}_e)V_q(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e}_{\text{Expected Quality}}\\ &-\underbrace{((1-d_p)M\cdot\left(\tfrac{b_i^T}{2}+\tfrac{b_e^T}{2}\right)+d_pM\cdot\int_0^1 f(\hat{q}_e)V_c(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e)}_{\text{Expected Costs}}\end{aligned} \tag{3.1}$$

$$\text{where } V_q(\hat{q}_e,d_p)=\begin{cases} q_i, & \text{for } q_i-b_i^T>\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T\\ \bar{q}_e^{post}(\hat{q}_e,d_p), & \text{for } q_i-b_i^T<\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T\end{cases}$$

$$\text{and } V_c(\hat{q}_e,d_p)=\begin{cases} b_i^T, & \text{for } q_i-b_i^T>\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T\\ b_e^T, & \text{for } q_i-b_i^T<\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T\end{cases}.$$

$\hat{q}_e$ in Eq. (3.1) denotes the entrant's quality that the buyer observed from the initial volume. $V_c(.)$ and $V_q(.)$ denote the price and expected quality per unit of the postponed volume, both depending on the realization of entrant quality in the initial volume. At the beginning of the process t_0, when the buyer sets d_p, this observation is uncertain; the buyer knows its realization $\hat{q}_e$ only after receiving and inspecting the entrant's initial volume.

Eq. (3.1) provides a relatively straightforward characterization of the buyer's problem. However, it entails an intricate interplay of competition and learning effects—both the expected quality and the expected costs depend on the buyer's learning, which is dependent on the postponed volume—that is difficult to disentangle, making it difficult for the buyer to determine her optimal postponed volume d_p. To shed light on this interplay we first analyze the two elements of the buyer's expected utility separately: in Section 3.3.1 we focus on the expected quality, study how the buyer benefits from learning about the entrant's quality based on the buyer's choice of the postponed volume d_p, and determine the postponed volume that maximizes the buyer's expected a priori quality. In Section 3.3.2 we study the terminal bids of the incumbent and the entrant and how these depend on the buyer's choice of the postponed volume d_p. In this section we also derive the postponed volume that minimizes the buyer's expected procurement costs. Section 3.3.3 uses these results to characterize the buyer's optimal postponement decision and find a surprising result: it is optimal for the buyer to pursue either a learning strategy and choose the quality-maximizing postponed volume (established in Section 3.3.1), or to pursue a competition strategy and to choose the cost-minimizing postponed volume (established in Section 3.3.2).

3.3.1 Buyer's learning and the quality-maximizing postponement volume

The buyer is uncertain about the entrant's product quality with regard to both the initial volume and the postponed volume. However, we assume that the buyer can learn from the quality the entrant provides in the initial award. To model the buyer's learning, we use a Bayesian updating approach that is similar to that of Tomlin (2009). Our model differs from Tomlin (2009)'s approach in that we assume the number of units observed by the buyer to be endogenous. Simply put, in our model, learning depends on the buyer's postponement decision. We model the uncertain product quality as a Bernoulli random variable. Suppose that each unit of the initial volume $\frac{1}{2}(1-d_p)M$ the buyer receives from the entrant is acceptable ($\rho_e = 1$) with probability $Q_e \in [0,1]$, and is unacceptable ($\rho_e = 0$) with probability $(1-Q_e)$, where Q_e is drawn from a probability distribution with probability density function f and distribution function F.

We assume that the entrant's quality follows a Beta-distribution $Beta(\alpha, \beta)$. Thus,

$$f(q_e) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(q_e)^{\alpha-1}(1-q_e)^{\beta-1}.$$

We choose the Beta-distribution because it is particularly suitable for modeling product quality and quality updating. It can represent a wide range of distributions, including the uniform distribution Beta($\alpha = 1, \beta = 1$) and bell-shaped distributions. It also has a bounded support on $[0,1]$, which corresponds well to our notion of uncertain quality. Moreover, the Beta-distribution has a specific property, which is convenient for our purposes, as it allows us to model Bayesian updating because it is a conjugate prior for the entrant's updated quality distribution. That is, if the prior distribution is a Beta(α, β)-distribution, and one conducts a single Bernoulli trial, the posterior distribution is Beta($\alpha+1,\beta$)-distributed if the trial was successful, and Beta($\alpha,\beta+1$)-distributed if the trial was unsuccessful (e.g. Press, 1989; Zhu and Lu, 2004; Tomlin, 2009).

Proposition 2.
Define $\bar{q}_e := \mathbb{E}[Q_e]$. *Suppose the buyer observes* $\frac{1}{2}(1-d_p)M$ *units from the entrant, and a fraction* $\hat{q}_e \in [0,1]$ *is of acceptable quality (i.e.,* $\rho_e = 1$*).*

a) Let Q_e^{post} *denote the posterior distribution of the entrant's quality.*
$Q_e^{post} \sim Beta(\alpha + \hat{q}_e\frac{1}{2}(1-d_p)M, \beta + (1-\hat{q}_e)\frac{1}{2}(1-d_p)M))$.

b) Let $\bar{q}_e^{post}(\hat{q}_e, d_p) := \mathbb{E}[Q_e^{post}]$ *denote the posterior mean quality of the entrant.*

$$\bar{q}_e^{post}(\hat{q}_e, d_p) = \frac{\alpha + \hat{q}_e\frac{1}{2}(1-d_p)M}{\alpha + \beta + \frac{1}{2}(1-d_p)M}. \tag{3.2}$$

c) Let $\mathbb{E}\big[\Delta(d_p)\big] := |\bar{q}_e^{post}(\hat{q}_e, d_p) - \bar{q}_e|$ *denote the buyer's expected learning.* $\mathbb{E}\big[\Delta(d_p)\big]$ *is concave and decreasing in* d_p*, and increasing in* M*.*

All proofs are relegated to appendix II.

Proposition 2 characterizes the buyer's posterior distribution of the entrant's quality and the buyer's expected learning after observing the entrant's initial delivery. From part b) we can see how the observed quality $\hat{q}_e$ and the entrant's initial volume $\frac{1}{2}(1-d_p)M$ determine the buyer's posterior mean of the entrant's quality: $\hat{q}_e$ adjusts the mean, and $\frac{1}{2}(1-d_p)M$ weighs this adjustment. This has an immediate impact on how much the buyer can expect to learn from the entrant's initial volume. As part c) of Proposition 1 states, the buyer's expected learning decreases at an increasing rate as the postponed volume d_p increases. Put differently, a buyer's expected learning is increasing in the initial volume of the entrant $(1-d_p)$ at a decreasing rate. Larger postponed volumes lead to smaller weights on the adjustment of the mean and, as a consequence, to smaller differences between the prior mean $\bar{q}_e$ and the posterior mean $\bar{q}_e^{post}(\hat{q}_e, d_p)$. In addition, expected learning is increasing in M because of a similar rationale: if M increases for a given postponed volume d_p, the initially observed volume $\frac{1-d_p}{2}M$ increases, so the buyer expects to learn more, resulting in a stronger update of the entrant's (prior) quality.

Obviously, the buyer can maximize expected learning by not postponing any volume, that is, when $d_p = 0$. However, the buyer will not enjoy any benefits from learning about the entrant's quality because she has already awarded the entire procurement volume M. Only if the buyer postpones at least some of the volume, such that $0 < d_p < 1$, will she benefit from learning. The buyer will be unable to learn if she opts for single-sourcing and postpones the entire volume. Therefore, we can conjecture that, while expected learning decreases with the postponed volume (see Proposition 2c), the benefits of learning increase in the postponed volume, as long as the buyer does not postpone the entire volume. To shed more light on the interplay of the two effects (expected learning vs. benefits of learning), we study how the buyer should set the postponed volume to maximize her expected a priori quality. In this case the buyer solves the following problem which results directly from Eq. (3.1):

$$\max_{0 \leq d_p \leq 1} \mathbb{E}[Q(d_p)] = (1-d_p)M \cdot \left(\frac{q_i}{2} + \frac{\mathbb{E}[Q_e]}{2}\right) + d_p M \cdot \int_0^1 f(\hat{q}_e) V_q(\hat{q}_e, d_p) \mathrm{d}\hat{q}_e$$
$$\text{where } V_q(\hat{q}_e, d_p) = \begin{cases} q_i, & \text{for } q_i > \bar{q}_e^{post}(\hat{q}_e, d_p) \\ \bar{q}_e^{post}(\hat{q}_e, d_p), & \text{for } q_i < \bar{q}_e^{post}(\hat{q}_e, d_p) \end{cases}. \tag{3.3}$$

Proposition 3.
Assume $Q_e \sim Beta(1,1)$ *and* $\bar{q}_e = q_i$. *Define* $d_p^Q := \operatorname{argmax}\{\mathbb{E}[Q(d_p)]\}$.

a) $\mathbb{E}[Q(d_p)]$ *is concave in* d_p.

b) $d_p^Q = \frac{1}{1+\frac{2}{\sqrt{4+M}}} > \frac{1}{2}$ *and is strictly increasing in* M.

The results presented in Proposition 3 exemplify our conjecture: expected quality has a minimum at $d_p = 1$ and $d_p = 0$, where the buyer either has no opportunity to learn or cannot benefit from learning. (See Figure 3.2, for an illustration of how $\mathbb{E}[Q(d_p)]$ depends on d_p.) For $0 < d_p \leq d_p^Q$, the expected quality increases as more volume is postponed; that is, the additional benefit of learning overcompensates for the decrease in expected learning. Recall that expected learning is decreasing in d_p at an increasing rate (Proposition 1c). Beyond d_p^Q, the decrease in expected learning is so strong that it offsets any benefits from learning —— the buyer learns too little to achieve a higher expected quality —— so expected quality decreases in d_p for $d_p^Q < d_p \leq 1$.

From part b) of Proposition 3, we see that d_p^Q is increasing in the overall procurement volume M because, as M increases, the same absolute number of units can be observed at lower levels of d_p, which makes it more attractive to postpone a larger share, at least from a quality perspective.

We now turn to the question of how differences in the a priori (expected) quality between incumbent and entrant affect learning and expected quality. More precisely, we explore whether our previous results change for $\bar{q}_e \neq q_i$.

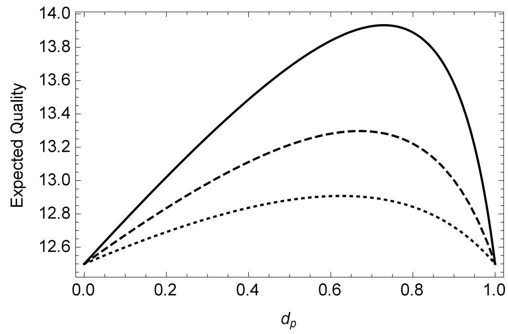


Figure 3.2: Expected quality dependent on d_p: $Q_e \sim Beta(1,1)$ (uniform, solid), $Beta(2,2)$ (symmetric bell-shaped, dashed), $Beta(2,5)$ (asymmetric bell-shaped, dotted) and $\bar{q}_e = q_i$, $M = 25$.

Proposition 4.
Assume $Q_e \sim Beta(1,1)$ and $q_i \neq \bar{q}_e$. Let $\tilde{d}_p = \frac{M-8\Delta_q-2M\Delta_q}{M(1-2\Delta_q)}$ and $\Delta_q := |q_i - \bar{q}_e|$.

a) $\mathbb{E}[Q(d_p)]$ is strictly increasing in d_p for $\tilde{d}_p \leq d_p \leq 1$.

b) $\tilde{d}_p$ is strictly decreasing in Δ_q.

Figure 3.3 shows that the shape of $\mathbb{E}[Q(d_p)]$ changes if the entrant's a priori expected quality and the incumbent's known quality differ. As the figure shows, the curve is no longer fully concave but has an increasing part and a concave part (as in Figure 3.2), as well as a local minimum, after which it is strictly increasing for $\tilde{d}_p \leq d_p \leq 1$. For $\tilde{d}_p \leq d_p \leq 1$, the buyer experiences the same learning effects as before, but does not benefit from learning at these small initial volumes: after updating, the buyer expects the quality of the entrant $\bar{q}_e^{post}$ always to remain below the incumbent's known quality q_i. Therefore, at these values of d_p, the buyer always expects to allocate the postponed volume to the incumbent. As a consequence, each additional unit of postponed volume increases the buyer's expected quality by a constant $\frac{1}{2}(q_i - \bar{q}_e) > 0$. Thus, for $d_p \geq \tilde{d}_p$, the expected quality is linearly increasing in d_p, as stated in Proposition 4a.

The results presented in Proposition 4b suggest that, for an increasing difference Δ_q in the a priori quality, $\tilde{d}_p$ decreases and the region of postponed volumes in which the buyer does not benefit from learning increases. Figure 3.3 illustrates this effect for increasing Δ_q. We observe that, for high Δ_q, the expected quality can even be monotonically increasing in d_p, suggesting that expected quality is maximized at $d_p = 1$ (i.e., when the buyer single-sources from the supplier with the higher a priori quality).

Our analysis shows that the effect of learning on expected quality is non-linear and is determined by the trade-off between a learning effect and the benefits of learning. Since

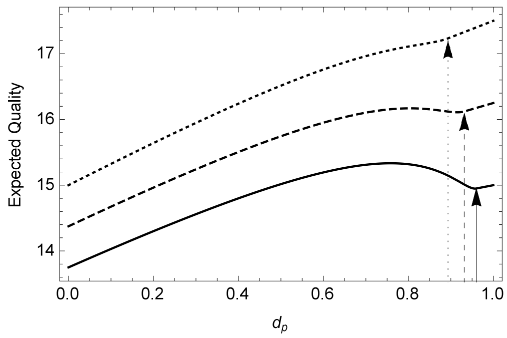


Figure 3.3: Expected quality dependent on d_p for $Q_e \sim Beta(1,1)$, $\bar{q}_e = 0.5 \neq q_i$ and $M = 25$. $\Delta_q = 0.1$ (solid), $\Delta_q = 0.15$ (dashed) and $\Delta_q = 0.2$ (dotted). Vertical arrows depict $\tilde{d}_p$.

both effects have an inverse relationship with the postponed volume, we can identify a d_p that maximizes the buyer's expected quality. The impact of the postponed volume on the expected quality depends on the quality risk of the entrant. Higher upside risks —— that is, higher probabilities for high positive differences in quality in conjunction with the ability to learn about the entrant's quality —— increase the positive impact that postponement can have on the expected quality. While the buyer always benefits from postponement if quality is the same, this is not the case for $q_i \neq \bar{q}_e$. For high quality differences the buyer does not enjoy benefits of learning when choosing very large postponed volumes. Under these circumstances, it may be more attractive for a buyer to single-source ($d_p = 1$) and to forgo any potential learning effects.

3.3.2 Competition and the cost-minimizing postponement volume

The previous section focused on expected quality and how it is impacted by the postponement tender. From a quality perspective, the optimal d_p depends on a number of factors, such as the difference between the entrant's expected quality and the incumbent's quality, its prior distribution, and the overall procurement volume, but our analysis and discussion of expected learning vs. the benefits of learning point to rather high values of d_p and, in certain instances, even to single-sourcing ($d_p = 1$). Intuition suggests that competitive pressure will increase at higher levels of postponed volume and that the incumbent's and the entrant's (terminal) bids will decrease as d_p increases. To determine whether and when this intuition holds true, in this section we introduce a competition model that reflects the negotiation process between the buyer and the two suppliers and that allows us to analyze how the terminal bids $b_i^T(d_p)$ and $b_e^T(d_p)$ depend on d_p. With this we will be able to characterize the postponement volume that minimizes the buyer's expected costs. We will see in Section 3.3.3 that this is an important ingredient for determining the optimal solution to Eq. (3.1). Therefore, in this section we will focus our analysis on the terminal bids and the buyer's expected costs in Eq. (3.1). For notational ease we omit the functional dependence of the terminal bids on d_p and write b_j^T instead.

The choice of our competition model is guided by the current practice of global-health procurers. Procurers of essential medicines for low- and low-middle-income countries are typically large institutional buyers like the Global Fund or UNICEF, smaller governmental buyers, and locally active relief organizations. In most instances, these buyers conduct bilateral negotiations with their suppliers. For example, in our motivating case of the Global Fund's tender of anti-malaria drugs, the Global Fund announced its intention to conduct negotiations with each supplier (Global Fund, 2017c) in a multi-round process during which suppliers would meet bilaterally with the buyer and adjust their bids in a descending auction, where suppliers consecutively submit price quotes publicly. Because of its close resemblance to the real-world negotiation process, we base our competition model on a descending auction format similar to that proposed by Santamaria (2015)[4], although our approach differs from that of the typical descending auction format in that we incorporate the buyer's ability to learn about the entrant's quality after the buyer receives the initial volume.

To keep matters simple, we assume that the incumbent and the entrant both have the same prior Q_e about the entrant's quality, are aware of the buyer's updating mechanism, and will incorporate the buyer's update into their bidding strategies. This assumption is justified if the entrant is a new supplier who has not yet had the opportunity to observe his own quality in a large-scale production process, so he does not have better information about his own quality than the buyer and the incumbent do.

Let $c_j \in [0,1]$ denote the marginal production costs of supplier $j \in \{i,e\}$. For now, we assume that the buyer knows these costs. (In the Appendix in Section 7.1.2 discusses the implications of relaxing this assumption.) The suppliers submit bids b_i^t and b_e^t in multiple rounds $t = 0, ..., T$. The buyer will contract a supplier only if the supplier provides a positive value (in expectation); therefore, the auction starts at $b_i^0 = q_i$ and $b_e^0 = \bar{q}_e$, the buyer's reservation prices. To avoid trivial solutions we assume that $q_i \geq c_i$ and $\bar{q}_e \geq c_e$. In each bidding round $t > 0$, a supplier observes the competitor's bid from the previous round $t-1$ and decides whether to lower his own bid by an increment $\delta > 0$

[4]The descending auction format is also widely used and well established in other procurement domains (Elmaghraby (2007)).

or to stop bidding.[5] We define T as the round in which both suppliers stop decreasing their bids and the auction terminates. Our objective is to characterize the terminal bids b_i^T and b_e^T depending on d_p.

We begin by characterizing the incentives of the incumbent to lower his bid. In bidding round $t \in \{1, \ldots, T\}$ the incumbent has an incentive to lower his bid b_i^{t-1} by δ if the lowered bid increases the incumbent's expected profit:

$$\mathbb{E}[\text{Profit of stopping at } b_i^{t-1}] < \mathbb{E}[\text{Profit of lowering } b_i^{t-1} \text{ by } \delta] \tag{3.4}$$

$$\begin{aligned}
\Leftrightarrow \quad & \overbrace{(b_i^{t-1} - c_i)\frac{1-d_p}{2}M}^{\text{Profit initial volume (current bid)}} + \overbrace{(b_i^{t-1} - c_i)d_p M P[q_i - b_i^{t-1} > \bar{q}_e^{post}(\hat{Q}_e, d_p) - b_e^{t-1}]}^{\text{Expected profit postponed volume (current bid)}} \\
< \quad & \underbrace{(b_i^{t-1} - \delta - c_i)\frac{1-d_p}{2}M}_{\text{Profit initial volume (lower bid)}} + \underbrace{(b_i^{t-1} - \delta - c_i)d_p M P[q_i - (b_i^{t-1} - \delta) > \bar{q}_e^{post}(\hat{Q}_e, d_p) - b_e^{t-1}]}_{\text{Expected profit postponed volume (lower bid)}}
\end{aligned} \tag{3.5}$$

Eq. (3.5) compares the incumbent's expected profit from stopping at the current bid b_i^{t-1} with the expected profit when lowering the bid b_i^{t-1} by δ. The total expected profit is the sum of the (deterministic) profit earned with the initial volume and the expected profit associated with the postponed volume. The latter is determined by the probability of winning the postponed volume $P[\cdot]$, which the incumbent will win only if his score $(q_i - b_i^{t-1})$ is larger than the entrant's expected score $(\bar{q}_e^{post}(\hat{Q}_e, d_p) - b_e^{t-1})$, which is the difference between the posterior mean and the entrant's bid. Note that because $P[\cdot]$ depends on the posterior mean $\bar{q}_e^{post}(\hat{Q}_e, d_p)$, the incumbent explicitly accounts for the buyer's learning. Lowering the bid by δ has two consequences for the incumbent: his margin will decrease and his probability of winning the postponed volume will increase. The incumbent will decrease his bid by δ if the net effect on the expected profit is positive. The entrant's condition for lowering his bid b_e^{t-1} by δ follows the same rationale and can

[5]We assume that the suppliers strictly follow the descending auction format, do not collude as in Fugger et al. (2015), and do not consider economic implications beyond the current tender.

be expressed as:

$$\mathbb{E}[\text{Profit of stopping at } b_e^{t-1}] < \mathbb{E}[\text{Profit of lowering } b_e^{t-1} \text{ by } \delta] \Leftrightarrow \tag{3.6}$$

$$\overbrace{(b_e^{t-1} - c_e)\frac{1-d_p}{2}M}^{\text{Profit initial volume (current bid)}} + \overbrace{(b_e^{t-1} - c_e)d_p M P[\bar{q}_e^{post}(\hat{Q}_e, d_p) - b_e^{t-1} > q_i - b_i^{t-1}]}^{\text{Expected profit postponed volume (lower bid)}}$$
$$< \underbrace{(b_e^{t-1} - \delta - c_e)\frac{1-d_p}{2}M}_{\text{Profit initial volume (lower bid)}} + \underbrace{(b_e^{t-1} - \delta - c_e)d_p M P[\bar{q}_e^{post}(\hat{Q}_e, d_p) - (b_e^{t-1} - \delta) > q_i - b_i^{t-1}]}_{\text{Expected profit postponed volume (lower bid)}} \tag{3.7}$$

Proposition 5.
Let $P(b_i^{t-1}|d_p) := P[q_i - b_i^{t-1} > \bar{q}_e^{post}(\hat{Q}_e, d_p) - b_e^{t-1}]$ denote the incumbent's probability of winning the postponed volume with a bid b_i^{t-1} and let $P(b_e^{t-1}|d_p) := P(\bar{q}_e^{post}(\hat{Q}_e, d_p) - b_e^{t-1}) > q_i - b_i^{t-1}]$ denote the entrant's probability of winning the postponed volume with a bid b_e^{t-1}. Supplier $j \in \{i, e\}$ lowers his bid b_j^{t-1} by δ in round $t \in \{1, \ldots, T\}$ if and only if

$$b_j^{t-1} - c_j > \underbrace{\delta \frac{\frac{1}{2}\frac{1-d_p}{d_p} + P(b_j^{t-1} - \delta|d_p)}{P(b_j^{t-1} - \delta|d_p) - P(b_j^{t-1}|d_p)}}_{\textit{supplier's markup} := \theta_j^t}, \quad \textit{for } 0 < d_p \leq 1, \ t \in \{1, \ldots, T\}. \tag{3.8}$$

Proposition 5 presents the condition under which a supplier will lower his bid in round t. The supplier will lower his bid by δ if his (per unit) margin $(b_j^{t-1} - c_j)$ in the previous round $t-1$ is larger than what we term the "markup" (θ_j^t) for round t. Otherwise he stops bidding. The markup in Eq. (3.8) features two interrelated elements: First, a supplier's decision to lower his bid depends on the relative size of the postponed volume $\frac{1-d_p}{d_p}$. Larger postponed volumes d_p result in low ratios of $\frac{1-d_p}{d_p}$, which, ceteris paribus, lead to lower markups. We refer to this as a *volume incentive*. Second, a supplier evaluates the probability of winning the postponed volume when lowering the bid $(P(b_j^{t-1} - \delta|d_p))$ and evaluates how lowering the bid improves his probability of winning the postponed volume $(P(b_j^{t-1} - \delta|d_p) - P(b_j^{t-1}|d_p))$. These probabilities depend on d_p through the buyer's learning process: As shown in Proposition 2, the buyer's posterior mean is a function of d_p, but since we do not know the direction in which the buyer will update her prior, there is no unidirectional relationship between the probabilities $P[.]$ and d_p. Thus, ceteris paribus, a low probability and/or a large improvement in the probability of winning the postponed volume result in a low markup. We refer to this as the as *risk incentive*. A high volume incentive and a high risk incentive lead to low markups, making it more attractive for a supplier to lower his bid.

Proposition 5 provides insight into a supplier's decision to lower his bid in a specific bidding round t, but we are more interested in the auction's outcome, that is, in the terminal bids b_j^T. A supplier stops lowering his bid if the current margin is equal to or below the markup; that is, in bidding round $t = T - 1$, the bidding condition in Eq. (3.8) is no longer fulfilled ($b_j^{T-1} - c_j \leq \theta_j^T$), which leads to a terminal bid $b_j^T = b_j^{T-1}$ and a terminal markup θ_j^T. The buyer pays a premium $b_j^T - c_j$ on top of the supplier's marginal cost c_j. For $\delta \to 0$, this premium is equal to the markup θ_j^T, so it is dependent on the volume and the risk incentive that the buyer sets by choosing a particular d_p. From the terminal bid b_j^T and the terminal markup θ_j^T, we can also infer the terminal probability of winning the postponed volume $P(b_j^T|d_p)$ for a given d_p, which will play an important role in our later analyses.

We now analyze how the suppliers' terminal bids depend on the postponed volume. We term the difference between the a priori (expected) quality and the costs of the suppliers as the "supplier value" and denote it by $v_i = q_i - c_i$ and $v_e = \bar{q}_e - c_e$, respectively. As in our analysis in Section 3.3.1, we begin by presenting the results for homogeneous suppliers, that is, suppliers' with the same values ($v_i = v_e$). Thereafter, we study how the terminal bids change when suppliers are heterogeneous (i.e., when $v_i \neq v_e$).

PROPOSITION 6.
Assume $Q_e \sim Beta(1, 1)$ and $v_i = v_e$.

a) There exists a threshold level $\underline{d}_p$. For $d_p \leq \underline{d}_p$ suppliers never lower their bids so that $b_i^T = q_i$ and $b_e^T = \bar{q}_e$.

b) For $d_p > \underline{d}_p$ terminal bids b_i^T and b_e^T are strictly decreasing in d_p.

c) $\underline{d}_p$ is strictly decreasing in v_j, $j \in \{i, e\}$, and strictly increasing in M.

Proposition 6 characterizes the incumbent's and the entrant's terminal bids when both suppliers have the same value, that is, when $v_i = v_e$. (See Figure 3.4 for an illustration.) The results presented in Proposition 6a suggest that both suppliers will bid their respective reservation prices $b_i^T = q_i$ and $b_e^T = \bar{q}_e$ and will not lower their bids for postponed volumes below a threshold $\underline{d}_p$. For postponed volumes below this threshold, the volume incentive and the risk incentive are low, resulting in a markup in the first bidding round that is so high that even in the first round the bidding conditions in Eq. (3.5) and Eq. (3.7) are not fulfilled. Only a postponed volume above the threshold

($\underline{d}_p < d_p$) provides sufficient volume and risk incentives; at these levels of d_p, the bidding conditions in Eq. (3.5) and Eq. (3.7) are met, and both suppliers lower their bids. Because the volume and the risk incentive both increase in d_p, terminal bids are strictly decreasing in d_p.

According to part c) of Proposition 6, the threshold $\underline{d}_p$ is decreasing in the supplier's values v_j, suggesting that suppliers with higher values will start to lower their bids at lower postponed volumes d_p. This effect of v_j becomes clear if we consider the condition in Eq. (3.8) in the first bidding round ($t = 1$): A supplier will start to quote bids below the reservation price if the supplier's value exceeds the markup in the first bidding round, that is, $b_j^0 - c_j = \mathbb{E}[q_j] - c_j = v_j \geq \theta_j^1$. We show in our proof of Proposition 6 that θ_j^1 is strictly decreasing in d_p, so if the left-hand side of the bidding condition increases, the threshold of d_p, at which the supplier starts to lower his bid, decreases. Proposition 6c also indicates that the lower threshold $\underline{d}_j^p$ increases in M. A larger M enables a buyer to learn more (for a given postponed volume d_p) because the buyer can observe more units. (See Proposition 2c.) The increase in expected learning decreases the probability of a supplier's winning the postponed volume in the bidding condition in Eq. (3.8) for a given d_p; that is, a supplier's lowering his bid by δ does not increase the probability of winning the postponed volume as much as it would for a lower M. For a higher M, suppliers will start quoting lower bids only for higher postponed volumes.

Proposition 6 suggests that a buyer can expect decreasing bids if she chooses a postponed volume that provides suppliers with sufficient economic incentives to lower their bids. Competition will be the strongest for $d_p = 1$, that is, if the buyer chooses single-sourcing, which is in line with previous research (e.g., Perry and Sákovics (2003) and Gong et al. (2012), who also find that suppliers provide the lowest bid in case of single-sourcing). However, a buyer must carefully choose the postponed volume so as not to undercut the threshold $\underline{d}_p$; if the buyer postpones too little volume, she will not benefit from competition and will be left paying reservation prices.

Corollary 1 follows from Proposition 6 and establishes first results for suppliers whose values v_j differ.

Corollary 1.
$\underline{d}_p^j < \underline{d}_p^{-j}$ *if* $v_j > v_{-j}$.

Suppliers whose values differ will start lowering their bids at different levels of d_p. Because $\underline{d}_p^j$ is decreasing in v_j (see Proposition 6c), supplier j, with the higher value $v_j > v_{-j}$, has a lower threshold $\underline{d}_p^j$ than his competitor, supplier $-j$, has. Proposition 7 complements these results and characterizes how suppliers will bid for postponed volumes

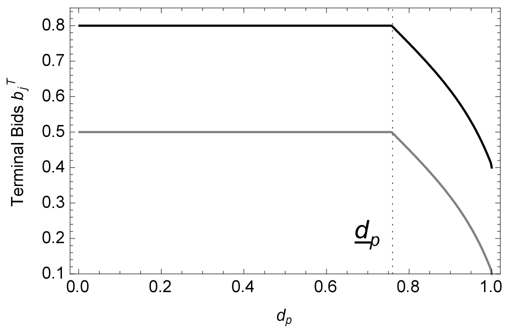


Figure 3.4: Terminal bids (b_j^T) of incumbent (black) and entrant (gray) depending on the postponed volume (d_p) for $Q_e \sim Beta(1,1)$, $q_i = 0.8$, $\overline{q}_e = 0.5$, $v_i = v_e = 0.4$, and $M = 25$.

$\underline{d}_p^j < d_p \leq 1$ if their values v_j differ. To ensure analytical tractability for the case of different supplier values, we assume, in Proposition 7 that supplier $-j$ has a value of zero and supplier j has a value greater than zero. This can be considered an extreme case of heterogeneous suppliers in which only one supplier can extract a positive margin.

Proposition 7.
Assume $Q_e \sim Beta(1,1)$ and $v_j > v_{-j} = 0$, $j \in \{i,e\}$.

a) There exists a threshold level $\overline{d}_p^j$ with $\underline{d}_p^j < \overline{d}_p^j < 1$. For $\underline{d}_p^j < d_p \leq \overline{d}_p^j$, b_j^T is strictly decreasing in d_p while for $\overline{d}_p^j < d_p \leq 1$, b_j^T is strictly increasing in d_p. $b_{-j}^T = \mathbb{E}[q_{-j}]$ for $0 \leq d_p \leq 1$.

b) The threshold $\overline{d}_p^j$ is strictly decreasing in v_j and strictly increasing in M.

Considering our previous results in Corollary 1 and the results established in Proposition 7a, we can characterize how the bids of suppliers with different values v_j depend on the postponed volume d_p. Figure 3.5 provides an illustration. For values below their individual (lower) thresholds $\underline{d}_p^j$, both suppliers will bid their reservation prices; the supplier with the positive value will lower his bids at postponed volumes higher than $\underline{d}_p^j$ but only up to the upper threshold $\overline{d}_p^j$. For $\overline{d}_p^j < d_p \leq 1$, the terminal bid of this supplier increases as d_p increases. The bids of the supplier with a value of zero are independent of d_p, as he will always quote his reservation price. Because our previous analysis suggested that the volume incentive lowers the supplier's markup and makes it more attractive to bid more aggressively for larger postponed volumes, it seems counterintuitive that the bids of the supplier with a positive value will increase for larger postponed volumes. However, for $\overline{d}_p^j < d_p \leq 1$, the risk incentive works in the opposite direction and offsets the effect of the volume incentive on the markup. Eq. (3.5) showed that the supplier will lower his bid by δ if the increase in the probability of winning the postponed volume leads to a higher expected profit than the profit the supplier expects if he remains with his previous bid. We know that the probability of winning the postponed volume depends on the incumbent's score $(q_i - b_i^T)$ and the entrant's expected score $(\overline{q}_e^{post}(\hat{q}_e, d_p) - b_e^T)$, after the buyer's update of the entrant's quality. Assume, without loss of generality, that $v_i > 0$ and $v_e = 0$. The entrant will bid his reservation price $b_e^T = \bar{q}_e$, and if the entrant delivers perfect quality in the initial volume $\frac{1}{2}(1 - d_p)M$, that is, if the realization of $\hat{Q}_e$ is 1, the entrant's maximum score is $\overline{q}_e^{post}(\hat{q}_e = 1, d_p) - \overline{q}_e$ (see Proposition 2). There exists a bid (denoted by b_i^G) that guarantees the incumbent wins the postponed volume after the buyer learns the entrant's quality:

$$q_i - b_i^G = \overline{q}_e^{post}(\hat{q}_e = 1, d_p) - \overline{q}_e \tag{3.9}$$

$$\Rightarrow b_i^G = q_i + \overline{q}_e - \overline{q}_e^{post}(\hat{q}_e = 1, d_p). \tag{3.10}$$

Because $P(b_i^G | d_p) = 1$, the incumbent will never quote a price lower than b_i^G, so b_i^G can be considered a natural lower bound on the incumbent's bid, which is dependent on d_p. The bid b_i^G increases in d_p because, as more volume is postponed, the buyer learns less from the initial volume because the difference between $\hat{q}_e$ and $\overline{q}_e^{post}$ decreases, such that, for very high values of d_p, b_i^G approaches $\overline{q}_i$, the incumbent's reservation price.

The upper threshold $\overline{d}_p^i$ can be interpreted as the postponed volume after which the risk incentive offsets the reduction in the markup that results from the volume incentive. For the case of uniform quality (Figure 3.5), this is the postponed volume at which the incumbent bids b_i^G. Thus, the incumbent's terminal bid b_i^T intersects with b_i^G, $P(b_i^G | d_p) = 1$, and it is not attractive for the incumbent to lower his bid any farther. The results presented in Proposition 7b indicate that, similar to the lower threshold, the upper threshold $\overline{d}_p^i$ decreases in the supplier's value v_i and increases in the procurement volume M. The rationale here is similar to the rationale we presented for the lower threshold: an increasing value impacts the bidding condition in Eq. (3.8) and allows a supplier to bid lower for a given d_p. Thus, the bids drop faster as d_p increases, and the bidding function has a larger negative slope. An increase in the incumbent's value can be the result of a higher quality q_i or lower costs c_i. Clearly, a lower cost does not impact the bid that guarantees the supplier will win the postponed volume (see Eq. (3.10)). An increase in the quality q_i leads to a proportional increase of the bid b_i^G that guarantees the incumbent will win the postponed volume. As a result, an increase in the value lowers $\overline{d}_p^i$, and a supplier begins to increase his bid at lower postponed volumes. Because the buyer's expected learning increases in the volume M (see Proposition 1c), the bid that guarantees the supplier will win the postponed volume decreases (see Eq. (3.10)), so $\overline{d}_p^i$ increases as the procurement volume M increases.

It is interesting to observe that $\underline{d}_p^j$ and $\overline{d}_p^j$ prescribe a "competition interval" that is predominantly determined by the suppliers' values v_j and the procurement volume M. The benefits of competition are not strictly increasing in the postponed volume; instead, the buyer can expect to see decreasing prices only if the buyer chooses a d_p that is within the competition interval, that is for $\underline{d}_p^j < d_p < \overline{d}_p^j$. Contrary to intuition, the buyer inhibits competition if she chooses high postponed volumes ($d_p > \overline{d}_p^j$) since single-sourcing under these conditions will not lead to the lowest bids from both suppliers. The competition interval shifts to the left (i.e., to lower levels of d_p) if the supplier's value v_j increases and the benefits of postponement in terms of competition materialize at lower levels of d_p. At the same time, the negative effects of choosing a d_p that is too high come into effect at lower levels of d_p.

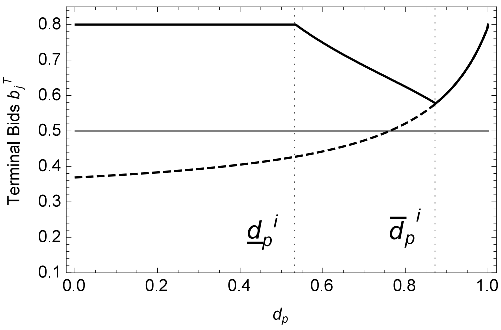


Figure 3.5: Terminal bids (b_j^T) of incumbent (black) and entrant (gray), depending on postponed volume (d_p) for $q_i = 0.8$, $c_i = 0.1$, $Q_e \sim Beta(1, 1)$, $v_i > v_e = 0$ and $M = 25$. The dashed line represents the bid b_i^G that guarantees the incumbent will win the postponed volume.

To ensure analytical tractability, we assumed that one supplier has a positive value and one has a value of zero. Of course, the question arises concerning whether our results from Proposition 7 change if both suppliers have a positive (but different) value. We cannot provide analytical proof, but we can show numerically that our structural insights continue to hold, even if $v_{-j} > 0$ for $j \in \{i, e\}$. (see Appendix 7.1.1)

Section 3.3.1 addressed how a buyer should set the postponed volume to maximize expected quality. Similarly, we now want to determine how the buyer should set the postponed volume in order to minimize her expected procurement costs. A buyer who seeks to minimize the expected procurement costs without considering the expected quality solves the following problem, which is directly based on Eq. (3.1):

$$\begin{aligned}\min_{0\le d_p\le 1} \mathbb{E}[C(d_p)] =&(1-d_p)M\cdot\left(\frac{b_i^T}{2}+\frac{b_e^T}{2}\right)+d_pM\cdot\int_0^1 f(\hat{q}_e)V_c(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e \\ \text{where } V_c(\hat{q}_e,d_p) =& \begin{cases} b_i^T, & \text{for } q_i-b_i^T>\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T \\ b_e^T, & \text{for } q_i-b_i^T<\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T \end{cases}.\end{aligned} \tag{3.11}$$

Proposition 8.
Assume $Q_e \sim Beta(1,1)$ and $q_i=\bar{q}_e$. Define $d_p^C := \operatorname{argmin}\{\mathbb{E}[C(d_p)]\}$.

a) $d_p^C = 1$ *for* $v_j = v_{-j}, j\in\{i,e\}$,

b) $d_p^C = \overline{d}_p^j$ *for* $v_j > v_{-j} = 0, j\in\{i,e\}$.

Our prior results showed that bids from suppliers with equal values are minimal at $d_p = 1$. As any deviation from $d_p = 1$ yields higher bids from each supplier, the buyer should postpone the full procurement volume and single-source, to minimize expected procurement costs. From Proposition 7 we know that, for $v_j > v_{-j} = 0$, supplier j's bid is minimal at $\overline{d}_p^j$. Because supplier $-j$'s bid is constant in d_p, the buyer minimizes expected procurement costs by postponing $d_p = \overline{d}_p^j$. The results, stated in Proposition 8, confirm this intuition. While we can not derive expressions of the terminal bids for the case $v_j \neq v_{-j} > 0$, and therefore can not determine d_p^Q analytically for this general case, further numerical analyses indicate that the results of Proposition 8 continue to hold when the suppliers' value is positive but differs ($v_j \neq v_{-j} > 0$) and for different probability distributions (see section 3.3.3 where we show that our findings also hold in the general case).

In summary, our analysis shows that a buyer should set the postponed volume carefully, as a competition interval exists for each supplier and only for postponed volumes within this interval can a buyer expect the suppliers to compete. If the buyer chooses a postponed volume that is too low, the suppliers have no incentive to place bids below their reservation prices, and the buyer will not benefit from competition. Likewise, if the buyer chooses a postponed volume that is higher than the upper bound of the competition interval, the supplier with the higher value will take into consideration that the buyer can learn about the entrant's quality, giving the higher-value supplier an incentive to increase his bids. Only in the case of homogeneous suppliers the buyer should opt for single-sourcing and choose a postponed volume of $d_p = 1$ in order to maximize competition. Our results show that the suppliers' values determine the size and position of the competition interval: If suppliers are largely homogeneous in terms of their values, a buyer will see strong price drops in a fairly large interval of postponed volumes, and the interval in which prices increase will be comparatively small; however, if suppliers are heterogeneous in terms of their values, the competition interval will be comparatively small. Therefore, the buyer must set the postponed volume to the right level or risk the benefits it could gain from the competition between the incumbent and the entrant.

In the pharmaceutical industry, generics suppliers (entrants) can often produce drugs with a similar (expected) quality at a much lower cost than incumbents can, so heterogeneous supplier value is not the exception but a common phenomenon in this particular domain. Therefore, the buyer must understand the competitive dynamics, the competition interval, and how to set the postponed volume to benefit from competition between the suppliers. The supplier's value is directly determined by his costs, which we assume to be known to the buyer. We made this assumption in order to obtain analytically tractable results for the terminal bids. However, it is more realistic to assume that the buyer may lack precise information about the suppliers' costs. We explore the impact of uncertain costs in Appendix 7.1.2 and find that this uncertainty does not change the structure of our results.

Our results provide useful insights for both management and theory. The results presented in Proposition 7 suggest that, for some postponed volumes, a supplier with positive value quotes bids below the reservation price, even if his competitor does not lower his bid to increase the probability of winning the postponed volume. Thus, the postponed volume provides a competitive lever for a buyer even if one supplier is not a competitive threat. Competition is often seen only as the pressure that emerges from another supplier's lowering his price, but with a postponement tender a supplier may lower his bid even if the competitor does not, as the mere threat of losing the postponed volume after the buyer learns about the entrant's quality introduces competitive pressure that persists even if one supplier has no room to lower his bid. As this effect is driven by the buyer's learning, it is affected by our simplifying assumption that the initial volume is split symmetrically. In Appendix 7.1.3 we discuss how this assumption impacts our result and show that relaxing this assumption does not change our results structurally. However, we do observe that an asymmetric initial volume split provides the buyer with an additional lever to impact learning and competition. This finding is structurally related to research on buyer-determined or non-binding auctions, where suppliers face the uncertainty that the buyer will set and reveal the evaluation criteria only after bidding has concluded. For example, Engelbrecht-Wiggans and Katok (2007) explain that, in many mechanisms that are employed in practice, the buyer does not commit to awarding the contract to the lowest bidder but reserves the right to evaluate other criteria after bidding has concluded. The uncertainty about the buyer's evaluation and final decision may incent suppliers to quote lower bids because they want to increase their chances of winning the auction. However, a buyer not clarifying the evaluation criteria can be seen as non-transparent and may even foster misconduct (Fugger et al., 2015). In contrast to traditional buyer-determined auctions, in our mechanism the evaluation criteria are set before the suppliers enter the negotiations, so the uncertainty stems only from their not knowing whether they will receive the postponed volume. Therefore, even without being vague about how suppliers are evaluated and volumes are awarded, the postponement tender equips the buyer with a lever to induce competitive pressure, even if one supplier is not very competitive.

3.3.3 The buyer's optimal strategy

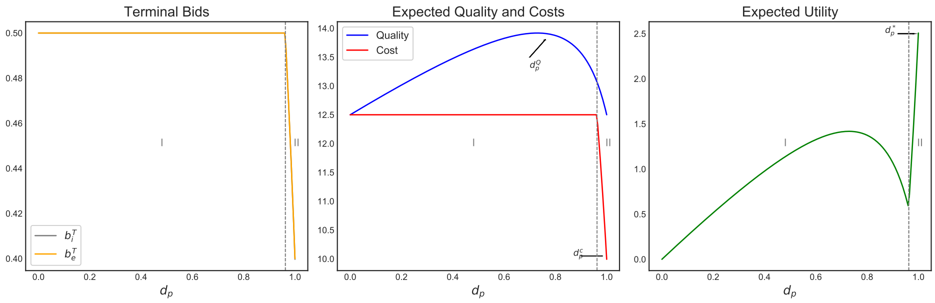
Sections 3.3.1 and 3.3.2 addressed how the choice of d_p impacts the buyer's expected quality and procurement costs and explained two characteristic values of d_p: the quality maximizing d_p^Q and the cost minimizing d_p^C. This section explains how the buyer should choose the postponed volume d_p to maximize her expected utility, as stated in Eq. (3.1).

While we can write separate expressions for the expected quality and the expected costs depending on the postponed volume (see Eq. (3.1)), it is clear that both are interrelated in a non-trivial way because the bids depend on the (expected) quality of both suppliers, and the choice of the supplier depends on both the bids and the (expected) quality. For this reason, it is difficult to determine the optimal postponed quantity d_p^* that maximizes the buyer's expected utility. We are, however, able to derive the optimal postponed volume for the case of homogeneous suppliers.

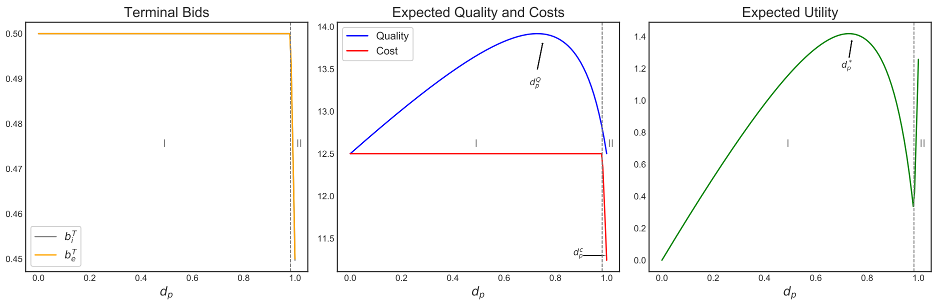
Proposition 9.
Let $Q_e \sim Beta(1,1)$ and $v_j = v_{-j} \geq 0$. Define $d_p^* := \text{argmax}\{\mathbb{E}[U(d_p)]\}$.

$$d_p^* = \begin{cases} d_p^Q < \underline{d}_p & \text{for } v_j \leq \overline{v}(M) = \frac{1}{2} - \frac{-8+3M+4\sqrt{4+M}}{8M} \\ d_p^C = 1 & \text{else} \end{cases} . \tag{3.12}$$

The result presented in Proposition 9 suggests that in the case of homogeneous suppliers it is optimal for the buyer to choose the postponed volume d_p^Q that maximizes her expected quality or the volume d_p^C that minimizes her expected procurement costs. Therefore, when $v_j = v_{-j}$ the buyer should either pursue a "learning strategy" or a "competition strategy" when setting d_p. From Proposition 8 we know that when the buyer faces homogeneous suppliers ($v_j = v_{-j}$), single-sourcing (i.e., $d_p^C = 1$) minimizes



(a) $v_i = v_e = 0.1$ and $d_p^* = d_p^C$.



(b) $v_i = v_e = 0.05$ and $d_p^* = d_p^Q$.

Figure 3.6: Suppliers' bids, expected quality, expected cost and expected utility for $M = 25$ and $Q_e \sim Beta(1, 1)$.

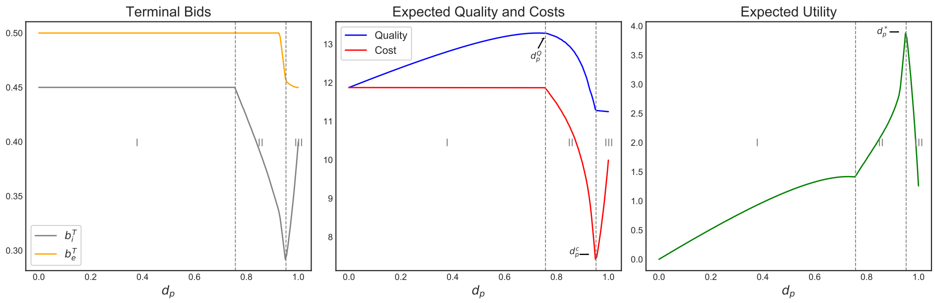
the expected procurement costs. As stated in Proposition 9, it is optimal for the buyer to single-source if the suppliers' values v_j exceed the threshold $\overline{v}(M)$, and to choose the quality-maximizing postponed volume d_p^Q if the suppliers' values are equal or below this threshold. The threshold is determined by the procurement volume M, so whether the buyer should opt for a learning or a competition strategy depends on a trade-off between suppliers' values and the purchasing volume. As Proposition 6 shows, higher supplier values lead to stronger competition and larger differences between the terminal bids and the reservation price, making the competition strategy more attractive. On the other hand, as we saw in Section 3.3.1, larger procurement volumes M increase the buyer's benefits from learning, making it more attractive to opt for a learning strategy and to dual-source and postpone some part of the procurement volume. However, this trade-off is not as straightforward as it may seem, as the procurement volume M not only has a direct impact on the buyer's learning, but also affects competition (Proposition 6). The threshold $\overline{v}(M)$ represents a tight bound on the suppliers' values that increases in M at a decreasing rate and converges to 0.125 for $M \to \infty$. Thus, our results suggest that, if the suppliers' values are not very low, a buyer should single-source to reap the benefits of competition, and should only opt for the postponement tender when suppliers' values are very low.

It is interesting to observe that—at least in the case of homogeneous suppliers—a buyer should always choose between the two local optima d_p^Q and d_p^C when trading-off the benefits of learning and competition, and pursue either a learning or a competition strategy. That is, a mixed learning and competition strategy is never optimal under these conditions. This dichotomy is a direct result of the interaction between learning and competition effects. To illustrate these effects, Figure 3.6 plots the suppliers' bids, expected quality and costs, and the expected utility, for homogeneous suppliers with values above (a) and below (b) the threshold $\overline{v}(M)$. In interval I ($d_p < \overline{d}_p$) there is no competition and the expected utility is only driven by the expected quality as a result of the learning effect, leading to a local maximum of the expected utility at d_p^Q. In interval II ($d_p > \overline{d}_p$) competition takes place and suppliers lower their bids. In Section 3.2 we saw that the suppliers' bids are driven by a volume incentive and a risk incentive and that in this interval both incentives lead to expected costs that are monotonically decreasing in d_p (see our discussion in conjunction with Proposition 6). Interestingly, as the results in Proposition 9 suggest, the cost decrease in interval II always offsets the decrease in expected quality—that is, the (negative) slope of the expected cost curve is larger than the (negative) slope of the expected quality curve and, as a consequence, the expected utility is strictly increasing in Interval II (Eq. (7.80) in the proof of Proposition 8 formalizes this property). The underlying reason for this result is that, taken together, the two effects—namely the volume incentive (a larger postponed volume is at stake) and the risk incentive (a larger postponed volume reduces the buyer's learning and allows the suppliers to have more impact on the outcome with their bids)—always have a stronger impact on expected costs than the reduced learning has on expected quality, which leads to a strictly increasing expected utility in Interval II. Therefore, the buyer's expected utility has a second local optimum at $d_p^C = 1$ and it is optimal for the buyer to choose either d_p^Q and pursue a learning strategy (for $v_j < \overline{v}(M)$), or to choose d_p^C and pursue a competition strategy (for $v_j > \overline{v}(M)$). At supplier values $v_j = \overline{v}(M)$ both local optima lead to the same expected utility for the buyer.

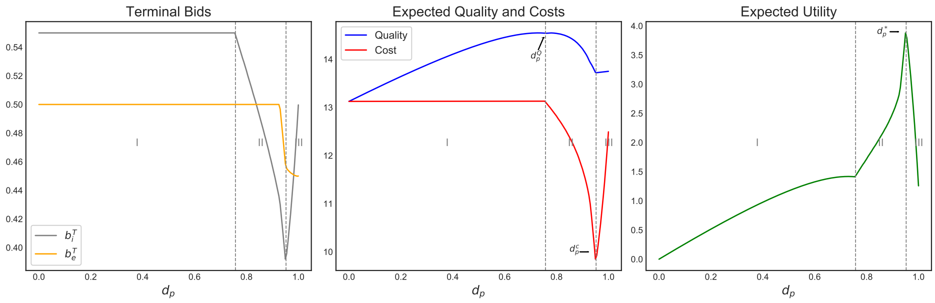
Of course, we want to find out whether this surprising result also extends to the case of heterogeneous suppliers ($v_j \neq v_{-j}$). However, as we discussed in Section 3.2, we cannot derive analytical expressions for d_p^C when $v_j \neq v_{-j} > 0$ and we can therefore not provide an analytical expression for the optimal postponement volume d_p^*. We can, however, evaluate Eq. (3.1) numerically and derive the optimal postponement volume for the general case $v_j \neq v_{-j}$. We carried out an extensive numerical study in which we determined d_p^Q, d_p^C and d_p^* for different supplier values v_i and v_e and procurement volumes M for various quality distributions $Q_e \sim Beta(a,b)$ with $a = b \geq 1$. More specifically, we varied v_i and v_e between 0 and 1 (in steps of 0.05) and M between 1 and 1000 (in increasing steps). A description of our evaluation procedure is provided in Appendix III.

The results of our numerical study suggest that the main results of Proposition 9 also hold for any combination of $v_j \neq v_{-j}$ and M: it is always optimal to either choose d_p^Q and to pursue a learning strategy, or to choose d_p^C and to pursue a competition strategy, and the optimal postponed volume d_p^* depends on both the suppliers' values v_i and v_e and the procurement volume M; it is never optimal to choose a mixed strategy with $d_p^Q < d_p < d_p^C$ or $d_p^C < d_p < d_p^Q$. We draw on our discussion of the results for homogeneous suppliers in Figure 3.6 to explain why this surprising result also holds true for heterogeneous suppliers. Figure 3.7 plots the suppliers' bids, the expected quality and costs and the buyer's expected utility for suppliers with unequal values. Figure 3.7a considers a case where the incumbent's quality is lower than the entrant's expected quality and Figure 3.7b a case where the entrant's expected quality is higher than the quality of the incumbent. In both instances the competition strategy is optimal, i.e. $d_p^* = d_p^C = \overline{d}_p$.

In Interval I ($d_p < \underline{d}_p^i$), we observe the same effects as in the case of homogeneous suppliers shown in Figure 3.6: the bids and the expected costs are constant because there is no competition and the expected quality and the expected utility is concave and increasing in d_p with a local optimum at d_p^Q. In Interval II ($\underline{d}_p^i \leq d_p \leq \overline{d}_p^e$), a buyer experiences decreasing expected quality as the incumbent is more likely to win the postponed volume and learning benefits decrease. The expected costs decrease because the incumbent's bid, and (for higher values of d_p) also the entrant's bid are decreasing in the postponed volume d_p. This behavior is again driven by the the volume and the risk incentive and we find the same structure as in the homogeneous supplier case: the two effects lead to strictly decreasing bids and therefore strictly decreasing expected costs that offset any negative changes in expected quality—hence, the buyer's expected utility is strictly increasing in Interval II. The main difference to the homogeneous case is that for heterogeneous suppliers it is not optimal to single source, i.e., to choose $d_p = 1$ (see Proposition 8), which is why we have a third interval that did not exist in Figure 3.6. In Interval III ($d_p > \overline{d}_p^i$), expected costs increase, because the incumbent's bid is increasing. Proposition 6 showed that in this interval the incumbent bids such that he is guaranteed to win the postponed volume. Because the incumbent is guaranteed to win, the buyer does not benefit from learning and, as d_p approaches a value of one,



(a) $q_i = 0.45 < \bar{q}_e$.

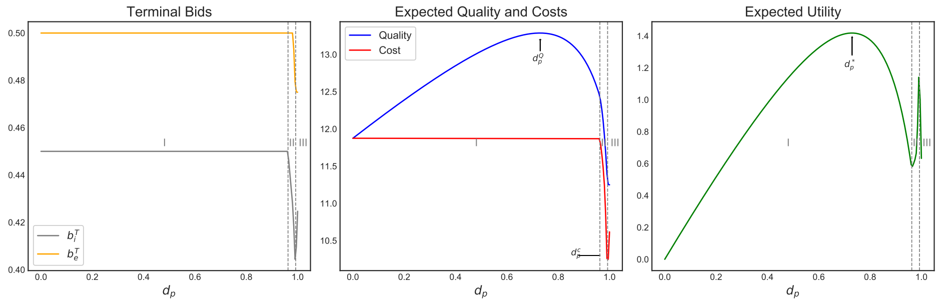


(b) $q_i = 0.55 > \bar{q}_e$.

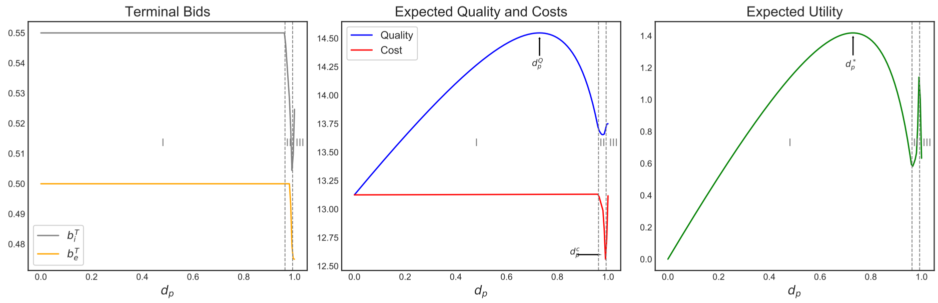
Figure 3.7: Suppliers' bids, expected quality, expected cost and expected utility for $v_i = 0.4$, $v_e = 0.05$, $M = 25$ and $Q_e \sim Beta(1,1)$.

the expected quality approaches the expected costs. At $d_p = 1$ the buyer does not learn (the initial volume is zero) and the incumbent only needs to decrease his bid infinitesimally below his reservation price (i.e., his quality) minus the entrant's value (i.e., his margin) to win the postponed volume. The expected quality is increasing when the incumbent's quality is higher than the entrant's expected quality, and decreasing when the incumbents quality is lower, because with increasing postponed volume more volume is allocated to the incumbent, as he is guaranteed to win. As a consequence, the buyer's expected utility is decreasing at postponed volumes $\overline{d}_p^i < d_p \leq 1$ and there is a second local optimum at $d_p^C = \overline{d}_p^i$.

The same holds true for instances in which the learning strategy is optimal, i.e. $d_p^* = d_p^Q$. In Figure 3.8 we illustrate such an instance similar to Figure 3.7: Figure 3.8a considers a case where the incumbent's quality is lower than the entrant's expected quality and Figure 3.8b a case where the entrant's expected quality is lower than the quality of the incumbent. As before, expected costs are constant in Interval I, because there is no competition and expected quality is concave due to the buyer's learning. Supplier values are lower than in Figure 3.7 and competition starts at a higher postponed



(a) $q_i = 0.45 < \bar{q}_e$.



(b) $q_i = 0.55 > \bar{q}_e$.

Figure 3.8: Suppliers' bids, expected quality, expected cost and expected utility for $v_i = 0.1$, $v_e = 0.025$, $M = 25$ and $Q_e \sim Beta(1,1)$.

volume, resulting in a local optimum of expected utility at d_p^Q in Interval I. In Interval II we can again see that expected costs are strictly decreasing and, as before, the decrease in the expected costs is stronger than the decrease in the expected quality—the buyer's expected utility is therefore strictly increasing in this interval. In Interval III the incumbent again increases his bid, which leads to increasing expected costs and decreasing expected utility. As a result we again observe a second local optimum of the expected utility at $d_p^C = \overline{d}_p^e$, just as in Figure 3.7. In this instance, however, the effect of competition is lower because of the low supplier values, which is why the buyer's expected utility is maximized at d_p^Q and not d_p^C, and the buyer should pursue a learning strategy. Summarizing our discussion of the six different cases depicted in Figures 7, 8 and 9, we find that the buyer's expected utility always has two local optima (d_p^Q and d_p^C) and that this shape of the expected utility function is induced by a) constant bids and concave expected quality in interval I, where the first local optimum (d_p^Q) occurs, b) a strictly increasing expected utility in Interval II where both the incumbent and the entrant compete strongly for the postponed volume and, c) decreasing expected utility in interval III (for $d_p > d_p^C$ in case of heterogeneous suppliers) where the incumbent is guaranteed to win the postponed volume. Although we can only prove this property of the expected utility function for the case of homogeneous suppliers, our numerical analyses show that this property also holds for heterogeneous suppliers, independent of the suppliers' values and the procurement volume, and also holds for various relevant quality distributions.

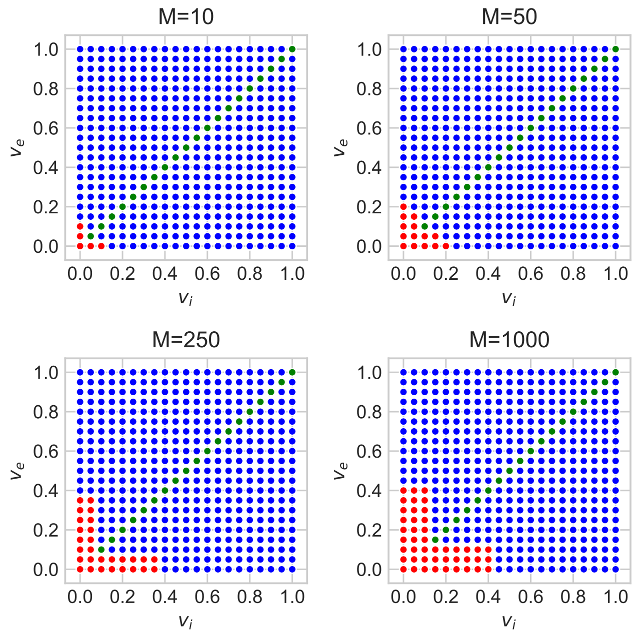


Figure 3.9: Optimal solutions (red: $d_p^* = d_p^Q$, blue: $d_p^* = d_p^C$, green: $d_p^* = d_p^C = 1$) dependent on supplier values v_i and v_e for different procurement volumes M.

Until now we explained *why* the results of Proposition 9 extend to the case of heterogeneous suppliers, and provided more insights into how expected quality and expected costs are intertwined and why there is no potential optimum between d_p^Q and d_p^C. Now we focus on explaining *when* it is optimal to choose a learning or a competition strategy. Figure 3.9 shows the optimal strategy depending on v_i and v_e for low, medium, and high values of the procurement volume M and also highlights when the competition strategy coincides with single sourcing ($d_p^C = 1$). The results are in line with those of Proposition 8 for homogeneous suppliers and we observe a similar interplay between the procurement volume M and the supplier values v_i and v_e: As in the case of homogeneous suppliers, the procurement volume appears to establish an upper bound on the supplier values below which it is optimal to pursue a learning strategy. At very low volumes (e.g., $M = 10$) it is only optimal to pursue a learning strategy when both suppliers have very low (but unequal) values. As M increases, a learning strategy is optimal even under higher and unequal supplier values. The results are rather intuitive and suggest that a learning strategy is attractive if suppliers are heterogeneous and exhibit low to moderate values and, at the same time, the procurement volumes are not very low. The buyer benefits less from learning when the suppliers are very similar in terms of their value or when the procurement volume is low (see our discussion in conjunction with Proposition 9). Our numerical results also show that single-sourcing is only optimal when suppliers are homogeneous and that a buyer facing heterogeneous suppliers should not single-source, even if it is optimal for her to pursue a competition strategy. These numerical results are in line with and extend the formal results established in Proposition 8. Proposition 8 showed that single-sourcing ($d_p = d_p^C = 1$) minimizes the expected costs when suppliers are homogeneous, and that the postponement tender (with $d_p^C < 1$) minimizes the expected costs for a special case of heterogeneous suppliers ($v_j > v_{-j} = 0$). As illustrated in Figure 3.9, our numerical results indicate that this is generally true. Whenever suppliers are heterogeneous and it is optimal to choose a competition strategy, the buyer should not single-source, but should employ the postponement tender to induce additional competition. This as an interesting and counter-intuitive finding because, contrary to managerial wisdom, we see that single-sourcing can inhibit competition compared to dual sourcing—at least if dual sourcing comes in the form of a postponement tender.

3.4 Managerial Implications

The main objective of our study is to determine under which conditions it is optimal for a buyer to employ a postponement tender instead of single-sourcing and how much to postpone. In the previous sections we carried out a formal analysis to determine how learning and competition affect the buyer's outcome and to characterize—both formally and numerically—the buyer's optimal postponement tender. This section addresses how the results of our analyses can be used to help decision-makers in practice.

The discussion of the results presented in Proposition 9 and the numerical results in Section 3.3.3 highlighted that the buyer should choose either a learning or a competition strategy, and that the optimal choice depends on the procurement volume M and the values of the suppliers v_j, $j \in \{i, e\}$. In many cases, the procurement volume M is closely related to a particular type of buyer and its size. Section 3.3 explained that M can be interpreted as the number of learning opportunities, such as the number of orders or units a buyer procures. Thus, we can generally distinguish two types of buyers: Institutional buyers like the Global Fund and UNICEF's supply division that consolidate the demand of many recipient countries, and smaller procurers like governmental buyers of low-income countries with smaller populations, and social marketing and relief organizations that operate decentralized procurement departments in certain countries or regions (e.g., Marie Stopes International). Institutional buyers usually have ample opportunities to learn because of their large procurement volumes and high numbers of orders. For example, USAID and UNFPA procure Depot-medroxyprogesterone acetate and place 30-90 orders per year (RHInterchange, 2016)). The smaller buyers typically have fewer learning opportunities because they have smaller procurement volumes and place fewer orders. Therefore, we assume that M is usually indicative of the buyer's size.

Global-health buyers encounter different supply market structures that are associated with specific (expected) supplier values v_j. Typically, the incumbent is a multi-national, branded manufacturer that supplies comparatively high quality at a comparatively high price. If the entrant is also a branded manufacturer, both suppliers will likely have similar expected quality and cost structures and will provide similar value to the buyer (i.e., $v_e = v_i$). We refer to this case as *homogeneous suppliers.* However, the entrant can also be a generics manufacturer that is focused on efficient operations and low overhead cost (for marketing and R&D), so its costs are often considerably lower than those of a branded manufacturer. Whether being a generics manufacturer makes the entrant have lower or higher value depends on the expected quality relative to the incumbent's quality. In general, two situations can occur: The buyer may expect the entrant to provide a similar (or even higher) quality as that of the incumbent, in which case the entrant's expected value is likely to be higher than that of the incumbent. There may also be cases in which the buyer expects the entrant's quality to be lower than that of the incumbent, such that $v_i > v_e$. We refer to the case of a higher or lower value as the *heterogeneous suppliers* case. Summarizing this discussion, we can broadly identify four scenarios with respect to the size of the buyer (high or low M) and the supply market structure (homogeneous or heterogeneous v_j, $j \in \{i, e\}$).

In the following sections we develop insights into the buyer's optimal postponement tender. These insights augment the results we derived in Section 3.3.3 by providing a more nuanced interpretation of the optimal postponement tender and deriving relevant managerial implications and recommendations. Finally, we discuss the managerial implications for these settings.

3.4.1 Optimal postponement for combinations of M and v_j

Figure 3.10 plots a buyer's optimal postponed volume d_p^* for three values of M depending on the expected entrant's value v_e and a fixed incumbent's value v_i and uniform entrant quality ($Q_e \sim Beta(1,1)$). Based on the results shown in Figure 3.10, we can identify the strategies developed in Section 3.3.3. When a small buyer faces heterogeneous suppliers with very low expected values ($v_e < v_i$), this buyer should choose a learning strategy. (We showed in Proposition 4 that d_p^Q is constant in v_j, which is reflected by the results presented in Figure 3.10.) The benefits of learning outweigh the benefits of competition, and the buyer should choose $d_p^* = d_p^Q$ if the entrant has a (net) disadvantage in terms of cost and quality. This disadvantage can occur when the supplier is either a branded manufacturer with similar quality but higher manufacturing costs or a generics manufacturer with comparatively low expected quality. In our example, it is optimal for the buyer to postpone approximately 65 percent of the volume.

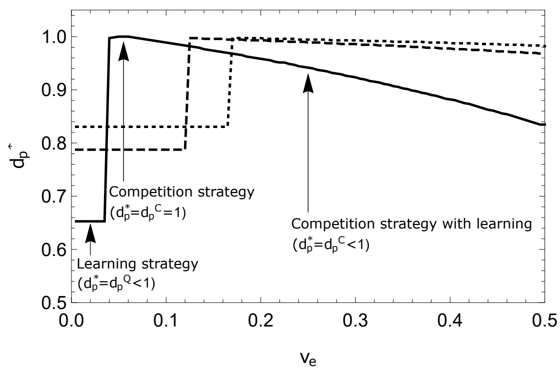


Figure 3.10: Optimal postponement d_p^* dependent on v_e for $v_i = 0.05$ and $Q_e \sim Beta(1,1)$ for procurement volumes $M = 10$ (solid), $M = 50$ (dashed) and $M = 90$ (dotted).

When suppliers are homogeneous (i.e., they have very similar expected values $v_e \approx v_i$), it is best for the small buyer to rely on a competition strategy with single-sourcing ($d_p^* = d_p^C = 1$). In this instance, the small buyer should not use a postponement tender. This is in line with the results of Proposition 9.

If suppliers are heterogeneous and the entrant has a higher expected value than the incumbent does ($v_e >> v_i$), small buyers should still rely on a competition strategy. However, if feasible, they should postpone most of the procurement volume and award only a small initial volume ($d_p^* = d_p^C < 1$). Here we observe the impact of learning on competition that emerges from the results of our discussion in Section 3.3.

For example, if the entrant is a generics manufacturer that is expected to have comparatively low production costs and quality similar to that of the incumbent, the buyer should exploit the benefits of competition. However, this conclusion does not suggest that single-sourcing is optimal; because of the effect that learning has on competition, the buyer benefits more from competition if the buyer initially awards a relatively small share of the volume and postpones the remaining quantity. Therefore, it is optimal for the buyer to opt for dual sourcing with a small initial award and a large postponed volume. In practice, it may not always be feasible to award small initial volumes. Appendix 7.1.4 discusses how a minimum initial volume impacts the buyer's optimal strategy.

The results are structurally the same for larger buyers, although their large procurement volumes make it more attractive to pursue a learning strategy even when the entrant promises (moderately) higher value than the incumbent does, that is, when suppliers are heterogeneous. Compared to smaller buyers' learning strategy, the fraction of the postponed volume should be higher for large buyers ($\geq 80\%$ in our example), a result that is in line with the results presented in Proposition 4.

If the entrant has substantially higher value than the incumbent, the buyer should pursue a competition strategy with a small initial volume and high postponed volumes, as shown in Figure 3.10 (e.g., for $M = 90$ and d_p^* just below 1). Once again, the reason for this outcome is the effect of learning on competition, which results in high postponed volumes when the competition strategy outperforms the learning strategy. (See our discussion in conjunction with Proposition 7.)

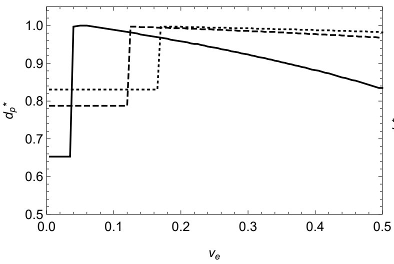
The high postponed volumes prompt a different managerial interpretation: If the buyer sets a high postponed volume, the initial volume must be small, so the postponement tender can be seen as a single-sourcing strategy with test lots from the suppliers. After evaluating the test lots, the buyer awards almost the entire procurement volume to the supplier with the higher value. Requiring test lots is common practice in pharmaceutical procurement when a buyer has little information about the quality of a new supplier. However, in the context of our mechanism, the proposed strategy differs from what is usually observed in practice, where buyers require test lots during the supplier-vetting process, that is, before suppliers compete (e.g. (USAID, 2016, p.5)). In the postponement tender, the suppliers have to quote a price before they provide the test lots, when they do not know how the buyer will evaluate their quality in comparison to their competitor's quality. This situation leads to additional competition because suppliers do not know whether they will be awarded the majority of the volume. In our mechanism, then, the buyer has an additional benefit when test lots are provided after the bidding process is completed. Of course, such a strategy is not feasible in all practical settings, as the buyer may be forced to award minimum initial volumes when opting for a postponement tender. We analyzed the impact of minimum initial volumes in a section in the the appendix (see Appendix 7.1.4). We show how these situations affect our structural insights and shape our managerial implications.

3.4.2 The impact of prior information

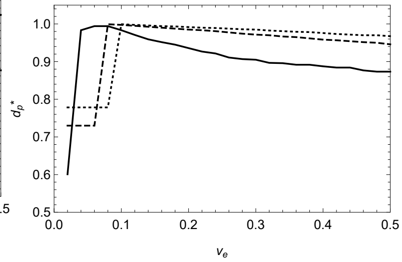
Until now we have assumed that entrants' quality is uniformly distributed ($Q_e \sim Beta(1,1)$). A uniform distribution represents a situation in which the buyer has very little information about the entrant's quality and all possible realizations are estimated to be equally likely. However, in reality, a buyer's a-priori information about suppliers' quality can vary substantially. When a buyer faces an entrant with which she has no experience and which is new to the market, the buyer has little information about the supplier's quality, and the buyer's uncertainty about the entrant's quality is likely to be high. In this situation, the uniform quality distribution is appropriate to represent the buyer's prior information. On the other hand, a buyer may have access to accurate information about the supplier's quality if the buyer has had prior contracts with the supplier for other products. Some buyers in the global-health domain also demand that suppliers undergo a strict quality-assurance process (e.g., the WHO-Prequalification and/or accreditation by a stringent regulatory authority like the US Food and Drug Administration (UNFPA, 2015; USAID, 2016; Global Fund, 2017a)). Prequalification and/or accreditation by a regulatory authority provide a strong signal of quality. In the context of our model, these conditions suggest prior quality distributions with lower dispersions and lower standard deviations. To explore how the optimal strategy is impacted by the buyer's having more prior information, we use a $Beta(\alpha, \beta)$ distribution with parameters $\alpha, \beta > 1$.

Figure 3.11 shows the buyer's optimal postponed volume d_p^* for different levels of prior information (i.e., different quality distributions $Beta(\alpha, \beta)$). If prior information increases — that is, if the coefficient of variation of the entrant's quality distribution decreases — there are fewer instances in which the buyer should pursue a learning strategy because the benefit from learning the entrant's quality decreases. (Recall our discussion in conjunction with Figure 3.3.)

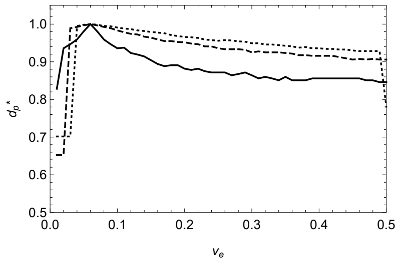
The results presented in Figures 3.11a, 3.11b, and 3.11c highlight this effect: for all types of buyers, it becomes more attractive to pursue a competition strategy than to pursue a learning strategy. If the amount of prior information increases (i.e., the CV decreases), a buyer who faces heterogeneous suppliers should postpone less volume and allocate higher initial volumes, so it becomes less attractive to require test lots. The reason for this effect lies in the impact of learning on competition, as the buyer needs



(a) $Q_e \sim Beta(1,1)$ and $CV = \frac{1}{\sqrt{3}}$.



(b) $Q_e \sim Beta(2,2)$ and $CV = \frac{1}{\sqrt{5}}$



(c) $Q_e \sim Beta(4,4)$ and $CV = \frac{1}{\sqrt{9}}$

Figure 3.11: Optimal d_p^* depending on v_e for $v_i = 0.05$ and $M = 10$ (solid), $M = 50$ (dashed), and $M = 90$ (dotted).

higher initial volumes to learn and impact competition. Hence, decreasing uncertainty decreases the value of a learning strategy, and all types of buyers should use a learning strategy if suppliers are expected to provide low value and little competition. However, in markets where suppliers are expected to be competitive, a buyer should still award initial volumes to use the effect of learning on competition, and should increase these initial volumes if uncertainty about quality decreases. Compared to our reference case with high prior uncertainty there are fewer cases in which buyers should postpone almost all volume to maximize expected utility.

Having analyzed the dimensions of the problem and what affects the optimal postponement decision, we can populate the market structure matrix (Figure 3.12). Our analyses show that the main strategy for small buyers and homogeneous suppliers should be a *competition strategy*. The same holds true for both small and large buyers and heterogeneous suppliers ($v_i < v_e$), where the optimal strategy is always a *competition strategy with learning*. For large buyers that face homogeneous suppliers, the optimal decision depends primarily on the level of uncertainty about quality. If the buyer has little prior information about the entrant's quality, she should focus on learning and implement a *learning strategy*, where she awards a large initial volume to both suppliers.

However, if the buyer has some information about the entrant's quality, she should forgo learning and opt for a competition strategy to increase competitive pressure during the bidding stage and maximize expected utility.

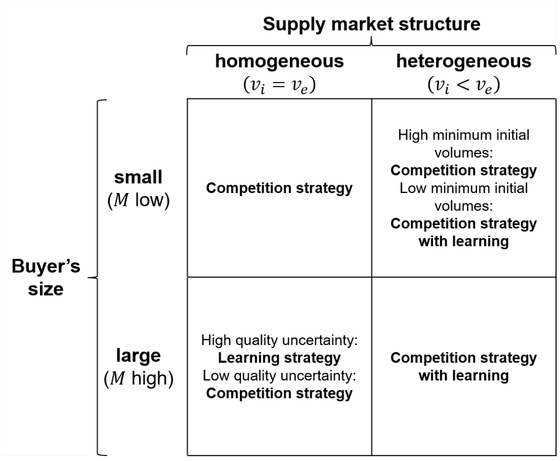


Figure 3.12: Market scenarios and optimal procurement strategies

3.5 Conclusion

This research addresses a new procurement mechanism, the *postponement tender*. With a postponement tender, the buyer allocates initial volumes to eligible suppliers, observes the delivered quality, updates her beliefs about the suppliers' quality, and then allocates the postponed volume to the supplier with the higher value. To induce competition, the buyer lets the suppliers bid before any volumes are allocated. This paper addresses two main questions: (1) Under what conditions is it optimal to prefer a postponement tender over single-sourcing? and (2) if it is optimal to employ a postponement tender, how much volume should be postponed? Based on our analytical and numerical analyses we find two surprising results: A postponement tender should always be preferred over a single-sourcing auction format when suppliers are heterogeneous in terms of their values, and the buyer should exclusively pursue either a "learning strategy" or a "competition strategy". The reason why in most cases a postponement tender should be preferred is that it induces an additional element of competition.

We are also able to identify the conditions under which it is optimal to pursue a learning or a competition strategy, and we find that the choice of the optimal strategy depends on the overall procurement volume and suppliers' values. We use our formal and numerical results and insights to derive managerial implications, especially for buyers in the global-health domain, and tie the postponement strategies to the types of buyers and the supply market structure.

To derive meaningful analytical results, we focused on certain elements of the problem, but ignored a number of incentive effects of the postponement tender. As highlighted in the introduction, the postponement tender not only strikes a balance between costs and uncertain quality, but may also incentivize new suppliers to enter the tendering process and to increase competitive pressure. This is an additional benefit of the postponement tender that we did not account for in our analysis. Moreover, suppliers may have an incentive to provide better quality in the initial volume to increase their chances of winning the postponed volume. Intuitively, larger postponed volumes should increase a supplier's incentive to invest into better quality control in the initial volume because the upside potential increases. However, a large postponed volume may lower suppliers' incentives to develop an appropriate production capacity for use with a smaller initial volume. We perceive a more extensive analysis of the incentive effects of a postponement tender to be an important and promising avenue for future research. Another aspect is our assumption of quality uncertainty. We have made the assumption that the buyer and both the incumbent and the entrant all share the same prior belief about the entrant's quality. However, it is reasonable to assume that the entrant might have a better prior for quality than the buyer and the incumbent. This information asymmetry could yield additional interesting implications for competition and could ultimately influence the buyers optimal postponement decision. Furthermore, we chose a descending auction to model the competition between the suppliers, which is likely very close to what we observe in practice. However, the choice of the auction format could actually be an additional lever for the buyer to influence competition, and a different format could ultimately provide better results for the buyer in terms of utility. Hence, it would be interesting to explore how different competition formats impact competition and the buyer's optimal choice. All of these avenues and their implications on the incentives and their effect on the postponement tender can potentially be addressed by future research.

Chapter 4

A Data-Driven Inventory Policy for Multi-Period, Multi-Product Inventory Planning Problems

This paper proposes a novel approach to dealing with multi-period and multi-product inventory problems, also known as joint replenishment problems, using sample average approximation. The approach is based on renewal theory and is motivated by the work of Çetinkaya and Lee (2000). Using real-world sales data from Kenyan pharmacies, the study shows that this approach can result in low inventory costs.

4.1 Introduction

Multi-period inventory problems have attracted a steady stream of research interest because it is difficult to find optimal inventory policies under generalized assumptions. In addition, in many inventory settings, multiple products must be managed simultaneously: they require joint optimization so they can share fixed ordering (or set-up) costs and quantity discounts offered by suppliers. This type of problem is also referred to as the Joint Replenishment Problem (JRP).

This paper uses the business setting of Maisha Meds[1], a Kenyan point-of-sale application provider for pharmacies and clinics, which provided historical sales data for this study. Imagine the inventory-management problem of a privately owned pharmacy: multiple products with random, non-stationary demand that can be correlated in time

[1] https://maishameds.org/

and between products must be ordered recurrently, with the potential to pool orders to reduce fixed costs and use suppliers' quantity discounts (based on order volume or invoice value). Using generalized assumptions would make it difficult to find the optimal replenishment policy, but there is a wide array of approaches to finding near-optimal policies. An approach that has gained attention is the (s, c, S) policy, a can-order policy introduced by Balintfy (1964) that features can- and must-order levels for each product. However, this type of policy class "is not the optimal coordinated replenishment policy class" (Liu and Yuan, 2000, p. 491). To make the problem manageable, most research makes simplifying assumptions, such as deterministic demand (Goyal (1974), Wildeman et al. (1997), Viswanathan (2002), Cha and Moon (2005), Zhang (2009)) or Poisson distributed demand (Atkins and Iyogun (1988), Viswanathan (1997), Çetinkaya and Lee (2000)). However, the true demand distribution is often unknown, and it is difficult to estimate demand distributions, let alone demand correlations (auto-correlation, as well as correlations between products), adding a new layer of difficulty to the problem. Therefore, data-driven inventory policies, such as the data-driven Newsvendor (Bertsimas and Thiele, 2005, 2006), have been used frequently because they do not require any particular type of demand distribution but find optimal decisions directly using the available data.

This paper proposes a novel data-driven inventory-management approach to the JRP using Sample Average Approximation (SAA) and real-world data. A first step compares the approach for the single-product setting without quantity discounts to the optimal dynamic-programming policy and the ex-post optimal policy. The dynamic-programming policy is theoretically optimal, so it serves as a first indicator of how well the proposed approach performs in a simple setting, while the ex-post optimal policy serves as a benchmark to make the results comparable in more complex settings. Then the approach is evaluated for the more complex multi-product setting with quantity discounts. As there is no theoretically optimal policy for this setting, the approach's performance is compared only to the ex-post optimal policy. Results show that the gap to optimality increases only slightly when one moves from the single-product setting to the multi-product setting with quantity discounts, despite the increased complexity of the planning problem. Combined with the finding that the proposed policy's performance is not far off the dynamic programming policy's performance in the single-product setting, this result suggests that the proposed approach works well in a multi-product setting with quantity discounts.

The rest of the paper is organized as follows. Section 4.2 presents a discussion of the related literature, while section 4.3 establishes the model for the single- and multi-product settings. In section 4.4 the proposed approach is analyzed using real-world data—first for the single-product setting to show that the approach is suitable for managing the underlying problem (Section 4.4.2) and then for the more complex multi-product setting (Section 4.4.3). Section 4.5 concludes and offers suggestions for future research on the evaluation, development, and application of the approach.

4.2 Relevant Literature

Multi-period inventory problems can be separated into problems with fixed costs and problems without fixed costs. While the optimal policy for the latter is usually an order-up-to policy, where in each period the inventory is replenished, such is not the case for the former, as the replenishment itself is costly, so the inventory should not necessarily be replenished each period. Scarf (1959) was the first to show that, in a single-product setting with a multi-period inventory with fixed order costs and independently distributed demand, an (S, s)- type policy is optimal, as it consists of an order-up-to point and a reorder level, and the inventory is replenished only if it falls below the reorder level. See Zipkin (2000) for a good overview of the classic single-product inventory problem. Research has recently started to focus on more data-driven approaches, where the inventory planner does not know the demand distribution but uses historical demand to find a solution. Ban (2019) introduced an SAA version of Scarf's inventory policy and showed that it is asymptotically optimal. I use Ban's policy later as a benchmark for the single-product setting.

The planning problem becomes significantly more difficult to solve when multiple products are managed simultaneously because scale economies in fixed-order costs and quantity discounts provide the potential for costs savings but also add complexity. The multi-product inventory problem is generally referred to as the JRP. Goyal (1974), Wildeman et al. (1997), Viswanathan (2002), Cha and Moon (2005) and Zhang (2009) studied the JRP under deterministic demand to find the optimal solution to the inventory problem using efficient algorithms and heuristics. Stochastic demand increases the difficulty of finding (optimal) inventory policies, as the deterministic problem itself is difficult to solve efficiently. Balintfy (1964) introduced a can-order policy, where each product has a must-order and a can-order level and is replenished at the can-order level only if another product is replenished at the same time. Atkins and Iyogun (1988) proposed a periodic replenishment policy, where products are replenished in fixed cycles, and showed that their approach can lower costs compared to the class of can-order policies. Viswanathan (1997) extended their suggested policy by adding individual and independent periodic review policies for each product and showed that doing so improves the results. Çetinkaya and Lee (2000) considered a similar problem in a vendor-managed inventory and proposed to model the inventory process as a stationary renewal process, where the expected time between two consecutive replenishments is a replenishment cycle. The present paper proposes an inventory policy that is based on their idea of minimizing the expected costs of the replenishment cycle but differs in considering only the next replenishment cycle, not a stationary one.

Although these policies are intuitive and perform well, they share the assumption that demand is stationary and Poisson-distributed, which is often not the case in practical settings, where demand is often not stationary and the true demand distribution is unknown. The current study contributes to the literature by proposing a novel approach that can deal with non-stationarity and does not have to assume a demand distribution, as it is a data-driven approach, where sample data can be used to calculate the optimal decisions for the inventory problem. Turgut et al. (2018) also considered a data-driven approach to optimize can-order inventory policies using mixed-integer linear programming, but they differed from the setting analyzed here in that they considered a different inventory process with a backroom effect and did not allow for quantity discounts.

This paper combines a renewal-theory view of the inventory process with SSA to estimate the expected inventory costs and find inventory decisions. Existing research that seeks to solve (joint) inventory management problems using sample-based approaches instead of parametric probability distributions has shown multiple advantages of sampling-based policies. Because it is often difficult to work with demand distributions, which may be either unknown or too complex, Levi et al. (2007, p. 821) contended that "a sampling-driven algorithmic framework is very attractive, both in practice and theory" and acknowledged that solving the SAA counterparts for stochastic multi-period problems is difficult, so they proposed a dynamic programming (DP) framework. Akçay and Xu (2004) considered a two-stage assemble-to-order inventory system and used a stochastic integer program to develop base-stock and component allocation policies. They also used an SAA approach to optimize their base-stock problem and showed that it outperforms related approaches. While they recognized, in general, their policy is not optimal for their problem, they found that it "has been adopted in analysis and practice due to its simple structure and easy implementation" (Akçay and Xu, 2004, p. 102).

The next section establishes modelling choices and derives a data-driven inventory policy for multi-period inventory problems in single- and multi-product settings. Section 4.4 provides a brief numerical evaluation of this novel inventory policy using a real-world data set, first for the single-product setting, for which an optimal benchmark exists

(Section 4.4.2), and then for the more complex multi-product setting with quantity discounts (Section 4.4.3).

4.3 Model

The general multi-period, multi-product inventory planning problem considered in this analysis requires definition. The term *multi-product* is used here synonymously with JRP, a more commonly used term. Define I_t^i as the inventory level of product $i \in \{1, \ldots, I\}$ in period $t \in \{1, \ldots, T\}$. For each unit held in period t, with $I_t^i > 0$, a holding cost of $h \geq 0$ per unit and period is incurred. If there are outstanding orders (i.e. $I_t^i < 0$), a back-order cost $b \geq 0$ per unit and period is incurred. At the beginning of a period t the inventory can be replenished with an order $Q_t^i \geq 0$. If an order is placed (i.e. $Q_t^i > 0$), fixed order cost $k(\mathbf{Q_t}) \geq 0$ and variable order costs $p_i(\mathbf{Q_t}) \geq 0$ are incurred, where $\mathbf{Q_t}$ is a vector representing the order quantities for all products $i \in \{1, \ldots, I\}$, and $p_i(Q)$ is the volume-dependent purchasing cost function. The purchase price of product i may be affected by the total amount of all products ordered because manufacturers may grant quantity discounts for larger orders. $k(Q)$ is the fixed cost function, which depends on all orders in one period, because there can be economies of scale if multiple orders are placed simultaneously. After the replenishment decision Q_t^i in period t has been made, random demand $D_t^i \sim F_t^i$ is realized, and the inventory level of the following period is calculated as $I_{t+1}^i = I_t^i + Q_t^i - D_t^i$. Define $I_0^i = 0$ as the starting inventory. The multi-period, multi-product inventory planning problem can then be formulated as

$$\min_{\mathbf{Q_t} \geq \mathbf{0}} \mathbb{E}_{D^i}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}(p(\mathbf{Q_t})Q_t^i + k(\mathbf{Q_t}) + (I_t^i)^+h + (-I_t^i)^+b)\right] \tag{4.1}$$

$$\text{s.t.} \quad I_0^i = 0, \quad i \in \{1, \ldots, I\}, \tag{4.2}$$

$$I_t^i = I_{t-1}^i + Q_t^i - D_t^i, \quad t \in \{1, \ldots, T\},\ i \in \{1, \ldots, I\}. \tag{4.3}$$

Determining the optimal orders Q_t^i for each product $i \in \{1, \ldots, I\}$ and period $t \in \{1, \ldots, T\}$ is obviously difficult. Even if the random variables D_t^i are assumed to be independently distributed in time and between products, the literature has not provided closed-form solutions for this inventory problem. If demand is correlated between periods and/or between products, the difficulty of this multi-period, multi-product inventory-planning problem increases sharply. In addition, the joint distribution function of demand is often unknown and is difficult to estimate.

To solve this problem, this study proposes a novel data-driven inventory policy that is motivated by the renewal theory approach of Çetinkaya and Lee (2000). This approach considers a (stationary) inventory process as a renewal process whose costs can be minimized by minimizing the expected average costs. A renewal cycle is the time between two consecutive inventory replenishments, and average costs are calculated as the cost of one replenishment cycle divided by the length of the cycle. Under general assumptions, the inventory process is not stationary, so the proposed approach estimates only the expected costs of the next replenishment cycle. Because the approach makes no assumption about how demand is distributed, future costs are estimated using historic demand samples and sample average approximation (SAA). Similar to Çetinkaya and Lee (2000), the assumption here is that the inventory uses a reorder-point, order-up-to policy. (See, e.g., (Scarf, 1959)). In the following the average replenishment cycle costs for the single product inventory is estimated, which is in itself a novel approach. Later, this formulation is extended to the multi-product setting.

Assume that problem (4.1) is reduced to the single-product setting (i.e. $I = 1$). The superscript i is kept to make the exposition more tractable. No assumptions are made about how the demand D^i_t in period t is distributed; instead, it is assumed that there are n demand samples $\Delta^{i,j}, j \in \{1, \ldots, n\}$,. Each demand sample consists of the historical demand of one selling season with T periods of demand, $\Delta^{i,j} = \{\delta^{i,j}_1, \ldots, \delta^{i,j}_T\}$. In period t the length of the next replenishment cycle $\tau^{i,j}_t$ for demand sample j is assumed to be the number of periods until the next replenishment is triggered, using a reorder point R^i_t. The starting inventory for this cycle, starting in period t, is assumed to be the order-up-to level $S^i_t \geq I^i_t$. If $S^i_t > I^i_t$, an order of size $Q^i_t = S^i_t - I^i_t$ is placed, and if $S^i_t = I^i_t$, no order is placed. The case $S^i_t < I^i_t$ does not make sense in this setting, as it is assumed that inventory can not be thrown away.

Using this structure results in the sample average costs of the replenishment cycle (starting in period t with inventory level I^i_t) for a reorder-point R^i_t and an order-up-to level $S^i_t \geq I^i_t$ as:

$$C^i_t(R^i_t, S^i_t) = \frac{1}{n} \sum_{j=1}^{n} \frac{c^{i,j}_t(R^i_t, S^i_t)}{\tau^{i,j}_t(R^i_t, S^i_t)}, \tag{4.4}$$

where $\tau^{i,j}_t(R^i_t, S^i_t)$ is the length of the replenishment cycle for the demand sample j, and $c^{i,j}_t(R^i_t, S^i_t)$ are the associated costs if the inventory is replenished up to level S^i_t in the current period t and a new order is expected to be placed as soon as the inventory reaches R^i_t. Both the length of the replenishment cycle and costs are functions of reorder level R^i_t and order-up-to level S^i_t. The length of the replenishment cycle for sample j

can be calculated as:

$$\tau_t^{i,j}(R_t^i, S_t^i) = \max\left[\inf\{u : \sum_{l=t}^{t+u} \delta_l^{i,j} \geq S_t^i - R_t^i\}, T\right] - t + 1, \tag{4.5}$$

which is the minimum number of periods until aggregate demand is equal to or larger than the difference between order-up-to point S_t^i and reorder point R_t^i, which is the period in which the next order would be triggered for the policy (R_t^i, S_t^i). Given the replenishment cycle length $\tau_t^{i,j}(R_t^i, S_t^i)$, it is now possible to calculate the replenishment cycle costs $c_t^{i,j}(R_t^i, S_t^i)$, which consist of multiple cost factors that will be established individually. First, holding costs $H_t^{i,j}(R_t^i, S_t^i)$ are:

$$H_t^{i,j}(R_t^i, S_t^i) = h \sum_{l=t}^{\tau_t^{i,j}(R_t^i, S_t^i)} (S_t^i - \sum_{m=t}^{l} \delta_m^{i,j})^+. \tag{4.6}$$

Similarly, backlogging costs $B_t^{i,j}(R_t^i, S_t^i)$ can be calculated as:

$$B_t^{i,j}(R_t^i, S_t^i) = b \sum_{l=t}^{\tau_t^{i,j}(R_t^i, S_t^i)} (\sum_{m=t}^{l} \delta_m^{i,j} - S_t^i)^+. \tag{4.7}$$

The purchasing costs depend on Q_t^i, the number of units ordered, and $p(Q)$, the purchasing cost function. As it is assumed that the inventory is replenished up to the level S_t^i if $S_t^i > I_t^i$, the order size $Q_t^{i,j}(S_t^i)$ depends only on the order-up-to level S_t^i and the inventory level I_t^i, so it is independent of the reorder point, and the demand sample:

$$Q_t^{i,j}(S_t^i) = \begin{cases} S_t^i - I_t^i, , & S_t^i \geq I_t^i, \\ 0, & S_t^i < I_t^i. \end{cases} \tag{4.8}$$

The purchasing costs $P_t^{i,j}(S_t^i)$ can then be calculated as:

$$P_t^{i,j}(S_t^i) = p(Q_t^{i,j}(S_t^i))Q_t^{i,j}(S_t^i). \tag{4.9}$$

Similarly, the fixed ordering costs are paid only if $S_t^i > I_t^i$, which yields the fixed ordering costs of:

$$K_t^{i,j}(S_t^i) = \begin{cases} k, & S_t^i > I_t^i, \\ 0, & S_t^i = I_t^i. \end{cases} \tag{4.10}$$

Therefore, the sum of the holding, backlogging, purchasing and fixed ordering costs yields the replenishment cycle costs in period t for sample j as:

$$c_t^{i,j}(R_t^i, S_t^i) = H_t^{i,j}(R_t^i, S_t^i) + B_t^{i,j}(R_t^i, S_t^i) + P_t^{i,j}(S_t^i) + K_t^{i,j}(S_t^i). \tag{4.11}$$

Now that all of the relevant costs have been established, the average cycle costs (4.4) can be calculated, and the minimization problem for the single-product problem in period $t \in \{1, \dots, T\}$ with inventory level I_t^i is:

$$\min_{(R_t^i, S_t^i)} C_t^i(R_t^i, S_t^i) = \frac{1}{n} \sum_{j=1}^{n} \frac{c_t^{i,j}(R_t^i, S_t^i)}{\tau_t^{i,j}(R_t^i, S_t^i)}, \tag{4.12}$$

$$\text{s.t.} \quad S_t^i \geq R_t^i, \tag{4.13}$$

$$S_t^i \geq I_t^i. \tag{4.14}$$

In each period t the minimization problem (4.12) is solved, and if the optimal order-up-to level S_t^i is larger than the current inventory level I_t^i, an order is triggered and the inventory is replenished up to S_t^i; otherwise, no order is placed. Before analyzing the performance of this multi-period, single-product inventory policy, we must first establish the extended policy for the multi-product setting. However, as section 4.4 shows, the single-product policy performs well.

In the multi-product setting, each product i is defined by its inventory level I_t^i in period t, an order-up-to point S_t^i and a reorder-point R_t^i. Similar to expression (4.4), we consider the sample average costs of the replenishment cycle starting in period t as

$$C_t(\mathbf{R_t}, \mathbf{S_t}) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{I} \frac{\mathbf{c}_t^{i,j}(R_t^i, \mathbf{S_t})}{\tau_t^{i,j}(R_t^i, S_t^i)} \tag{4.15}$$

where $\mathbf{R_t}$ and $\mathbf{S_t}$ are the vector representations of all inventories' reorder points and order-up-to levels for period t. The lengths of the individual cycles $\tau_t^{i,j}(R_t^i, S_t^i)$ are the same as in (4.5), while the cycle costs $\mathbf{c}_t^{i,j}(R_t^i, \mathbf{S_t})$ per product depend on all order-up-to levels.

Similar to the single-product case, each cost component is first established individually. Inventory holding costs and back-order costs are equivalent to those of the single-product model and can be calculated as:

$$\mathbf{H}_t^i(R_t^i, S_t^i) = h \sum_{l=t}^{\tau_t^j(R_t^i, S_t^i)} (S_t^i - \sum_{m=t}^{l} \delta^{i,j})^+, \tag{4.16}$$

and

$$\mathbf{B}_t^j(R_t^i, S_t^i) = b \sum_{l=t}^{\tau_t^j(R_t^i, S_t^i)} (\sum_{m=t}^{l} \delta^{i,j} - S_t^i)^+. \tag{4.17}$$

Purchasing costs now depend on the total volume ordered for all inventories that have order-up-to levels larger than their respective inventory levels. Hence, the total number of units ordered is:

$$\mathbf{Q}_t^j(\mathbf{S_t}) = \sum_{i=1}^{I} Q_t^{i,j}(S_t^i), \tag{4.18}$$

where the order quantity function from the single-product setting is used, resulting in purchasing costs of

$$\mathbf{P}_t^{i,j}(\mathbf{S_t}) = p(\mathbf{Q}_t^j(\mathbf{S_t}))Q_t^{i,j}(S_t^i) \tag{4.19}$$

for product i. Finally, fixed ordering costs depend on how these costs can be split among multiple products if more than one order is triggered. Therefore, assume that there is a function $\mathbf{K}_t^j(\mathbf{S_t})$ that defines the amount of fixed ordering costs incurred for each product. For example, if no scale economies are possible, $\mathbf{K}_t^j(\mathbf{S_t})$ would be equal to k if $S_t^i > I_t^i$ and equal to 0 for $S_t^i = I_t^i$. The sum of the holding, backlogging, purchasing and fixed ordering costs gives us the approximated replenishment cycle costs in period t as:

$$\mathbf{c}_t^{i,j}(R_t^i, \mathbf{S_t}) = \mathbf{H}_t^{i,j}(R_t^i, S_t^i) + \mathbf{B}_t^{i,j}(R_t^i, S_t^i) + \mathbf{P}_t^{i,j}(\mathbf{S_t}) + \mathbf{K}_t^{i,j}(\mathbf{S_t}). \tag{4.20}$$

Now that we have formulated the approximated cost function for individual inventories, we can formulate the minimization problem for the multi-product problem in period

$t \in \{1, \dots, T\}$ as

$$\min_{(\mathbf{R_t}, \mathbf{S_t})} C_t(\mathbf{R_t}, \mathbf{S_t}) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{I} \frac{\mathbf{c}_t^{i,j}(R_t^i, \mathbf{S_t})}{\tau_t^{i,j}(R_t^i, S_t^i)}, \tag{4.21}$$

$$\text{s.t.} \quad S_t^i \geq R_t^i, \quad i \in \{1, \dots, I\}, \tag{4.22}$$

$$S_t^i \geq I_t^i, \quad i \in \{1, \dots, I\}. \tag{4.23}$$

As in the single-product setting, in each period t the minimization problem (4.21) is solved, and the inventory of each product i is replenished if the cost-minimizing S_t^i is larger than the product's inventory I_t^i. Now that the inventory policies for both the single- and multi-product settings are established, the next section presents a brief analysis of the performance of these policies for the Maisha Meds data set. The costs of these policies are compared to the ex-post optimal costs under deterministic planning and, when possible, to other benchmark policies.

4.4 Analysis

In this section the performance of the proposed inventory policy, referred to in this article as the renewal theory (RT) policy, is evaluated for the single- and multi-product settings in a real-world application using a data set provided by Maisha Meds. Section 4.4.1 explains the structure of the data and describes how it is used for the numerical evaluation of the inventory policy. Section 4.4.2 evaluates the RT policy's performance for the single-product setting, and Section 4.4.3 extends the analysis to the multi-product setting. This procedure reveals the applicability of the RT policy in a simple setting for which there is a theoretically optimal benchmark, Ban (2019)'s data-driven DP policy, which is based on Scarf (1959)'s fundamental work, to which this article refers as the DP policy. However, there is no extension to this approach for the multi-product setting, and no theoretically optimal policy in general, so the ex-post optimal (EPO) policy, the inventory policy that minimizes costs if demand is known ex-ante (i.e., under deterministic planning) is used as a second benchmark. The EPO policy can also be evaluated for the multi-product setting, making possible a comparison of the RT policy and the EPO policy in the multi-product setting and in the single-product setting. If the difference between RT and EPO policy is stable between the single- and multi-product settings, and the RT policy's performance is close to that of the DP policy in the single-product setting, the RT policy is likely to be a suitable approach to dealing with these types of inventory problems.

4.4.1 Evaluation Procedure

The first benchmark is the data-driven DP approach that Ban (2019) proposed, a sample-average approximation version of the original dynamic programming inventory model that Scarf (1959) introduced. This approach in particular provides a perfect benchmark, as it uses the same input data and, under the assumption that demand in each period is independently distributed, is the theoretically optimal policy for the single-product setting. The second benchmark is the EPO policy, which is implemented using linear programming. (See the appendix in Chapter 4 for the formal model of the EPO policy and Ban (2019)'s paper for details of the DP policy.) All numerical evaluations are implemented and executed in R.

A data set that contains daily sales of multiple products from multiple pharmacies between January 2017 and October 2018 (22 months of daily sales) is used for the numerical evaluation. The planning period is assumed to be a thirty-one-day month (i.e., $T = 31$). The data set is split into a training set that consists of sixteen months and a test set of six months. Each month corresponds to one demand sample with T = 31 consecutive days of demand values. The inventory decisions are calculated using only samples from the training data, and inventory costs are evaluated for the test data. The evaluated data contain 100 pairs of products, so 200 instances are evaluated for the single-product setting (each product individually) and 100 instances are evaluated for the multi-product setting (each product pair individually). The average inventory cost for each setting and policy (RT, DP, and EPO) is then calculated as follows: For each product (or product pair in the multi-product setting), the inventory process and the corresponding costs are calculated based on the respective policy for each sample from the test data. A policy's average inventory cost is then calculated as the average inventory cost of all evaluated products (or product pairs) for all test data samples. Average costs of policy $x \in \{\mathrm{RT}, \mathrm{DP}, \mathrm{EPO}\}$ are defined as C_x.

4.4.2 Single-Product Setting

For the single-product setting, the inventory policies are evaluated for a number of parameter settings. Holding costs, back-order costs, and fixed order costs will be varied as $h \in \{0.05, 0.1, 0.15\}$, $b \in \{1, 1.5, 2\}$, and $k \in \{50, 75, 100\}$, respectively. The variable order costs will be fixed at $p = 0.5$ because the impact of p changes with the changes in the other parameters. For the single-product setting, variable order costs do not change with the order quantity because the DP policy is defined only for constant order costs. This assumption is relaxed for the multi-product setting. Table 4.1 shows all of the evaluated parameter constellations and their numerical results. The costs of the EPO policy are not sensitive to changes in b because the policy plans deterministically, so it can avoid back orders.

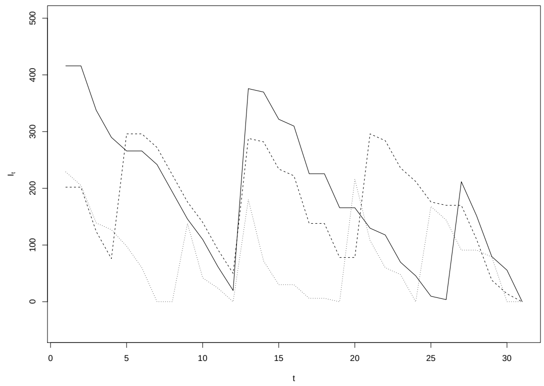


Figure 4.1: Inventory processes of RT policy (solid), DP policy (dashed) and the EPO policy (dotted) for a randomly selected product and sample from the data set.

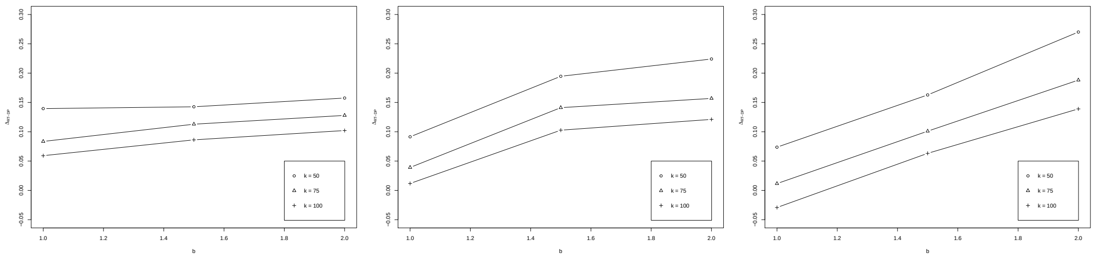
The RT policy is expected to perform, on average, worse than the DP policy. The DP policy is theoretically the optimal policy if demand is independently distributed between periods, so no other policy can be expected to result in lower average costs. In addition, the RT policy is derived by assuming that the inventory process is a renewal process, so it will not be as efficient in a finite-horizon setting as it will be in an infinite-horizon setting. The assumption of time independence also has a major impact on how the RT policy uses the available data compared to how the DP policy does so: While the RT policy uses the time-series structure of the data samples as they are and considers each data sample individually (yielding n samples), the DP policy considers all possible combinations of each period's data sample (yielding all n^T combinations of samples). Thus, the RT policy is supposedly more likely to suffer from small data sets than the DP policy is, especially if the demand correlation between periods is small. Figure 4.1 shows an example of the RT policy, the DP policy and the EPO policy for the inventory processes over time for a randomly selected product from the data set.

The results in Table 4.1 show that the average costs C_{RT} of the RT policy are higher than the average costs C_{DP} of the DP policy, which supports the intuition that the DP is theoretically optimal. However, when the policies' gap to optimality, $\Delta_{RT} = C_{RT}/C_{EPO}$ and $\Delta_{DP} = C_{DP}/C_{EPO}$, are compared, the difference, $\Delta_{RT-DP} = \Delta_{RT} - \Delta_{DP}$, ranges between -2.92 and 27.01 percent, depending on the cost parameters, with an average

Policy Evaluation Single Product						
(h, b, k)	C_{RT}	Δ_{RT}	C_{DP}	Δ_{DP}	C_{EPO}	Δ_{RT-DP}
(0.05, 1, 50)	380.80	51.32 %	345.70	37.37 %	251.66	13.95 %
(0.05, 1.5, 50)	396.03	57.37 %	360.14	43.11 %	251.66	14.26 %
(0.05, 2, 50)	415.19	64.98 %	375.56	49.23 %	251.66	15.75 %
(0.05, 1, 75)	415.97	44.24 %	391.87	35.89 %	288.38	8.35 %
(0.05, 1.5, 75)	436.90	51.50 %	404.33	40.21 %	288.38	11.29 %
(0.05, 2, 75)	454.84	57.72 %	417.94	44.93 %	288.38	12.79 %
(0.05, 1, 100)	454.24	40.98 %	435.17	35.06 %	322.21	5.92 %
(0.05, 1.5, 100)	474.93	47.40 %	447.16	38.78 %	322.21	8.62 %
(0.05, 2, 100)	491.91	52.67 %	459.02	42.46 %	322.21	10.21 %
(0.1, 1, 50)	432.66	52.24 %	406.68	43.10 %	284.19	9.14 %
(0.1, 1.5, 50)	483.29	70.06 %	427.98	50.60 %	284.19	19.46 %
(0.1, 2, 50)	510.85	79.76 %	447.17	57.35 %	284.19	22.41 %
(0.1, 1, 75)	478.38	45.16 %	465.48	41.24 %	329.56	3.92 %
(0.1, 1.5, 75)	532.46	61.57 %	485.96	47.46 %	329.56	14.11 %
(0.1, 2, 75)	553.35	67.91 %	501.65	52.22 %	329.56	15.69 %
(0.1, 1, 100)	521.26	40.75 %	516.89	39.57 %	370.35	1.18 %
(0.1, 1.5, 100)	573.36	54.82 %	535.34	44.55 %	370.35	10.27 %
(0.1, 2, 100)	593.81	60.34 %	548.96	48.23 %	370.35	12.11 %
(0.15, 1, 50)	471.51	53.33 %	448.86	45.96 %	307.52	7.37 %
(0.15, 1.5, 50)	520.29	69.19 %	470.22	52.91 %	307.52	16.28 %
(0.15, 2, 50)	576.90	87.60 %	493.85	60.59 %	307.52	27.01 %
(0.15, 1, 75)	519.19	44.30 %	515.02	43.14 %	359.81	1.16 %
(0.15, 1.5, 75)	573.33	59.34 %	536.98	49.24 %	359.81	10.10 %
(0.15, 2, 75)	625.26	73.78 %	557.60	54.97 %	359.81	18.81 %
(0.15, 1, 100)	564.47	38.99 %	576.31	41.91 %	406.11	-2.92 %
(0.15, 1.5, 100)	620.60	52.82 %	594.94	46.50 %	406.11	6.32 %
(0.15, 2, 100)	673.59	65.86 %	617.11	51.96 %	406.11	13.90 %
						∅ 11.39 %

Table 4.1: Numerical results for the single-product setting.

difference of 11.39 percent. Although the amount of data evaluated does not make these results statistically significant, they still show that our suggested approach does perform well compared to the theoretically optimal DP policy, with even one instance of lower average costs for the RT policy. Figure 4.2 visualizes the difference in relative costs between the RT and the DP policies. For this setting, the difference decreases in fixed order costs, increases in back-order costs and either increases or decreases in holding costs, depending on the holding costs. Increasing back-order costs creates more uncertainty because stock-outs are punished more heavily. Therefore, a policy that can



(a) $h = 0.05$ (b) $h = 0.1$ (c) $h = 0.15$

Figure 4.2: Difference in relative performance Δ_{RT-DP} for the single-product setting depending on b for different values of k.

deal with uncertainty better will likely result in lower average inventory costs.

As mentioned earlier, one main difference between the RT and the DP policy is how historical data evaluated: While for the DP policy it is always assumed that there is no correlation between the various periods' demands — demand does not have to be stationary, but it is assumed to be i.i.d . — so the available data's time series structure is ignored, and the RT policy uses the demand samples with their implied timely structure. Therefore, the DP approach considers the potential combinations of demands in different periods from different demand samples, while the RT approach considers each time series of demands in each demand sample individually. Practically speaking, the DP policy solution uses the same amount of available data but considers more combinations of periodic demands, so it considers n^T possible demand samples instead of only n. The DP policy cannot assume correlated demands between each period, but the RT policy can relax this assumption by, for example, considering all n^T possible demand samples or generating demand samples based on certain probabilities. With respect to the data used in this analysis, the correlation between consecutive days is low — the mean auto-correlation coefficient is 0.0153 — and no obvious correlation between the auto-correlation and the performance of the RT policy compared to the DP policy could be identified.

In summary, this first brief analysis of the RT policy in the single-product setting shows that the approach performs well and is not too far away from the theoretically optimal DP policy in terms of costs. While this brief analysis does not explain the intricacies in the difference between RT policy and DP policy, it shows that the RT can be a suitable approach to this type of inventory problem. Future research may consider analyzing the influence of, for example, sample size, demand correlation, and length of planning period on the RT policy's performance.

4.4.3 Multi-Product Setting

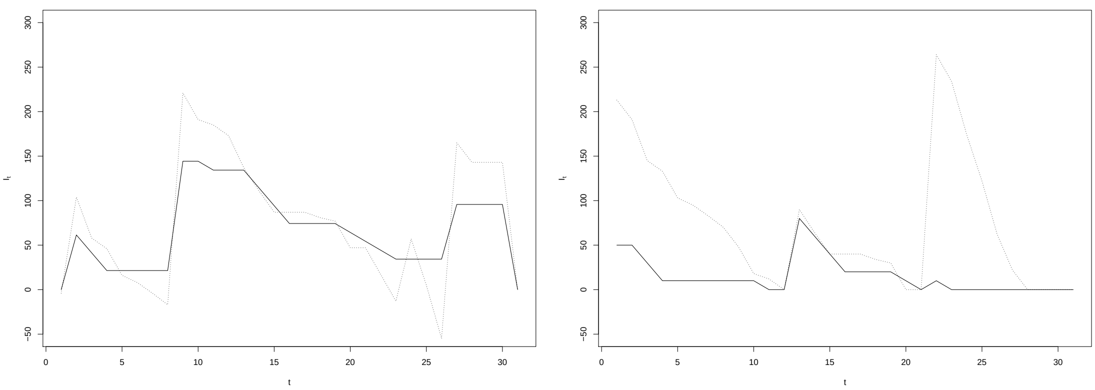
While the RT policy delivers satisfactory results in the single-product setting,whether its results are equally satisfactory in the multi-product setting remains in question. As multiple inventories are managed simultaneously, this problem is complex so, to keep the analysis simple, in each instance the inventory of two products is managed simultaneously. The fixed order cost function is defined as:

$$\mathbf{K}_t^{i,j}(\mathbf{S}_t) = \begin{cases} \frac{1}{I}k, & \text{if } S_t^i > I_t^i \text{ for any } i \in \{1, \dots, I\}, \\ 0, & \text{else,} \end{cases} \tag{4.24}$$

such that fixed order costs k are incurred if an order for either product is placed (and these costs are shared between all products $i \in \{1, \dots, I\}$), and no fixed order costs are incurred if no order is placed. The purchasing cost function is defined as:

$$p(q) = \begin{cases} p, & q < \theta, \\ (1-\epsilon)p, & q \geq \theta, \end{cases} \tag{4.25}$$

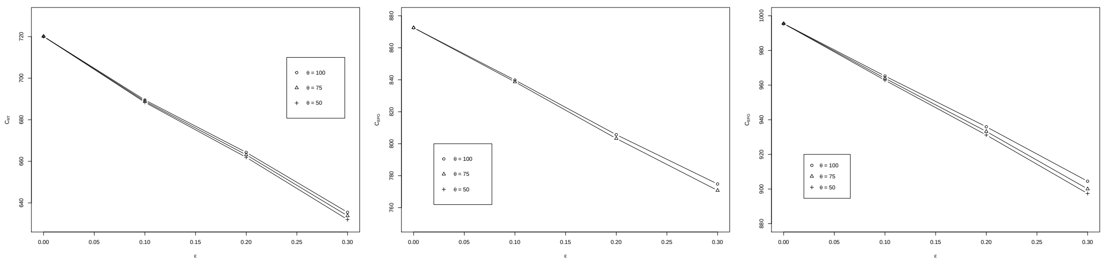
with θ as the order quantity threshold at which the discount $\epsilon \geq 0$ is triggered. To keep the analysis simple, the discount $\epsilon \in \{0, 0.1, 0.2, 0.3\}$, the threshold $\theta \in \{50, 75, 100\}$ and the holding cost $h \in \{0.05, 0.1, 0.15\}$ are varied, while the remaining parameters are fixed at $b = 1$, $k = 50$, and $p = 0.5$, respectively.



(a) RT policy (b) EPO policy

Figure 4.3: Exemplary inventory processes of a two-product inventory (solid line: product 1, dotted line: product 2) over time for a randomly selected product pair.

Figure 4.3 shows as an example, for randomly selected products from the data set, how the inventories of two products develop over the course of one planning period (31 days), for the RT policy (Figure 4.3a) and the EPO policy (Figure 4.3b). As in the single-product setting, the EPO policy avoids shortages and orders just the right amount. In addition, the orders for both products are synchronized, so fixed costs are shared for each replenishment and quantity discounts can be utilized. In comparison, the RT policy does not plan deterministically, resulting in shortages as well as excess inventory. Replenishments are aligned in three out of four replenishments, and there is one individual order.



(a) $h = 0.05$ (b) $h = 0.1$ (c) $h = 0.15$

Figure 4.4: Average costs of the RT policy for the multi-product setting.

Table 4.2 contains the results of the evaluation for the data set. The relative gap to optimality Δ_{RT} of the RT policy varies between 54.51 percent and 61.16 percent, which is only slightly higher than the gap in similar settings in the single-product setting (compare the results in Table 4.1). When the base settings without quantity discounts ($\epsilon = 0$) are compared, the RT policy's relative gap to optimality in the multi-product setting increases by only 3.19 percent, 6.57 percent, and 8.7 percent, respectively, compared to the single-product setting. Acknowledging that, because of the problem's higher dimensionality it becomes significantly more difficult to solve (which is why there is no go-to benchmark), the RT policy performs well. In the setting without quantity discounts, the inventory costs for the two products can be reduced compared to the individually managed products only if orders are synchronized, as only then can the fixed ordering costs be shared. This results in average costs of 720.07 per product pair (for $h = 0.05$), which is an average of $720.07/2 = 360.035$ per product. In the single-product analysis, the average inventory costs C_{RT} per product were 380.80 for the same setting, which is approximately 5.77 percent higher. When the possibility of quantity discounts are added ($\epsilon > 0$ and $\theta > 0$), the average costs C_{RT} are decreasing in the

Policy Evaluation Multi Product			
(ϵ, θ)	C_{RT}	Δ_{RT}	C_{EPO}
$h = 0.05$			
(0, 100)	720.07	54.51 %	461.08
(0.1, 100)	689.51	55.62 %	434.83
(0.2, 100)	664.30	56.62 %	408.30
(0.3, 100)	635.47	58.72 %	381.62
(0.1, 75)	688.89	55.84 %	434.19
(0.2, 75)	663.13	57.12 %	407.17
(0.3, 75)	633.79	59.23 %	380.04
(0.1, 50)	688.37	55.96 %	433.73
(0.2, 50)	661.92	57.25 %	406.35
(0.3, 50)	631.98	59.22 %	378.95
$h = 0.1$			
(0, 100)	872.52	58.81 %	518.67
(0.1, 100)	839.82	59.30 %	493.58
(0.2, 100)	805.66	59.80 %	468.07
(0.3, 100)	774.82	61.17 %	442.28
(0.1, 75)	838.69	59.38 %	492.56
(0.2, 75)	803.27	59.94 %	466.19
(0.3, 75)	770.78	61.16%	439.64
(0.1, 50)	837.81	59.59 %	491.61
(0.2, 50)	801.78	60.49 %	464.45
(0.3, 50)	768.54	61.95 %	437.23
$h = 0.15$			
(0, 100)	995.48	62.03 %	559.58
(0.1, 100)	965.24	62.57 %	535.50
(0.2, 100)	935.94	63.55 %	510.81
(0.3, 100)	904.49	64.63 %	485.76
(0.1, 75)	963.94	62.67 %	534.18
(0.2, 75)	933.25	63.73 %	508.39
(0.3, 75)	900.05	64.79 %	482.36
(0.1, 50)	962.81	62.94 %	532.80
(0.2, 50)	931.19	64.22 %	505.91
(0.3, 50)	897.41	65.67 %	478.94

Table 4.2: Numerical results for the multi-product setting.

discount ϵ and increasing in the discount threshold θ, confirming that the policy uses economies of scale that are due to fixed order costs and quantity discounts (Figure 4.4). Furthermore, it is interesting to note that the relative gap to optimality slightly increases in the discount factor ϵ, which is intuitive: as the potential quantity discount increases, it becomes more important to coordinate the orders of different inventories. While this coordination is complex for a stochastic inventory policy, the EPO policy does always achieve the lowest possible inventory costs, so the gap between stochastic and ex-post optimal policies is expected to increase in ϵ.

In summary, the RT policy's relative gap to optimality Δ_{RT} in the multi-product setting increases only slightly compared to the results in the single-product setting (compare tables 4.1 and 4.2); that the RT policy's costs are close to those of the DP policy in the single-product setting suggests that the RT policy's costs are close to the (unknown) theoretically optimal policy for the multi-product setting. Besides confirming that the policy uses economies of scale in both fixed order costs and quantity discounts, this evaluation shows that the RT policy might be a suitable approach to managing multi-product inventory problems. However, further analyses are needed to confirm these early results.

4.5 Conclusion

This paper proposes a novel data-driven approach, referred to as the RT policy, to dealing with multi-period inventory problems, also known as JRPs. A first analysis using real-world data shows that the proposed model performs well. In fact, for simple settings with single-product replenishment and no quantity discounts, the model's costs are close to the costs of the theoretically optimal approach. For complex multi-product settings (JRPs) with quantity discounts, where optimal policies cannot be derived using general assumptions, the performance is still similar to EPO costs under deterministic planning, which suggests that the policy performs well in complex settings.

The first analyses of the RT policy look promising and lead to several avenues for future research. First, the policy could be extended to include lead time, which would improve its applicability. This extension could easily be made by including the time an order takes until it arrives into the replenishment cycle. Second, the amount of data and how the policy handles data samples (in sequence) affect the expected costs, so a more rigorous analysis with either a simulation or more data is necessary to confirm the policy's performance. Furthermore, for the multi-product setting this study compared the policy only to the EPO policy, but identifying the most significant (data-driven) approaches for this setting and comparing their performance to that of the RT policy is important. With respect to real-world applications, a study that combines the RT policy with forecasting tools like machine learning models would be useful. The weighted SAA Bertsimas and Kallus (2018) introduced could be naturally suited to such as study.

Chapter 5

Conclusion

To reach global health goals like the SDGs the global health industry must make the most out of limited funds and healthcare budgets by decreasing procurement costs so they can increase the supply of essential medicines and medical equipment. However, a focus on costs can introduce additional risks, such as poor product quality and supply shortages, because low-cost suppliers often do not fulfill all the requirements of sustainable, high-quality production, so already tight budgets might be "wasted on products of unknown quality with potentially devastating effects for public health" (FM't Hoen et al., 2014, p. 22). A profound understanding of the various drivers of risk and competition can help decision-makers to optimize their allocation of procurement volumes, set the right incentives via procurement mechanisms, and determine the optimal order timing. The three articles in this thesis explore and analyze decision problems in these three areas to provide the insights and tools needed to improve decision-making in global health procurement.

Chapter 2 analyzes how a buyer should allocate procurement volumes between an incumbent supplier and a new-entrant supplier when lead times are uncertain, when the entrant may not be registered in all eligible countries, and when production capacities are limited. The chapter shows how the entrant's value to the buyer changes depending on several assumptions. The buyer's objective function can have multiple local maxima and minima, which renders the process of making the optimal decision intricate, so simple rules of thumb (e.g., a 70/30% split) that are regularly used in practice may not be the best way to solve these types of procurement problems. From a methodological standpoint, we contribute to the literature on bargaining models by incorporating dual sourcing into a bilateral bargaining model with capacity constraints. However, our results have some limitations, as we rely on simulation to keep our results practically relevant. While this approach helped us incorporate more aspects of the actual problem, from a scientific standpoint it limits our findings' general applicability. Therefore, we suggest that future research develop a closed-form model to derive generalizable findings.

In Chapter 3 we analyzed a procurement mechanism called the postponement tender, which allows a buyer to postpone part of the volume allocation so the buyer can learn about the suppliers' uncertain quality before allocating the rest of the volume. In this mechanism, the suppliers' bid their prices, and the buyer allocates initial volumes to them. Then, after the suppliers have delivered the initial volume, the buyer observes the suppliers' quality and allocates the postponed volume to the supplier with the higher value (based on quality and price). Our analysis shows that a buyer can benefit from this mechanism more than it can from simple single-sourcing and shows how the buyer should set the initial and postponed volume optimally based on the procurement volume, the supplier's costs, and the quality uncertainty. Using a postponement tender, the buyer can induce competition and learn about uncertain quality, and substantially improve its procurement outcome. In our research we analyze the postponement tender and compare its performance to single sourcing. Future research could compare the postponement tender to other procurement formats, such as repeated tenders and (repeated) negotiations, to shed more light on the question concerning which procurement mechanism a buyer should use in which procurement situation.

The third problem, considered in Chapter 4, is the Joint Replenishment Problem (JRP) with quantity discounts, an inventory-management problem in which multiple inventories of multiple products are ordered on a recurrent basis. Economies of scale in both fixed and variable order costs can be used by coordinating orders for multiple products. We developed novel, data-driven inventory policy using sample average approximation that reduces the complexity of the underlying coordination problem. A first numerical evaluation using real-world data from Kenyan pharmacies showed that, in the single-product setting, the policy's costs are similar to the costs of the theoretically optimal policy. In the complex multi-product setting, where no optimal policies are known, the policy still yields lower costs than the ex-post optimal costs. The results presented are promising, but additional research is needed to clarify the applicability and performance of this approach. This data-driven policy could also be combined with a machine-learning methodology to improve the approach's performance by adding forecasting techniques to the inventory policy.

Bibliography

Y. Akçay and S. H. Xu. Joint inventory replenishment and component allocation optimization in an assemble-to-order system. *Management Science*, 50(1):99–116, 2004.

J.J. Anton and D.A. Yao. Split awards, procurement, and innovation. *The RAND Journal of Economics*, 20(4):538–552, 1989.

D. R. Atkins and P. O. Iyogun. Periodic versus "can-order" policies for coordinated multi-item inventory systems. *Management Science*, 34(6):791–796, 1988.

M. Ausubel. An efficient ascending-bid auction for multiple objects. *American Economic Review*, 94(5):1452–1475, 2004.

V. Babich, A. N. Burnetas, and P. H. Ritchken. Competition and diversification effects in supply chains with supplier default risk. *Manufacturing & Service Operations Management*, 9(2):123–146, 2007a.

V. Babich, A.N. Burnetas, and P.H. Ritchken. Competition and diversification effects in supply chains with supplier default risk. *Manufacturing & Service Operations Management*, 9(2):123–146, 2007b.

J. L. Balintfy. On a basic class of multi-item inventory problems. *Management science*, 10(2):287–297, 1964.

G.-Y. Ban. Confidence intervals for data-driven inventory policies with demand censoring. *Available at SSRN 2654014*, 2019.

D.A. Berry. Bayesian clinical trials. *Nature reviews Drug discovery*, 5(1):27–36, 2006.

D. Bertsimas and A. Thiele. A data-driven approach to newsvendor problems. *Working Paper, Massachusetts Institute of Technology*, 2005.

D. Bertsimas and A. Thiele. Robust and data-driven optimization: modern decision making under uncertainty. In *Models, methods, and applications for innovative decision making*, pages 95–122. INFORMS, 2006.

Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *arXiv preprint arXiv:1402.5481v4*, 2018.

K. Binmore, A. Rubinstein, and A. Wolinsky. The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 17(2):176–188, 1986.

S. Çetinkaya and C.-Y. Lee. Stock replenishment and shipment scheduling for vendor-managed inventory systems. *Management Science*, 46(2):217–232, 2000.

B. C. Cha and I. K. Moon. The joint replenishment problem with quantity discounts under constant demand. *OR Spectrum*, 27(4):569–581, 2005.

Aadhaar Chaturvedi, Damian R Beil, and Victor Martínez-de Albéniz. Split-award auctions for supplier retention. *Management Science*, 60(7):1719–1737, 2014.

C. Chiang and W.C. Benton. Sole sourcing versus dual sourcing under stochastic demands and lead times. *Naval Research Logistics (NRL)*, 41(5):609–624, 1994.

Chisale, M. Sixth Meeting of African Ministers of Health of Small Island Developing States - Pooled Procurement, 2017. URL `http://www.aho.afro.who.int/networks/sites/default/files/ro_pooled_procurement_0.pdf`.

A. Collard-Wexler, G. Gowrisankaran, and R.S. Lee. "Nash-in-Nash" bargaining: A microfoundation for applied work. *Journal of Political Economy*, forthcoming in 2018.

CSP/RHSC. Coordinated Supply Planning(CSP) Working Group/RHSC – Intramuscular DMPA Demand Forecast, Supplier Meeting, August 17, 2015 (New York City), 2015.

M. Dada, N. C. Petruzzi, and L. B. Schwarz. A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management*, 9 (1):9–32, 2007.

W. Elmaghraby. Auctions within e-sourcing events. *Production and Operations Management Society*, 16(4):409–422, 2007.

R. Engelbrecht-Wiggans and E. Katok. Regret in auctions: Theory and evidence. *Economic Theory*, 33(1):81–101, 2007.

European Parliament Research Service. Addressing shortages of medicines, European Parliament Research Service, PE 649.402 - April 2020, 2020.

A. Federgruen and N. Yang. Selecting a portfolio of suppliers under demand and supply risks. *Operations Research*, 56(4):916–936, 2008.

Q. Feng and L.X. Lu. Supply chain contracting under competition: Bilateral bargaining vs. stackelberg. *Production and Operations Management*, 22(3):661–675, 2013.

E. FM't Hoen, H.V. Hogerzeil, J.D. Quick, and H.B. Sillo. A quiet revolution in global public health: The World Health Organization's prequalification of medicines programme. *Journal of Public Health Policy*, 35(2):137–161, 2014.

FP2020. Family Planning 2020 Focus Countries, 2016. URL `http://www.familyplanning2020.org/entities`.

N. Fugger, E. Katok, and A. Wambach. Collusion in dynamic buyer-determined reverse auctions. *Management Science*, 62(2):518–533, 2015.

Gavi. Supply and Procurement Strategy 2016-20, 2016. URL `https://www.gavi.org/library/gavi-documents/supply-procurement/supply-and-procurement-strategy-2016-20--overview/`.

Global Drug Facility. Ivitation to Bid: ITB-IDA/GDF - FLD/2013/005, 2013.

Global Drug Facility. Invitation to Bid: ITB-IDA/GDF - FLD/2014/001, 2014.

Global Fund. Request for Proposal (RFP) TGF-14-400 Antiretroviral Medicins, 2014a.

Global Fund. Request for Proposal (RFP) TGF-14-010 Artimisinan Based Combination Thearpies, 2014b.

Global Fund. Guide to Global Fund Policies on Procurement and Supply Management of Health Products, 2017a. URL `https://www.theglobalfund.org/media/5873/psm_procurementsupplymanagement_guidelines_en.pdf`.

Global Fund. Global fund – price & quality reporting, 2017b. URL `https://www.theglobalfund.org/en/sourcing-management/price-quality-reporting/`.

Global Fund. Request for Proposal (RFP) TGF-17-001 Antimalarial Medicines, 2017c.

Global Fund. Request for Proposal (RFP) TGF-18-010 Antiretrovirals and other selected medicines used in HIV programs, 2018.

J. Gong, J. Li, and R.P. McAfee. Split-award contracts with investment. *Journal of Public Economics*, 96(1):188–197, 2012.

S. K. Goyal. Determination of optimum packaging frequency of items jointly replenished. *Management Science*, 21(4):436–443, 1974.

A.K. Gupta and S. Nadarajah. *Handbook of beta distribution and its applications.* Marcel Dekker, Inc., 2004.

E. Haruvy and E. Katok. Increasing revenue by decreasing information in procurement auctions. *Production and Operations Management*, 22(1):19–35, 2013.

H. S. Heese. Single versus multiple sourcing and the evolution of bargaining positions. *Omega*, 54:125–133, 2015.

H. Horn and A. Wolinsky. Bilateral monopolies and incentives for merger. *The RAND Journal of Economics*, 19(3):408–419, 1988.

E. Iakovou, D. Vlachos, C. Keramydas, and D. Partsch. Dual sourcing for mitigating humanitarian supply chain disruptions. *Journal of Humanitarian Logistics and Supply Chain Management*, 4(2):245–264, 2014.

S.D. Jap. Online reverse auctions: Issues, themes, and prospects for the future. *Journal of the Academy of Marketing Science*, 30(4):506–525, 2002.

E. Katok and A. Wambach. Collusion in dynamic buyer-determined reverse auctions. *Management Science*, pages 1–27, 2008.

P. Kelle and P.A. Miller. Decreasing expected shortages through order splitting. *Engineering Costs and Production Economics*, 19(1):351–357, 1990.

P. Kelle and P.A. Miller. Stockout risk and order splitting. *International Journal of Production Economics*, 71(1–3):407–415, 2001.

D.E. Klotz and K. Chatterjee. Dual sourcing in repeated procurement competitions. *Management Science*, 41(8):1317–1327, 1995.

J. Komrska, L.R. Kopczak, and J.M. Swaminathan. When supply chains save lives. *Supply Chain Management Review*, 17(1):42–49, 2013.

P. Kouvelis and J. Li. Flexible backup supply and the management of lead-time uncertainty. *Production and Operations Management*, 17(2):184–199, 2008.

V. Krishna. *Auction theory.* Academic press, 2009.

M. Kumar, P. Basu, and B. Avittathur. Pricing and sourcing strategies for competing retailers in supply chains under disruption risk. *European Journal of Operational Research*, 265(2):533–543, 2018.

F. Lauton, A. Rothkopf, and R. Pibernik. The value of entrant manufacturers: A study of competition and risk for donor-funded procurement of essential medicines. *European Journal of Operational Research*, 272(1):292–312, 2019.

H.L. Lee and C.S. Tang. Modelling the costs and benefits of delayed product differentiation. *Management science*, 43(1):40–53, 1997.

R. Levi, R. O. Roundy, and D. B. Shmoys. Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, 32(4):821–839, 2007.

L. Liu and X.-M. Yuan. Coordinated replenishments in inventory systems with correlated demands. *European Journal of Operational Research*, 123(3):490–503, 2000.

Management Sciences for Health. *MDS-3: Managing Access to Medicines and Health Technologies*. Management Sciences for Health, Arlington, VA, 2013.

S. Minner. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81:265–279, 2003.

C. Moellers, H. Normann, and C.M. Snyder. Communication in vertical markets: Experimental evidence. *International Journal of Industrial Organization*, 50:214–258, 2017.

J.F. Nash. The bargaining problem. *Econometrica: Journal of the Econometric Society*, pages 155–162, 1950.

K.V. Natarajan and J.M. Swaminathan. Inventory management in humanitarian operations: Impact of amount, schedule, and uncertainty in funding. *Manufacturing & Service Operations Management*, 16(4):595–603, 2014.

M.K. Perry and J. Sákovics. Auctions for split-award contracts. *The Journal of Industrial Economics*, 51(2):215–242, 2003.

S.J. Press. *Bayesian Statistics: Principles, Models and Applications.* Wiley, New York, 1989.

L. Qi, J. J. Shi, and X. Xu. Supplier competition and its impact on firm s sourcing strategy. *Omega*, 55:91–110, 2015.

R.V. Ramasesh, J.K. Ord, J.C. Hayya, and A. Pan. Sole versus dual sourcing in stochastic lead-time (s, Q) inventory models. *Management Science*, 37(4):428–443, 1991.

RHInterchange. RHInterchange, 2016. URL `https://www.unfpaprocurement.org/rhi-home`.

RHSC. Reproductive Health Supplies Coalition Family Planning Market Report 2016, 2016. URL `http://www.clintonhealthaccess.org/2016-fp-market-report/`.

J. D. Sachs and J. W. McArthur. The millennium project: a plan for meeting the millennium development goals. *The Lancet*, 365(9456):347–353, 2005.

N. Santamaria. An analysis of scoring and buyer-determined procurement auctions. *Production and Operations Management*, 24(1):147–158, 2015.

H. Scarf. The optimality of (s, S) policies in the dynamic inventory problem. *Mathematical Methods in the Social Science, KJ Arrow, S. Karlin, P. Suppes, eds. (Stanford UniversityPress, Stanford)*, 1959.

L. V. Snyder, Z. Atan, P. Peng, Y. Rong, A. J. Schmitt, and B. Sinsoysal. Or/ms models for supply chain disruptions: A review. *IIE Transactions*, 48(2):89–109, 2016.

J.M. Swaminathan and H.L. Lee. Design for postponement. *Handbooks in Operations Research and Management Science*, 11:199–226, 2003.

J.M. Swaminathan and S.R. Tayur. Managing broader product lines through delayed differentiation using vanilla boxes. *Management Science*, 44(12-part-2):S161–S172, 1998.

T.A. Taylor and W. Xiao. Subsidizing the distribution channel: Donor funding to improve the availability of malaria drugs. *Management Science*, 60(10):2461–2477, 2014.

B. Tomlin. On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, 52(5):639–657, 2006.

B. Tomlin. Impact of supply learning when suppliers are unreliable. *Manufacturing & Service Operations Management*, 11(2):192–209, 2009.

B. Tomlin and Y. Wang. On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing & Service Operations Management*, 7(1):37–57, 2005.

R. Tren, K. Hess, and R. Bate. Drug procurement, the global fund and misguided competition policies. *Malaria Journal*, 8(1):305, 2009.

Ö. Turgut, F. Taube, and S. Minner. Data-driven retail inventory management with backroom effect. *OR Spectrum*, 40(4):945–968, 2018.

UNFPA. Quality Assurance Policy for Reproductive Health Medicines, 2012. URL `http://www.unfpa.org/resources/reproductive-health-medicines`.

UNFPA. Policy and Procedures for Regular Procurement, 2015. URL `http://www.unfpa.org/admin-resource/policy-and-procedures-regular-procurement`.

UNICEF Supply Division. Yellow Fever Vaccine: Current Outlook, 2013. URL `https://www.unicef.org/supply/files/Yellow_Fever_Vaccine_Current_Outlook_Nov_2013.pdf`.

USAID. Measuring Supply Chain Performance, 2010. URL `http://pdf.usaid.gov/pdf_docs/Pnadx524.pdf`.

USAID. Procurement Performance Indicators Guide, 2013. URL `http://apps.who.int/medicinedocs/documents/s20157en/s20157en.pdf`.

USAID. Healthy Markets for Global Health: A Market Shaping Primer, 2014. URL `https://www.usaid.gov/sites/default/files/documents/1864/healthymarkets_primer.pdf`.

USAID. HIV Rapid Test Kits - USAID Approval and Technical Guidance, 2016. URL `https://www.usaid.gov/sites/default/files/documents/1864/Guidance-on-HIV-RTKs-Dec2016-508\_0.pdf`.

J.A.A Van Mieghem and M. Dada. Price versus production postponement: Capacity and competition. *Management Science*, 45(12):1639–1649, 1999.

S. Viswanathan. Note. periodic review (s, s) policies for joint replenishment inventory systems. *Management Science*, 43(10):1447–1454, 1997.

S. Viswanathan. On optimal algorithms for the joint replenishment problem. *Journal of the Operational Research Society*, 53(11):1286–1290, 2002.

Z. Wan, D. R. Beil, and E. Katok. When does it pay to delay supplier qualification? theory and experiments. *Management Science*, 58(11):2057–2075, 2012.

WDI/CF. William-Davidson-Institute/Concept Foundation – Intramuscular DMPA Demand Forecast, Supplier Meeting, August 17, 2015 (New York City), 2015.

WHO. Health in 2015: from MDGs, millennium development goals to SDGs, sustainable development goals, 2015. URL https://www.who.int/gho/publications/mdgs-sdgs/en/.

WHO. Prequalification Program – A United Nations program managed by the WHO, 2016. URL http://apps.who.int/prequal.

WHO. Procurement mechanisms and systems, 2019. URL https://www.who.int/immunization/programmes_systems/procurement/mechanisms_systems/pooled_procurement/en/index1.html.

R. E. Wildeman, J. B. G. Frenk, and R. Dekker. An efficient optimal solution method for the joint replenishment problem. *European Journal of Operational Research*, 99 (2):433–444, 1997.

Z. Yang, G. Aydin, V. Babich, and D.R. Beil. Using a dual-sourcing option in the presence of asymmetric information about supplier reliability: Competition vs. diversification. *Manufacturing & Service Operations Management*, 14(2):202–217, 2012.

H. Yu, A.Z. Zeng, and L. Zhao. Single or dual sourcing: decision-making in the presence of supply chain disruption risks. *Omega*, 37(4):788–800, 2009.

J. Zhang. Cost allocation for joint replenishment models. *Operations research*, 57(1): 146–156, 2009.

M. Zhu and A. Y. Lu. The counter-intuitive non-informative prior for the bernoulli family. *Journal of Statistics Education*, 12(2):1–10, 2004.

P. H. Zipkin. *Foundations of inventory management*. McGraw-Hill, Boston, 2000.

Chapter 6

Appendix to Chapter 2

Figures

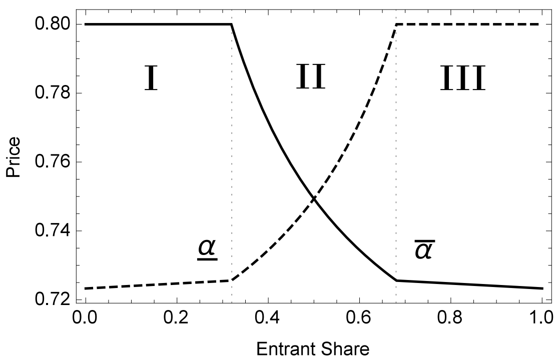


Figure 6.1: Equilibrium prices $p_E^*(\alpha)$ (solid) and $p_I^*(\alpha)$ (dashed) depending on the entrant's share (α) in the reference case.

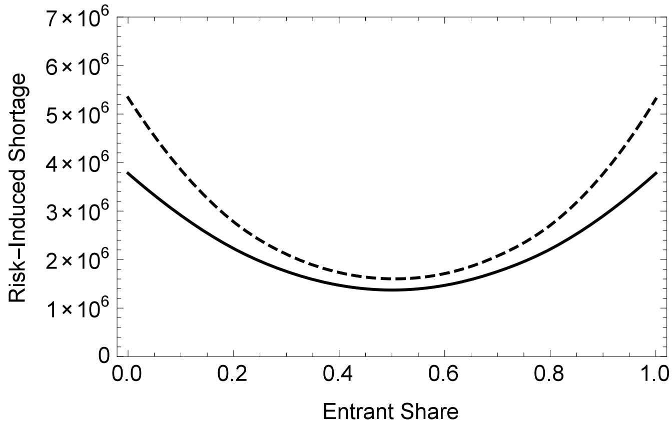


Figure 6.2: Risk-induced shortage (solid) and corresponding standard deviation depending on the entrant's share for the reference case.

Tables

Country (c)	Full Reg.	Lim. Reg.
Afghanistan	x	
Angola	x	
Bangladesh	x	
Benin	x	
Burkina Faso	x	x
DR Congo	x	
Ethiopia	x	
Guinea	x	
Haiti	x	
Jordan	x	
Kenya	x	x
Liberia	x	
Madagascar	x	x
Malawi	x	x
Mali	x	
Mozambique	x	
Nepal	x	x
Nigeria	x	x
Pakistan	x	
Rwanda	x	
Senegal	x	x
Sierra Leone	x	
Tanzania	x	x
Togo	x	
Uganda	x	x
Zambia	x	x

Table 6.1: Entrant's registered countries in case of full registration and limited registration for USAID.

Variable	Description
α, $\hat{\alpha}$	volume split (in percent) allocated to entrant and adjusted volume split
b	PO's procurement budget
cap_j	(absolute) production capacity of manufacturer $j \in \{I, E\}$
c_j, $\hat{c}_j$	production costs of manufacturer $j \in \{I, E\}$, and lower boundary of costs
$d(\alpha)$, d^I, d^{IE}	actual procurement volume, actual procurement volume from countries where only the incumbent (I) or incumbent and entrant (IE) are registered
D_t^I, D_t^{IE}, d_t^I, d_t^{IE}	(random) demand at time t from countries where only the incumbent (I) or incumbent and entrant (IE) are registered; corresponding realizations: d_t^I, d_t^{IE}
L_j, l_j	(random) supply lead time from manufacturer $j \in \{I, E\}$, corresponding realization
O_t^I, $O_t^{IE}(\alpha)$, $O_t^E(\alpha)$	(random) order at time t for countries where only the incumbent (I) is registered, for countries where both manufacturers (IE) are registered and are send to the incumbent, and for countries where both manufacturers are registered and are send to the entrant (E)
$p_j(\alpha)$	per-unit price of manufacturer $j \in \{I, E\}$
$\overline{p}(\alpha)$	weighted average per-unit price
$\pi_j(\alpha)$, π_j^{sq}	profit and status-quo profit of manufacturer $j \in \{I, E\}$
q, q^I, q^{IE}	target volume, target volume of countries where only the incumbent (I) is registered, and target volume of countries where incumbent and entrant (IE) are registered
r	reservation price
R_t^I, R_t^{IE}	(random) residual virtual inventory for countries where only the incumbent (I) is registered, or where both manufacturers (IE) are registered
S_t^I, S_t^{IE}, S_t^E	(random) supply at time t send from the incumbent to countries where only the incumbent (I) is registered, or send from the incumbent to countries where both manufacturers (IE) are registered, or send from the entrant (E) to countries where both manufacturers are registered
T	planning horizon
τ	shortage threshold
$u(\alpha)$, u^{sq}	utility and status-quo utility of PO
V_t^I, V_t^{IE}	(random) virtual inventory for countries where only the incumbent (I) is registered, or where both manufacturers (IE) are registered
w	value of the entrant
$\xi(\alpha)$, $\xi^p(\alpha)$, $\xi^r(\alpha)$	total shortage, price-induced shortages, and (expected) risk-induced shortage
z_j	(absolute) volume procured from manufacturer $j \in \{I, E\}$

Table 6.2: List of important variables for the competition and the coordination model

Chapter 7

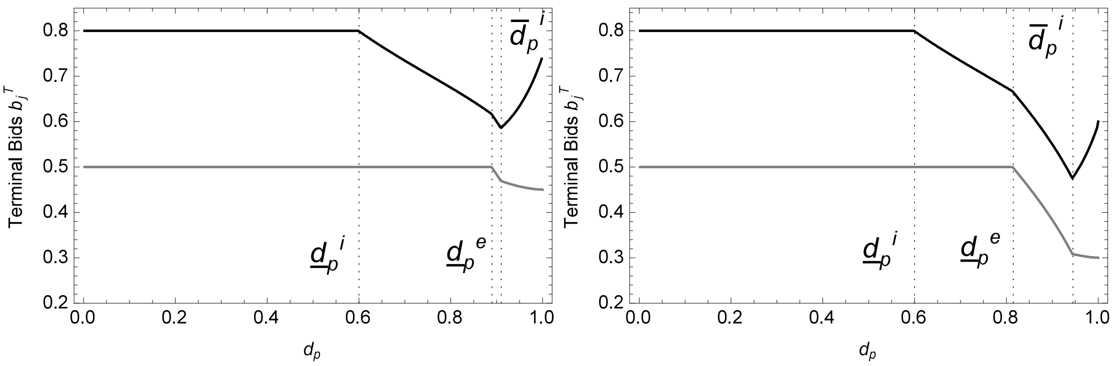
Appendix to Chapter 3

7.1 Appendix I

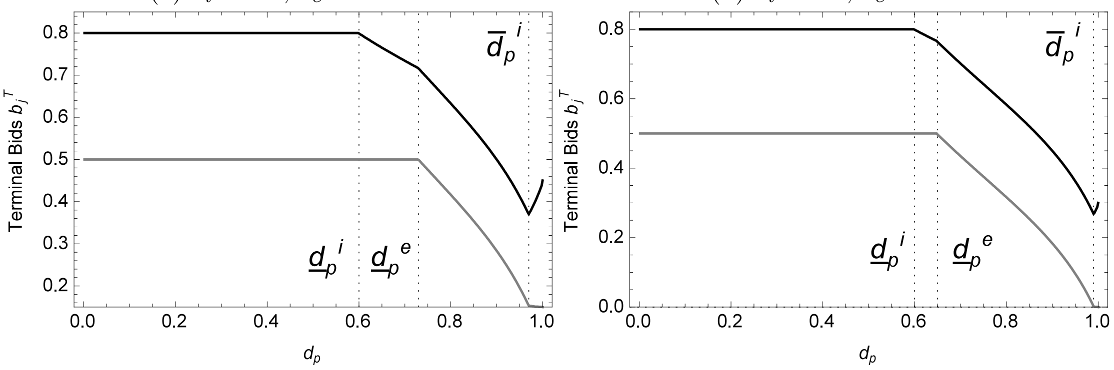
7.1.1 Further Numerical Analyses

Figure 7.1 plots the terminal bids depending on d_p for suppliers' different (positive) values. The bids of the supplier with the higher value are in line with our previous results, as the threshold $\underline{d}_p^j$ of that supplier is, in each instance, lower than the threshold of the other supplier. In addition, the bids decrease for $\underline{d}_p^j < d_p < \overline{d}_p^j$ and increase for $\overline{d}_p^j < d_p \leq 1$. For $\underline{d}_p^j < d_p < \underline{d}_p^{-j}$, supplier j lowers his bid to increase his chances of winning the postponed volume without having to consider changes in supplier $-j$'s bid. Notably, from $\underline{d}_p^{-j}$ on, supplier j's bid decreases at a higher rate because when $d_p > \underline{d}_p^{-j}$, supplier j's probability of winning the postponed volume decreases as supplier j starts to lower his bids and has to bid more aggressively to increase his expected profit (Eq. (3.5)).

More interesting are the bids of the supplier with the lower value, for whom we assumed $v_{-j} = 0$ in Proposition 7. For $0 < d_p < \overline{d}_p^j$, this supplier's bids depend on d_p, as predicted by Proposition 6, but his bids continue to decrease for $\overline{d}_p^j < d_p \leq 1$; that is, in contrast to the bids of the supplier with the higher value, the supplier with the lower value does not submit higher bids for high values of d_p. To understand supplier $-j$'s rationale first note that he still seeks to increase his probability of winning the postponed volume, even at $P(b_{-j}^{t-1}|d_p) = 0$, as long as his expected profit from lowering his bid increases (see Eq. (3.7)). But, because supplier j has a larger value he can always outbid supplier $-j$ who at some point stops bidding because the bidding condition in Eq. (3.8) is no longer fulfilled. The increasing volume incentive results in supplier $-j$'s decreasing his bid for $d_p > \overline{d}_p^j$. Thus, supplier $-j$ still lowers his bid in an (unsuccessful)



(a) $v_i = 0.6$, $v_e = 0.05$. (b) $v_i = 0.6$, $v_e = 0.2$.

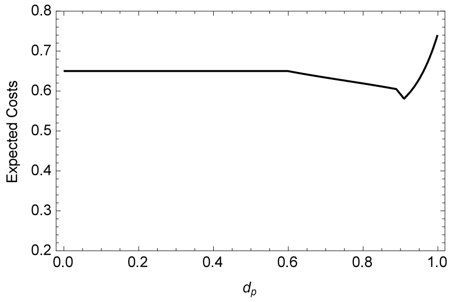


(c) $v_i = 0.6$, $v_e = 0.35$. (d) $v_i = 0.6$, $v_e = 0.5$.

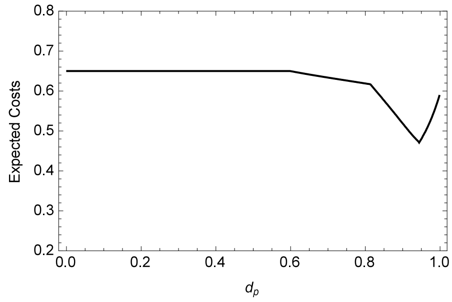
Figure 7.1: Terminal bids (b_j^T) of incumbent (black) and entrant (gray) depending on postponed volume (d_p) for $v_i \neq v_e$, $Q_e \sim Beta(1,1)$, $\bar{q}_e = 0.5$, $q_i = 0.8$ and $M = 25$.

attempt to increase the probability that he will win the (higher) postponed volume. For $d_p > \overline{d}_p^j$, supplier $-j$'s bid decreases at a lower rate because he does not have to consider the decreasing bids of supplier j in calculating the probability that he will win the postponed volume.

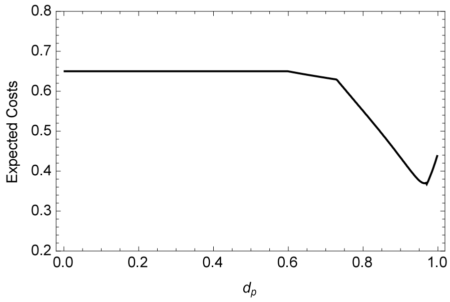
In line with findings from Proposition 6, Figure 7.1 shows that increasing the value v_{-j} reduces supplier $-j$'s lower threshold $\underline{d}_p^{-j}$. Also, decreasing v_j decreases the upper threshold $\overline{d}_p^j$, which is intuitive because as supplier $-j$ provides a lower value, his competitor j can lower his bid and be guaranteed to win the postponed volume for lower d_p. A comparison of Figures 7.1a through 7.1d shows that the region $\underline{d}_p^{-j} < d_p < \overline{d}_p^j$, where a buyer can expect strong competition, increases as supplier $-j$'s value grows closer to the higher value of supplier j, and as v_{-j} increases, the region in which a buyer observes increasing bids from supplier j decreases. This result suggests that a buyer should prefer suppliers with similar values because then she can benefit from lower prices for larger ranges of postponed volumes.



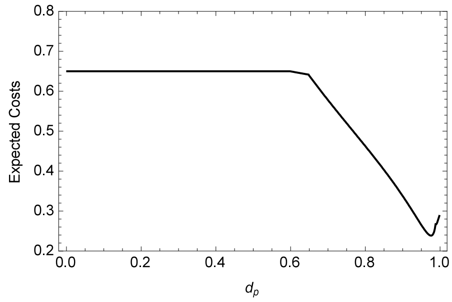
(a) $v_i = 0.6$, $v_e = 0.05$.



(b) $v_i = 0.6$, $v_e = 0.2$.



(c) $v_i = 0.6$, $v_e = 0.35$.

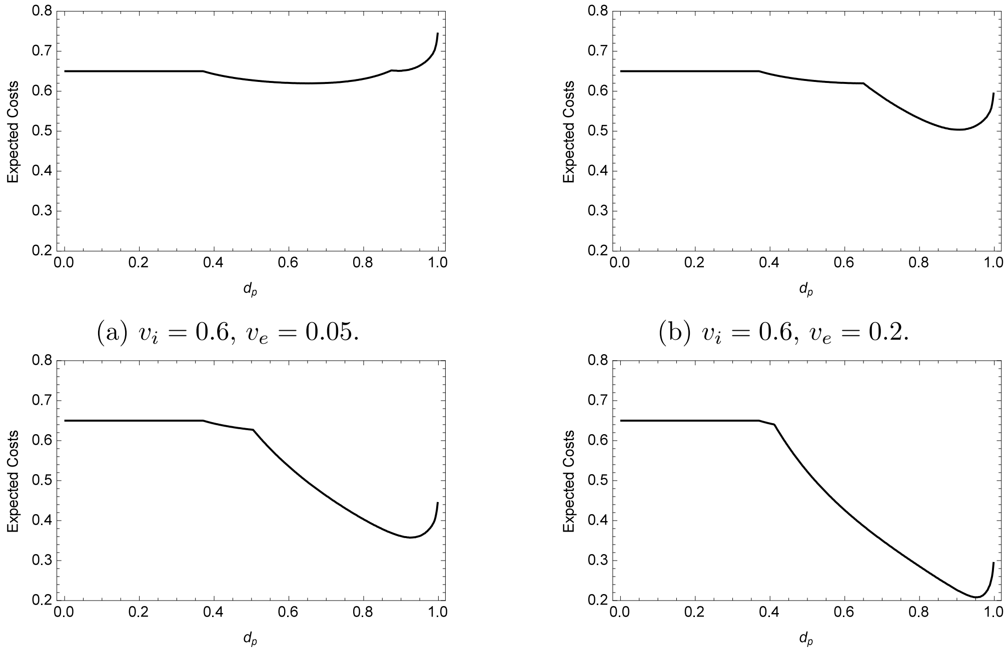


(d) $v_i = 0.6$, $v_e = 0.5$.

Figure 7.2: Expected costs ($\mathbb{E}[C(d_p)]$) depending on postponed volume (d_p) for $v_i \neq v_e$, $Q_e \sim Beta(1,1)$, $\bar{q}_e = 0.5$, $q_i = 0.8$ and $M = 25$.

7.1.2 Extension I: Uncertain costs

We have assumed that the buyer knows the suppliers' production costs c_i and c_e. Although this assumption may hold when the buyer has accurate estimates of the suppliers' cost structures, buyers often lack this information. Do our previous results hold when the buyer does not have perfect information about the suppliers' costs? Clearly, such cost uncertainty does not impact the results presented in Section 3.3.1, as what the buyer learns about the entrant's quality is independent of the entrant's costs. However, uncertainty about costs does affect the incumbent's terminal bids because the bidding condition in Eq. (3.8) is directly impacted by the costs of supplier j. Loosely speaking, supplier j's margin (LHS of Eq. (3.8)) becomes a random variable, and its realization depends on supplier j's true costs. Because the suppliers' bids are interrelated, the terminal bids are also random variables, and their distributions depend on the joint distribution of both suppliers' costs. Because we assume a risk-neutral buyer, it is sufficient to characterize how the expected terminal bids depend on d_p (similar to Proposition 7 for



(a) $v_i = 0.6$, $v_e = 0.05$.

(b) $v_i = 0.6$, $v_e = 0.2$.

(c) $v_i = 0.6$, $v_e = 0.35$.

(d) $v_i = 0.6$, $v_e = 0.5$.

Figure 7.3: Expected costs ($\mathbb{E}[C(d_p)]$) depending on postponed volume (d_p) for $v_i \neq v_e$, $Q_e \sim Beta(2,2)$, $\bar{q}_e = 0.5$, $q_i = 0.8$ and $M = 25$.

the case of known costs), but such characterization can be achieved only under highly restrictive and not very realistic assumptions about the suppliers' cost distributions. More general and realistic assumptions about these distributions render our model analytically intractable. Despite these challenges, we want to determine how uncertain costs impact competition and the buyer's optimal postponement decision, so we proceed as follows: First, we assume a simple cost distribution that allows us to obtain structural results that correspond to Proposition 7. Then we explain how the expected terminal bids can be determined numerically under realistic assumptions about the suppliers' cost distributions. Based on a numerical example, we show that, even under more realistic assumptions, we obtain structural results that are similar to those in Section 3.3.2. The optimal postponement decision can be determined in a similar fashion as described in Section 3.3.3.

Proposition 10.
Assume that suppliers' costs C_j, $j \in \{i, e\}$, are random variables distributed according to a discrete distribution such that with probability p we observe the case $C_e = c_e' = \overline{q}_e$ and $C_i = c_i' < q_i$, and with probability $1-p$ we observe the case $c_e = c_e' < \overline{q}_e$ and $c_i = c_i' = q_i$.

Then $\mathbb{E}[b_i^T] = pb_i^T(d_p, c_i < q_i, c_e = \overline{q}_e) + (1-p)q_i$ *and* $\mathbb{E}[b_e^T] = p\overline{q}_e + (1-p)b_e^T(d_p, c_i = q_i, c_e < \overline{q}_e)$.

a) *There exists a threshold level* $\overline{d}_p^j$, $j \in \{i, e\}$, *with* $\underline{d}_p^j < \overline{d}_p^j < 1$. *For* $\underline{d}_p^j < d_p \leq \overline{d}_p^j$, $\mathbb{E}[b_j^T]$ *is strictly decreasing in* d_p *while for* $\overline{d}_p^j < d_p \leq 1$, $\mathbb{E}[b_j^T]$ *is strictly increasing in* d_p.

b) *The threshold* $\overline{d}_p^j$ *is strictly decreasing in* v_j *and strictly increasing in* M.

Proposition 10 considers only two potential realizations of the uncertain suppliers' costs with corresponding probabilities p and $1-p$. As a result, b_j^T can have two outcomes, where the expected terminal bids $\mathbb{E}[b_j^T]$ are computed as their weighted sums.

The results presented in Proposition 10 are similar to those established in Proposition 7 for known costs. Suppliers do not lower their bids for postponed volumes up to a threshold level $\underline{d}_p^j$ and reduce their bids up to the threshold level $\overline{d}_p^j$. The only difference from our previous results is that now the bids of *both* suppliers can be increasing in the postponed volume above the threshold $\overline{d}_p^j$ because the expected terminal bids are weighted averages of all possible outcomes for the realizations of the costs C_i and C_e. For each supplier, there is one instance, ($v_j > 0$ and $v_{-j} = 0$), in which the supplier's bid is increasing for high postponed volumes $\overline{d}_p^j < d_p \leq 1$, while in the other instance, $v_j = 0$ and $v_{-j} > 0$, the supplier's bid is constant. Because the expected bid is simply a weighted average, both suppliers' expected bids are increasing for high postponed volumes.

Of course, the question arises concerning whether these results also hold if we make less restrictive and more realistic assumptions about the buyer's information on the suppliers' costs. In many practical instances, a buyer has access to some information about the suppliers' cost structures that allows the buyer to specify a lower bound $\underline{c}_j$ and an upper bound $\overline{c}_j$ of the costs of supplier j and some distribution $f(c_j)$ with non-negative support $[\underline{c}_j, \overline{c}_j]$.[1] In this case, the calculation of the supplier j's expected terminal bids becomes more complex:

$$\mathbb{E}[b_j^T] = \int_{\underline{c}_j}^{\overline{c}_j} \int_{\underline{c}_{-j}}^{\overline{c}_{-j}} b_j^T(d_p, c_j, c_{-j}) f(c_j) f(c_{-j}) dc_j dc_{-j} \text{ for } j \in \{i, e\}. \tag{7.1}$$

Thus, any combination of the two suppliers' cost realizations yields a specific terminal bid, and these terminal bids are weighted with the corresponding joint probability. While we provide no analytical solutions to Eq. (7.1), the underlying logic remains the same as

[1]Absent more accurate information, we could assume that $C_j \sim U(\underline{c}_j, \overline{c}_j)$.

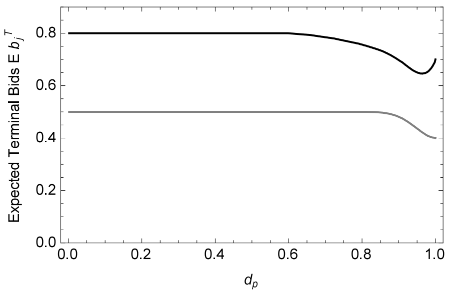
that for Proposition 10, and we can evaluate Eq. (7.1) numerically. Figure 7.4 plots the expected terminal bids, depending on d_p, for different cases of uniform costs (from highly unequal expected supplier values in Figure 7.4a to equal expected values in Figure 7.4d). Comparing these numerical results to the results in Figure 7.1 shows that, structurally, suppliers' (expected) terminal bids are the same as they are when their costs are known. For low postponed volumes, neither supplier lowers its bid, but higher postponed volumes provide sufficient incentive to lower the bids, and there exists an upper threshold past which a supplier's bid can increase because he anticipates the buyer's quality-updating process and expects to win the contract. Whether the bid increases for high postponed volumes is determined by the weighting of the number of instances, similar to our simple example above.

These analyses indicate that our structural insights continue to hold when the buyer faces uncertainty about the suppliers' costs. As a result, the rationale that underlies the optimal postponement tender that we presented in Section 3.3.3 remains the same: On the one hand, the buyer can choose a learning strategy and postpone d_p^Q, an option that remains unchanged because it is not affected by cost uncertainty; on the other hand, the buyer can choose a competition strategy and postpone a high volume $d_p^{C'}$ that corresponds to the cost minimum (similar to d_p^C in Proposition 8). Although we cannot provide a formal characterization of the buyer's optimal solution, the intuition remains the same as that presented in conjunction with Proposition 9: The size of the procurement volume M and the (expected) supplier values $\mathbb{E}[v_j]$ determine whether the learning strategy or the competition strategy is optimal. Thus, when the suppliers' costs are uncertain and the buyer can specify a cost distribution, she can still rely on our model to determine the optimal postponement tender numerically.

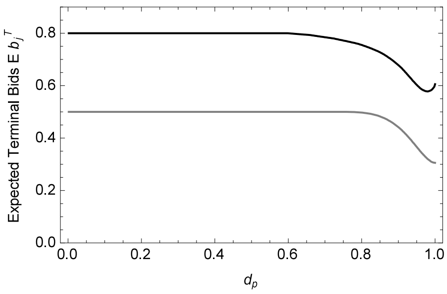
7.1.3 Extension II: Unequal initial volume shares

We have assumed that the buyer allocates the initial volume $(1 - d_p)M$ equally between the incumbent and the entrant, but a buyer may be interested in allocating unequal volume shares. This section shows that increasing or decreasing the initial volume share awarded to the entrant does not change the results of our analysis structurally and that increasing the initial volume share of the entrant impacts our results in the same way that increasing M does.

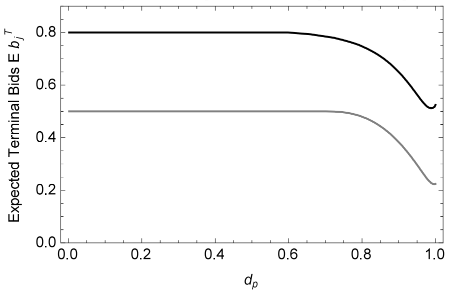
Suppose the buyer allocates $\gamma(1 - d_p)M$ units of the initial volume to the entrant and $(1 - \gamma)(1 - d_p)M$ units to the incumbent, with $\gamma \in [0; 1]$. That is, $\gamma > 0.5$ indicates that the entrant receives a larger share of the initial volume. We first investigate how



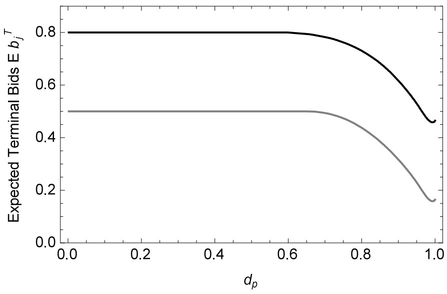
(a) $\mathbb{E}[v_i] = 0.4$, $\mathbb{E}[v_e] = 0.1$,
$C_i \sim U[0.2, 0.6]$, $C_e \sim U[0.3, 0.5]$.



(b) $\mathbb{E}[v_i] = 0.4$, $\mathbb{E}[v_e] = 0.2$,
$C_i \sim U[0.2, 0.6]$, $C_e \sim U[0.2, 0.4]$.



(c) $\mathbb{E}[v_i] = 0.4$, $\mathbb{E}[v_e] = 0.3$,
$C_i \sim U[0.2, 0.6]$, $C_e \sim U[0.1, 0.3]$.



(d) $\mathbb{E}[v_i] = 0.4$, $\mathbb{E}[v_e] = 0.4$,
$C_i \sim U[0.2, 0.6]$, $C_e \sim U[0, 0.2]$.

Figure 7.4: Expected Terminal bids ($\mathbb{E}[b_j^T]$) of incumbent (black) and entrant (gray) depending on postponed volume (d_p) for uniformly distributed costs, $Q_e \sim Beta(1,1)$, $\bar{q}_e = 0.5$, $q_i = 0.8$ and $M = 25$.

the initial volume share impacts the buyer's learning and then the impact of the share on competition. To better understand the impact on the buyer's learning, recall the posterior distribution of the entrant's quality, derived in Proposition 2, modified for unequal initial shares:

$$Q_e^{post} \sim Beta(\alpha + \hat{q}_e \gamma(1 - d_p)M, \beta + (1 - \hat{q}_e)\gamma(1 - d_p)M). \tag{7.2}$$

Eq. (7.2) shows that, if the buyer allocates more of the initial volume to the entrant (i.e., $\gamma > 0.5$), learning increases and the buyer can make a more accurate estimate of the entrant's quality distribution than it could if it allocated an equal share (and vice versa if she allocates less to the entrant). This effect is structurally the same as increasing M (see our discussion of Proposition 2): observing more initial volume results in higher learning.

The effect of unequal initial volume shares on competition is more difficult to explain. We begin by formulating the bidding conditions, modified for unequal shares. For the incumbent the condition is:

$$b_i^{t-1} - c_i > \underbrace{\delta \frac{(1-\gamma)\frac{1-d_p}{d_p} + P(b_i^{t-1} - \delta | d_p)}{P(b_i^{t-1} - \delta | d_p) - P(b_i^{t-1} | d_p)}}_{\text{incumbent's markup} := \theta_i^t}, \quad \text{for } 0 < d_p \leq 1,\ t \in \{1, \dots, T\}, \tag{7.3}$$

and the entrant's condition is

$$b_e^{t-1} - c_e > \underbrace{\delta \frac{\gamma\frac{1-d_p}{d_p} + P(b_e^{t-1} - \delta | d_p)}{P(b_e^{t-1} - \delta | d_p) - P(b_e^{t-1} | d_p)}}_{\text{entrant's markup} := \theta_e^t}, \quad \text{for } 0 < d_p \leq 1,\ t \in \{1, \dots, T\}. \tag{7.4}$$

In conjunction with Proposition 5, we referred to the relative size of the postponed volume $\frac{1-d_p}{d_p}$ as the volume incentive and the probabilities of winning (i.e., $P(\cdot)$-terms) as the risk incentive. Clearly, splitting the initial volume allocation affects both the volume incentive and the risk incentive in the suppliers' bidding conditions. The effect on the volume incentive tends to be intuitive: a supplier's incentive to decrease his bid decreases if he receives a larger share of the initial volume. For example, if the entrant receives the larger share (i.e., $\gamma > 0.5$), the entrant's markup increases (see Eq. (7.4)) because the first term in the numerator increases in the initial volume share, and if the incumbent receives the smaller share $(1-\gamma)$, his markup decreases (see Eq. (7.3)), increasing his incentive to quote lower bids.

The effect of unequal initial volume shares on the risk incentive is more difficult to explain. Similar to our discussions regarding the risk incentive in conjunction with Proposition 5, we cannot tie changes in γ directly to the risk incentive. However, we know that the risk incentive, represented by the probabilities of winning the postponed volume $P(\cdot)$ in the bidding conditions, is connected to the buyer's learning. As we explained, increasing (decreasing) the initial volume share increases (decreases) the buyer's learning, similar to increasing or decreasing M, because both have the same impact on the updated quality distribution in Eq. (7.2). Therefore, we conclude that a higher (lower) share of the initial volume allocated to the entrant has the same structural impact on the risk incentive as a higher (lower) M; that is, the findings in Proposition 6 and Proposition 7 also apply to increasing values of γ: the lower threshold $\underline{d}_p^j$ and the upper threshold $\overline{d}_p^j$ of both suppliers are strictly increasing if the entrant receives a higher share of the initial volume. The competition interval for which suppliers decrease their bids decreases with the entrant's share of the initial volume γ. Figure 7.5, which shows numerical results for various initial volume shares, confirms this intuition: For large shares γ allocated to the entrant, the interval in which suppliers decrease their bids decreases.

In summary, the initial split, whether even or uneven, does not change the structure of our results. (Compare Figure 7.5 to the results in Section 3.3.2.) However, the initial split does provide the buyer with a second lever with which to impact learning and competition.

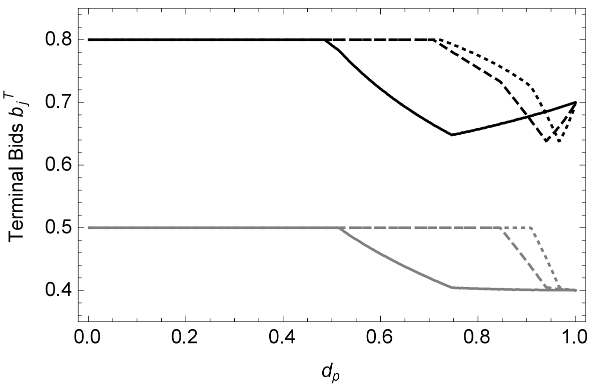


Figure 7.5: Terminal bids b_i^T (black) and b_e^T (gray) for $v_i = 0.3 > v_e = 0.1$, $M = 10$ and $Q_e \sim Beta(1, 1)$. $\gamma = 0.1$ (solid), $\gamma = 0.5$ (dashed) and $\gamma = 0.9$ (dotted).

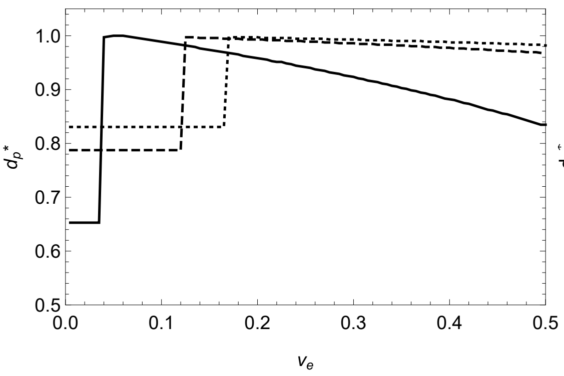
7.1.4 Extension III: Minimum initial volumes

Our results have shown that, especially for large buyers, it may be attractive to award small initial volumes (basically test lots) and to postpone almost the entire procurement volume. However, in practice, doing so may not be feasible. Practitioners who employ mechanisms similar to a postponement tender award larger initial volumes (e.g., Global Fund (2014a)) than the small initial volumes suggested by our approach, perhaps to incent the entrant to make a substantial capacity investment. Small initial volumes and the risk of not winning the postponed volume may prevent the entrant from investing in increasing his capacity. Another important reason for a buyer's being hesitant to award test lots emerged from our discussions with global-health experts, who indicated that suppliers may have an incentive to supply high-quality lab-produced test lots that have no relationship with the quality of the remaining volume that is produced on a larger scale. To avoid these issues, the buyer may be inclined to choose a relatively high initial volume. In this section, we assume that the buyer can determine a minimum initial volume that ensures that the entrant and incumbent partake in the tender and do not provide test lots whose quality is not matched in the postponed volume. We include the minimum initial volume as a constraint in our model and study its impact on the optimal strategy.

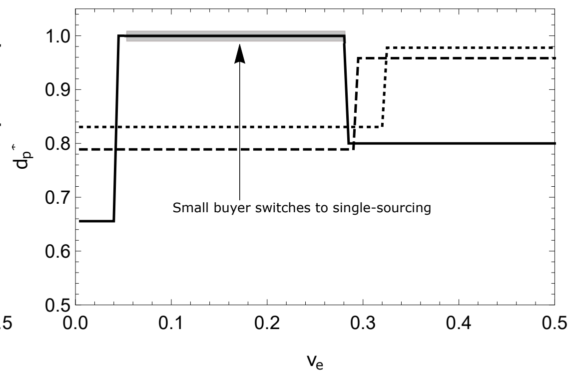
Let $o_{min} \leq M$ denote the initial volume each supplier must receive, such that $\frac{1}{2}(1-d_p)M \geq o_{min}$. In this case, the buyer's feasible postponed volumes are restricted to $d_p \in [0; 1-2\frac{o_{min}}{M}] \cup \{1\}$, where $d_p \in [0; 1-2\frac{o_{min}}{M}]$ is the interval that satisfies the minimum initial volume condition, and $d_p = 1$ is the single-sourcing case in which one supplier always receives at least the minimum initial volume o_{min}, and the other supplier receives no allocation.

Figure 3.10 shows how a minimum initial volume changes the buyer's optimal sourcing strategies. Because the minimum initial volume restricts the buyer's postponement options, the optimal strategy is now strictly divided into the three possible strategies: while the learning strategy (d_p^Q) and the single-sourcing strategy ($d_p^C = 1$) derived in Section 3.3.3 do not change, the competition strategy with a high postponement volume ($d_p^C < 1$) is affected by the restriction on the initial volume (for $M = 50$ and $M = 90$ there exists no single-sourcing optimum, because $v_e > v_i$ for all cases where $d_p^* = d_p^C$). Therefore, the optimal strategy is to either to single-source or to postpone a volume such that the initial volumes are equal to the minimum initial volume o_{min}. This has a distinctive effect on the optimal strategy for both small buyers and large buyers. The results in Figure 7.6b show that a small supplier should choose to single-source when the minimum initial volume increases; the explanation for this effect is that, as o_{min} grows, the volume that can be postponed (i.e., $M - 2 * o_{min}$) declines, either making it less attractive to postpone because the learning effects are low (Figures 7.6a and 7.6b) or rendering postponement infeasible (Figures 7.6b and 7.6c). The minimum initial volume has the opposite effect on larger buyers' optimal strategy, as larger minimum initial volumes (Figures 7.6c and 7.6d) do not allow the buyer to choose d_p close to 1 (see Figure 7.6a). At a certain minimum initial volume, a larger postponed volume should always be chosen, regardless of differences in suppliers' values (Figure 7.6d). The effect of the minimum initial volume on larger buyers' optimal strategy is driven by the large learning opportunities that are associated with a large M.

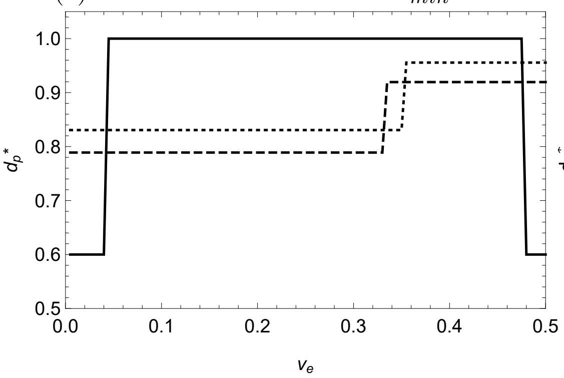
Clearly, the effect of a minimal initial volume on the optimal strategy differs with the buyer's size, which seems counter-intuitive. A comparison of Figures 7.6a and 7.6d reveals that the effect of the minimum initial volume for small buyers is opposite the effect for large buyers. At first glance, one would assume that a large buyer should switch from low initial volumes (i.e., test lots) to single-sourcing because a low initial volume represents only a small deviation from the optimal postponed volume d_p^*. However, in doing so, the buyer would forego all learning benefits and her overall utility would decrease sharply. Because of the high volume of the large supplier the learning benefits are so strong that it is optimal to deviate substantially from the optimal (unconstrained) postponed volume and to choose instead a comparatively low postponed volume than to single source. By contrast, a small buyer can afford to forego the benefits of learning, which are comparatively low because of her small volume. Instead of offering an initial volume, then, the small buyer should exploit the benefits of competition and opt for single-sourcing. From a managerial point of view, this finding means that a minimum initial volume, which restricts the action space, effectively reverses the optimal strategies for the different types of buyers, and decision-makers should be careful when facing these restrictions in allocating volume.



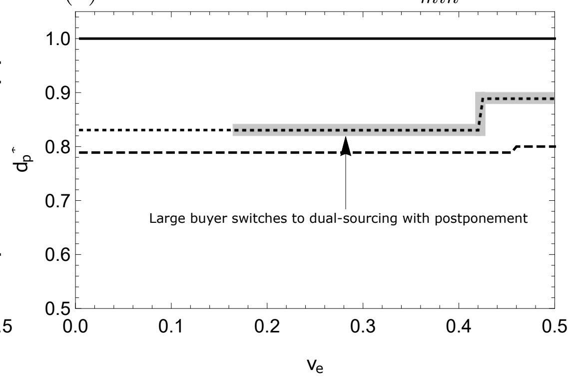
(a) Minimum initial volumes $o_{min} = 0$.



(b) Minimum initial volumes $o_{min} = 1$.



(c) Minimum initial volumes $o_{min} = 2$.



(d) Minimum initial volumes $o_{min} = 5$.

Figure 7.6: Optimal postponement d_p^* dependent on v_e for $v_i = 0.05$, $Q_e \sim Beta(1, 1)$ and procurement volumes $M = 10$ (solid), $M = 50$ (dashed) and $M = 90$ (dotted).

7.2 Appendix II

Proof of Proposition 2

The result in a) follows directly from the characteristics of the Beta-distribution. If one conducts a single Bernoulli trial with a prior distribution $Beta(\alpha, \beta)$ and this trial is successful, the resulting (posterior) Beta-distribution is $Beta(\alpha+1, \beta)$ and if the trial is unsuccessful the resulting (posterior) Beta-distribution is $Beta(\alpha, \beta+1)$ (Gupta and Nadarajah (2004); Tomlin (2009); Zhu and Lu (2004)). In our case the buyer observes $\hat{q}_e \frac{1}{2}(1-d_p)M$ successful trials and $(1-\hat{q}_e)\frac{1}{2}(1-d_p)M$ unsuccessful trials; as a result the posterior distribution of the entrant's quality is distributed according to $Q_e^{post} \sim$Beta$(\alpha + \hat{q}_e\frac{1}{2}(1-d_p)M, \beta + (1-\hat{q}_e)\frac{1}{2}(1-d_p)M))$. The result in b) follows by calculating the expected posterior quality using the probability distribution from a):

$$\bar{q}_e^{post}(\hat{q}_e, d_p) = \frac{\alpha + \hat{q}_e\frac{1}{2}(1-d_p)M}{\alpha+\beta+\frac{1}{2}(1-d_p)M}. \tag{7.5}$$

We defined in Proposition 2 c) expected learning as

$$\mathbb{E}[\Delta(d_p)] := \mathbb{E}[|\bar{q}_e^{post}(\hat{Q}_e, d_p) - \bar{q}_e|] = \int_0^1 |\bar{q}_e^{post}(\hat{q}_e, d_p) - \bar{q}_e| f(\hat{q}_e)\ \mathrm{d}\hat{q}_e. \tag{7.6}$$

with $f(q_e) \sim Beta(\alpha, \beta)$ being the a priori quality density function. Because

$$\frac{\partial \bar{q}_e^{post}(\hat{q}_e, d_p)}{\partial \hat{q}_e} = \frac{M(1-d_p)}{2(\alpha+\beta)+M(1-d_p)} > 0, \tag{7.7}$$

$\bar{q}_e^{post}(\hat{q}_e, d_p)$ is strictly increasing in $\hat{q}_e$ for $d_p < 1$. Using the fact that $\bar{q}_e^{post}(\hat{q}_e, d_p)$ is continuous in $\hat{q}_e$ and $\bar{q}_e^{post}(\bar{q}_e, d_p) = \bar{q}_e$, we can write (7.6) as

$$\begin{aligned}\mathbb{E}[|\bar{q}_e^{post}(\hat{Q}_e, d_p) - \bar{q}_e|] &= \int_0^{\bar{q}_e} (\bar{q}_e - \bar{q}_e^{post}(\hat{q}_e, d_p)) f(\hat{q}_e)\ \mathrm{d}\hat{q}_e + \int_{\bar{q}_e}^1 (\bar{q}_e^{post}(\hat{q}_e, d_p) - \bar{q}_e) f(\hat{q}_e)\ \mathrm{d}\hat{q}_e \\ &= \int_0^{\bar{q}_e} (\bar{q}_e - \frac{\alpha+\hat{q}_e\frac{1-d_p}{2}M}{\alpha+\beta+\frac{1-d_p}{2}M}) f(\hat{q}_e)\ \mathrm{d}\hat{q}_e + \int_{\bar{q}_e}^1 (\frac{\alpha+\hat{q}_e\frac{1-d_p}{2}M}{\alpha+\beta+\frac{1-d_p}{2}M} - \bar{q}_e) f(\hat{q}_e)\ \mathrm{d}\hat{q}_e\end{aligned}$$

Differentiation of $\mathbb{E}[|q_e^{post}(\hat{q}_e, d_p) - \bar{q}_e|]$ w.r.t. d_p yields:

$$\frac{\partial \mathbb{E}[\Delta(d_p)]}{\partial d_p} = -\int_0^{\bar{q}_e} \frac{2M(\alpha-(\alpha+\beta)\hat{q}_e)}{(2(\alpha+\beta)+M(1-d_p))^2} f(\hat{q}_e)\ \mathrm{d}\hat{q}_e + \int_{\bar{q}_e}^1 \frac{2M(\alpha-(\alpha+\beta)\hat{q}_e)}{(2(\alpha+\beta)+M(1-d_p))^2} f(\hat{q}_e)\ \mathrm{d}\hat{q}_e \tag{7.8}$$

The first term evaluates realizations $0 < \hat{q}_e < \bar{q}_e$ and is negative because $\alpha-(\alpha+\beta)\hat{q}_e > 0$ for $\hat{q}_e < \frac{\alpha}{\alpha+\beta} = \bar{q}_e$, and the second term (evaluating realizations $\bar{q}_e < \hat{q}_e < 1$) is also negative because $\alpha-(\alpha+\beta)\hat{q}_e < 0$ for $\hat{q}_e > \frac{\alpha}{\alpha+\beta} = \bar{q}_e$. Therefore both terms in Eq. (7.8) are negative and $\frac{\partial \mathbb{E}[\Delta(d_p)]}{\partial d_p} < 0$ (strictly decreasing).

The second derivative of $\mathbb{E}[|\bar{q}_e^{post}(\hat{Q}_e, d_p) - \bar{q}_e|]$ w.r.t. d_p:

$$\frac{\partial^2 \mathbb{E}[\Delta(d_p)]}{\partial^2 d_p} = -\int_0^{\bar{q}_e} \frac{4M^2(\alpha-(\alpha+\beta)\hat{q}_e)}{(2(\alpha+\beta)+M(1-d_p))^3} f(\hat{q}_e)\ \mathrm{d}\hat{q}_e + \int_{\bar{q}_e}^1 \frac{4M^2(\alpha-(\alpha+\beta)\hat{q}_e)}{(2(\alpha+\beta)+M(1-d_p))^3} f(\hat{q}_e)\ \mathrm{d}\hat{q}_e \tag{7.9}$$

For the same reason as before, both terms in Eq. (7.9) are negative and $\frac{\partial^2 \mathbb{E}[\Delta(d_p)]}{\partial^2 d_p} < 0$ (concave). From the above the result in Proposition 2c follows: $\mathbb{E}[|\bar{q}_e^{post}(\hat{Q}_e, d_p) - \bar{q}_e|]$ is concave decreasing in d_p.

Differentiation of $\mathbb{E}[|q_e^{post}(\hat{q}_e, d_p) - \bar{q}_e|]$ w.r.t. M yields:

$$\frac{\partial \mathbb{E}[\Delta(d_p)]}{\partial M} = \int_0^{\bar{q}_e} \frac{2(1-d_p)(\alpha-(\alpha+\beta)\hat{q}_e)}{(2(\alpha+\beta)+M(1-d_p))^2} f(\hat{q}_e)\ \mathrm{d}\hat{q}_e - \int_{\bar{q}_e}^1 \frac{2(1-d_p)(\alpha-(\alpha+\beta)\hat{q}_e)}{(2(\alpha+\beta)+M(1-d_p))^2} f(\hat{q}_e)\ \mathrm{d}\hat{q}_e \tag{7.10}$$

For the same reason as in Eq. (7.8) and (7.9) both terms in Eq. (7.10) are positive. Therefore, $\frac{\partial \mathbb{E}[\Delta(d_p)]}{\partial M} > 0$ (strictly increasing) and $\frac{\partial^2 \mathbb{E}[\Delta(d_p)]}{\partial^2 dM} > 0$ (convex). From the above the result in Proposition 2c follows: $\mathbb{E}[|\bar{q}_e^{post}(\hat{Q}_e, d_p) - \bar{q}_e|]$ is concave increasing in M. □

Proof of Proposition 3

In the following we assume that $Q_e \sim Beta(1,1)$. We defined Eq. (3.3) as:

$$\mathbb{E}[Q(d_p)] = (1-d_p)M \cdot \left(\frac{q_i}{2} + \frac{\mathbb{E}[Q_e]}{2}\right) + d_p M \cdot \int_0^1 f(\hat{q}_e) V_q(\hat{q}_e, d_p) \mathrm{d}\hat{q}_e$$
$$\text{where } V_q(\hat{q}_e, d_p) = \begin{cases} q_i, & \text{for } q_i > \bar{q}_e^{post}(\hat{q}_e, d_p) \\ \bar{q}_e^{post}(\hat{q}_e, d_p), & \text{for } q_i < \bar{q}_e^{post}(\hat{q}_e, d_p) \end{cases}. \tag{7.11}$$

The proof of Proposition 2c shows that $\bar{q}_e^{post}(\hat{q}_e, d_p)$ is strictly increasing in $\hat{q}_e$. This indicates that there exists a unique quality observation $x^* \in [0,1]$ such that

$$\max\{\bar{q}_e^{post}(\hat{q}_e, d_p), q_i\} = \begin{cases} \bar{q}_e^{post}(\hat{q}_e, d_p) & \hat{q}_e \geq x^* \\ q_i & \hat{q}_e < x^* \end{cases}. \tag{7.12}$$

Solving $\overline{q}_e^{post}(x^*, d_p) \overset{!}{=} q_i$ results in $x^* = \frac{(4+(1-d_p)M)q_i - 2}{(1-d_p)M}$ and we obtain

$$\begin{aligned}\mathbb{E}[Q(d_p)] = (1-d_p)(\tfrac{1}{2}q_i + \tfrac{1}{2}\overline{q}_e)M \\ + d_pM\Big(\int_0^{x^*} q_i f(\hat{q}_e)\ \mathrm{d}\hat{q}_e + \int_{x^*}^1 \overline{q}_e^{post}(\hat{q}_e, d_p) f(\hat{q}_e)\ \mathrm{d}\hat{q}_e\Big).\end{aligned} \tag{7.13}$$

With $q_i = \bar{q}_e$ it holds that $x^* = \overline{q}_e$ (if suppliers are a priori equal, they are a posteriori equal if the update equals q_i) and Eq. (7.13) is reduced to

$$\mathbb{E}[Q(d_p)] = (1-d_p)\overline{q}_e M + d_pM\Big(\int_0^{\overline{q}_e} \overline{q}_e f(\hat{q}_e)\ \mathrm{d}\hat{q}_e + \int_{\overline{q}_e}^1 \overline{q}_e^{post}(\hat{q}_e, d_p) f(\hat{q}_e)\ \mathrm{d}\hat{q}_e\Big) \tag{7.14}$$

With $Q_e \sim Beta(1,1)$ (that is quality is uniformly distributed on the interval $[0,1]$) we know $\bar{q}_e = \frac{1}{2}$ and $f(q_e) = 1$. This yields

$$\mathbb{E}[Q(d_p)] = (1-d_p)\tfrac{1}{2}M + d_pM\Big(\int_0^{\frac{1}{2}} \tfrac{1}{2}\ \mathrm{d}\hat{q}_e + \int_{\frac{1}{2}}^1 \overline{q}_e^{post}(\hat{q}_e, d_p)\ \mathrm{d}\hat{q}_e\Big) \tag{7.15}$$

$$= \frac{M(16+(1-d_p)(4+d_p)M)}{32+8(1-d_p)M}. \tag{7.16}$$

Differentiation of Eq. (7.16) w.r.t. d_p yields

$$\frac{\partial\mathbb{E}[Q(d_p)]}{\partial d_p} = \frac{M^2(4+M+d_p(-8+(-2+d_p)M))}{8(4+M(1-d_p))^2} \overset{!}{=} 0 \leftrightarrow d_p^Q = \frac{1}{1+\frac{2}{\sqrt{4+M}}}$$

$$\frac{\partial^2\mathbb{E}[Q(d_p)]}{\partial d_p^2} = -\frac{M^2(4+M)}{(4+(1-d_p)M)^3} < 0.$$

Therefore the expected quality $\mathbb{E}[Q(d_p)]$ in Eq. (7.11) is concave with a maximum at d_p^Q. Differentiation of d_p^Q w.r.t. M yields

$$\frac{\partial d_p^Q}{\partial M} = \frac{1}{\sqrt{4+M}(2+\sqrt{4+M})^2} > 0 \tag{7.17}$$

□

Proof of Proposition 4

Assume that $Q_e \sim Beta(1,1)$ and $q_i \neq \overline{q}_e = \frac{1}{2}$. Our proof builds on the following observation: we know from our prior analyses that large postponed volumes only allow the buyer to learn little about the entrant's quality. If the a prior quality between incumbent and entrant differs there are cases of large postponed volumes in which the buyer will never receive a sufficiently strong update to change the allocation decision. First, we are interested in identifying the postponed volume $\tilde{d}_p$ that separates the interval where the buyer will never change her allocation decision from the interval where she may change her allocation decision. To do so, we have to separate the case where the incumbent has a priori a higher quality than the entrant from the case where the entrant as (a priori) a higher quality than the incumbent. Second, once we have identified $\tilde{d}_p$ we will characterize the buyer's expected quality depending on d_p.

There exists a postponed volume $\tilde{d}_p \in [0,1]$ such that

$$\max\{\overline{q}_e^{post}(\hat{q}_e, d_p), q_i\}_{d_p > \tilde{d}_p} = \begin{cases} \overline{q}_e^{post}(\hat{q}_e, d_p), & \forall \hat{q}_e \in [0,1] \text{ for } \overline{q}_e > q_i \\ q_i, & \forall \hat{q}_e \in [0,1] \text{ for } \overline{q}_e < q_i \end{cases}. \quad (7.18)$$

To characterize $\tilde{d}_p$ we want to identify the postponed volume where even the worst (best) update does not change the allocation decision. That is $\tilde{d}_p$ is the solution to

$$\overline{q}_e^{post}(\hat{q}_e = 0, \tilde{d}_p) \stackrel{!}{=} q_i \rightarrow \tilde{d}_p = \frac{M - 8(\overline{q}_e - q_i) - 2M(\overline{q}_e - q_i)}{M(1 - 2(\overline{q}_e - q_i))} \text{ for } \overline{q}_e > q_i \quad (7.19)$$

and

$$\overline{q}_e^{post}(\hat{q}_e = 1, \tilde{d}_p) \stackrel{!}{=} q_i \rightarrow \tilde{d}_p = \frac{M - 8(q_i - \overline{q}_e) - 2M(q_i - \overline{q}_e)}{M(1 - 2(q_i - \overline{q}_e))} \text{ for } \overline{q}_e < q_i. \quad (7.20)$$

Therefore, there exists an interval $d_p \in (\tilde{d}_p, 1]$ for which the posteriori quality of the entrant is either always lower or always higher than the incumbent's quality, and the buyer will always choose to allocate the postponed volume to the a priori better supplier despite the outcome of learning. With this information we can write the expected quality from Eq. (7.11) as

$$\mathbb{E}[Q(d_p)]_{d_p > \tilde{d}_p} = \begin{cases} (1 - d_p)(\frac{1}{2}q_i + \frac{1}{2}\overline{q}_e)M + d_p M \overline{q}_e, & \text{for } \overline{q}_e > q_i \\ (1 - d_p)(\frac{1}{2}q_i + \frac{1}{2}\overline{q}_e)M + d_p M q_i, & \text{for } \overline{q}_e < q_i \end{cases} \quad (7.21)$$

and differentiation w.r.t. d_p yields

$$\frac{\partial \mathbb{E}[Q(d_p)]_{d_p > \tilde{d}_p}}{\partial d_p} = \begin{cases} \frac{1}{2}(\overline{q}_e - q_i)M, & \text{for } \overline{q}_e > q_i \\ \frac{1}{2}(q_i - \overline{q}_e)M, & \text{for } \overline{q}_e < q_i \end{cases}, \quad (7.22)$$

which is positive in both cases. Hence $\mathbb{E}[Q(d_p)]$ is strictly increasing in d_p for $d_p > \tilde{d}_p$.

To simplify our presentation we define $\Delta_q = |\overline{q}_e - q_i|$ and we can write $\tilde{d}_p$ as

$$\tilde{d}_p = \frac{M-8\Delta_q-2M\Delta_q}{M(1-2\overline{q}_e)} \qquad \text{for } \overline{q}_e \neq q_i \tag{7.23}$$

and differentiation w.r.t. Δ_q yields

$$\frac{\partial \tilde{d}_p}{\partial \Delta_q} = -\frac{8}{M(1-2\Delta_q)^2} < 0. \tag{7.24}$$

□

Proof of Proposition 5

The bidding condition in Proposition 5 follows from algebraic manipulation of inequality (3.5) and (3.7). □

Proof of Proposition 6

Assume that $Q_e \sim Beta(1,1)$ and $\overline{v}_e = v_i$. It is further reasonable to assume that $v_j > \delta$, $j \in \{i, e\}$.

a) We want to find $\underline{d}_p$ for which both suppliers start to lower their bids. The probability for the incumbent to win the postponed volume, given bids b_i^{t-1} and b_e^{t-1} in bidding round $t-1$, can be calculated as

$$\begin{aligned}
P[q_i - b_i^{t-1} > \overline{q}_e^{post} - b_e^{t-1}] &= P[q_i - b_i^{t-1} + b_e^{t-1} > \overline{q}_e^{post}] && (7.25)\\
&= P[q_i - b_i^{t-1} + b_e^{t-1} > \frac{1+\hat{q}_e\frac{1-d_p}{2}M}{2+\frac{1-d_p}{2}M}] && (7.26)\\
&= P[\frac{(q_i-b_i^{t-1}+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M} > \hat{q}_e] && (7.27)\\
&= F(\frac{(q_i-b_i^{t-1}+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M}), && (7.28)
\end{aligned}$$

where $F(q_e) \sim Beta(1,1)$. The probability for the entrant to win the postponed volume is therefore simply

$$\begin{aligned}
P[\overline{q}_e^{post} - b_e^{t-1} > q_i - b_i^{t-1}] &= 1 - P[q_i - b_i^{t-1} > \overline{q}_e^{post} - b_e^{t-1}] && (7.29)\\
&= 1 - F(\frac{(q_i-b_i^{t-1}+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M}). && (7.30)
\end{aligned}$$

With these probabilities we can calculate the markups θ_i^{t-1} and θ_e^{t-1} from Proposition 5:

$$\theta_i^{t-1} = \delta \frac{\frac{1}{2}\frac{1-d_p}{d_p} + F(\frac{(q_i-(b_i^{t-1}-\delta)+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M})}{F(\frac{(q_i-(b_i^{t-1}-\delta)+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M}) - F(\frac{(q_i-b_i^{t-1}+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M})} \tag{7.31}$$

and

$$\theta_e^{t-1} = \delta \frac{\frac{1}{2}\frac{1-d_p}{d_p} + (1-F(\frac{(q_i-(b_i^{t-1}-\delta)+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M}))}{(1-F(\frac{(q_i-(b_i^{t-1}-\delta)+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M})) - (1-F(\frac{(q_i-b_i^{t-1}+b_e^{t-1})(2+\frac{1-d_p}{2}M)-1}{\frac{1-d_p}{2}M}))} \tag{7.32}$$

At time $t = 0$ suppliers bids are equal to the reservation prices, hence, $b_i^0 = q_i$ and $b_e^0 = \overline{q}_e$. With these starting bids, we can calculate the bidding condition (3.8) for both suppliers in the first round ($t = 1$):

$$v_i = q_i - c_i > \underbrace{\delta + \frac{(1-d_p)M}{2d_p(4+(1-d_p)M)}}_{\theta_i^1} \tag{7.33}$$

and

$$v_e = \frac{1}{2} - c_e > \underbrace{\delta + \frac{(1-d_p)M}{2d_p(4+(1-d_p)M)}}_{\theta_e^1} \tag{7.34}$$

We observe that $\theta_i^1 = \theta_e^1 = \theta_j^1$. Because $v_i = v_e$ both suppliers start to lower their bid for the same $\underline{d}_p$. To show that $\underline{d}_p \in (0;1]$ we observe

$$\frac{\partial \theta_j^1}{\partial d_p} = -\frac{M(4+(1-d_p)^2 M)}{2d_p^2(4+M(1-d_p))^2} < 0, \tag{7.35}$$

that is markups are both strictly decreasing in d_p. As v_j are constant there can only exists one $\underline{d}_p \in (0;1]$. It is simple to show that $\lim_{d_p \to 0^+} \theta_j = \infty$ and $\lim_{d_p \to 1^-} \theta_j = \delta$; therefore $\underline{d}_p$ always exists in $(0;1]$ if $v_j > \delta$. In summation suppliers decrease their bids below the reservation price for $\underline{d}_p < d_p < 1$, and for $d_p \leq \underline{d}_p$ suppliers do not lower their bids. □

b) We want to find the terminal bids of the incumbent and the entrant for $d_p > \underline{d}_p$.

In the proof of Proposition 6a we show that both suppliers start to lower their bids at the same $\underline{d}_p$ and the markups at the start of the auction ($t = 0$) are the same for the incumbent and the entrant ($\theta_i^1 = \theta_e^1$). The difference between the bids at the start of the auction is $b_e^0 - b_i^0 = \overline{q}_e - q_i$. Because suppliers' values are equal, i.e. $v_i = v_e$, the left-hand side of both suppliers' bidding condition (3.8) is equal. Therefore, after suppliers lowered their bid by δ in the second round, the difference in bids is still $b_e^1 - b_i^1 = \overline{q}_e - q_i$. Plugging this difference into the incumbent's and entrant's markups (7.31) and (7.32) gives us the markups at the second bidding round:

$$\theta_i^2 = \underbrace{\delta + \tfrac{(1-d_p)M}{2d_p(4+(1-d_p)M)}}_{\theta_i^1} \tag{7.36}$$

and

$$\theta_e^2 = \underbrace{\delta + \tfrac{(1-d_p)M}{2d_p(4+(1-d_p)M)}}_{\theta_e^1}. \tag{7.37}$$

Clearly, because the difference in bids remains the same, the markups for both suppliers are still the same ($\theta_i^2 = \theta_e^2$) and the suppliers have the same incentive to bid down. Additionally, the left hand side of the bidding conditions (3.8) is also the same for both suppliers after they have lowered their bid by δ ($v_i - \delta = v_e - \delta$ for $v_i = v_e$). Hence, in the case of equal values, in each round both suppliers have the same incentive to decrease their bids by δ, i.e. the difference in bids for equal values stays the same over the course of the auction. Therefore, it holds that $b_e^{t-1} - b_i^{t-1} = b_e^0 - b_i^0 = \overline{q}_e - q_i \; \forall t \in \{1, \ldots, T\}$ and we can calculate the terminal bids by solving the bidding conditions for each supplier for the terminal bid b_j^T such that

$$b_i^{T-1} - c_i \stackrel{!}{=} \underbrace{\delta + \tfrac{(1-d_p)M}{2d_p(4+(1-d_p)M)}}_{\theta_i^T} \tag{7.38}$$

and

$$b_e^{T-1} - c_e \stackrel{!}{=} \underbrace{\delta + \tfrac{(1-d_p)M}{2d_p(4+(1-d_p)M)}}_{\theta_e^T}. \tag{7.39}$$

Therefore, suppliers stop to lower their bid in round T and $b_j^T = b_j^{T-1}$. Solving Eq.

(7.38) and (7.39) gives us the terminal bids

$$b_i^T = c_i + \delta + \tfrac{1-d_p}{2d_p}\tfrac{M}{4+M(1-d_p)} \quad (7.40)$$

$$b_e^T = c_e + \delta + \tfrac{M(1-d_p)}{2d_p(4+M(1-d_p))}. \quad (7.41)$$

To show that these are decreasing in d_p we differentiate w.r.t. d_p:

$$\tfrac{\partial b_i^T}{\partial d_p} = \tfrac{\partial b_e^T}{\partial d_p} = -\tfrac{M(4+M(1-d_p)^2)}{2d_p^2(4+M(1-d_p)^2)} < 0. \quad (7.42)$$

□

c) In the proof for Proposition 6a we show that the left hand side (v_j) of the suppliers' bidding condition in Eq. (3.8) is constant, and the right hand side (θ_j^1) is strictly decreasing in d_p. We argued that if θ_j^1 is smaller than the right hand side the suppliers start to lower their bids and referred to this intersection point as $\underline{d}_p$. It is straightforward to see that $\underline{d}_p$ decreases in v_j.

To show that $\underline{d}_p$ is increasing in M we differentiate θ_j^1 w.r.t. M:

$$\tfrac{\partial \theta_j^1}{\partial M} = \tfrac{2(1-d_p)}{d_p(4+M(1-d_p))^2} > 0, \quad (7.43)$$

hence, θ_j is strictly increasing in M and the intersection point $\underline{d}_p$ between v_j and θ_j^1 is increasing in M. □

Proof of Corollary 1

This follows directly from Proposition 6a. If $v_j > v_{-j}$, $j \in \{i, e\}$, the left-hand side of supplier j's bidding condition in the first round is larger than that of supplier $-j$ (see Eq. (7.33) and (7.34)). Hence, supplier j will start to bid down for a smaller d_p than supplier $-j$. While the decrease of supplier j's bid will change supplier $-j$'s right-hand side of the bidding condition (θ_{-j}^t), because there exists a $\lambda > 0$ such that $v_j + \lambda = v_{-j}$, there also exists an $\epsilon > 0$ such that $\underline{d}_p^j + \epsilon = \underline{d}_p^{-j}$. In other words we can always find an $\epsilon > 0$ such that supplier j will lower his bid by δ while supplier will not. □

Proof of Proposition 7

a) Without loss of generality we assume that $v_i > v_e = 0$. The entrant e has no margin and therefore can not lower his bid, hence, $b_e^T = \bar{q}_e \; \forall d_p \in [0; 1]$.

We insert $b_e^T = \bar{q}_e$ into i's bidding condition and get the condition for the terminal bid b_i^T of the incumbent as

$$b_i^T = c_i + \delta \frac{\frac{1-d_p}{2d_p} + \frac{(q_i - b_i^T + \bar{q}_e + \delta)(2 + \frac{1-d_p}{2}M) - 1}{\frac{1-d_p}{2}M}}{\frac{(q_i - b_i^T + \bar{q}_e + \delta)(2 + \frac{1-d_p}{2}M) - 1}{\frac{1-d_p}{2}M} - \frac{(q_i - b_i^T + \bar{q}_e)(2 + \frac{1-d_p}{2}M) - 1}{\frac{1-d_p}{2}M}}. \tag{7.44}$$

Solving for b_i^T yields

$$b_i^T = \frac{M - 2Md_p^2(\delta + q_i + c_i) + d_p(-M + 2(4+M)\delta + 2(4+M)c_i + 2(4+M)q_i)}{4d_p(4 + M(1-d_p))} := b_i^x \tag{7.45}$$

Differentiation with respect to d_p yields

$$\frac{\partial b_i^x}{\partial d_p} = -\frac{M(4 + M(1-d_p)^2)}{4d_p^2(4 + M(1-d_p))^2} < 0, \tag{7.46}$$

hence, b_i^x is strictly decreasing in d_p. However, b_i^x from Eq. (7.45) may not be the terminal bid for all $d_p \in [\underline{d}_p^i; 1]$ because the incumbent will stop lowering his bid if his probability of winning the postponed volume is 1. In Eq. (3.10) we defined the bid that guarantees the incumbent to win the postponed volume as $b_i^G = q_i + \overline{q}_e - \overline{q}_e^{post}(\hat{q}_e = 1, d_p)$. If $b_i^x < b_i^G$ for $d_p \in [\overline{d}_p^i; 1]$, the incumbent will bid $b_i^T = b_i^G$ for $d_p > \overline{d}_p^i$, because quoting b_i^G ensures that he wins the postponed volume and he has no incentive to bid below b_i^G.

$$\frac{\partial b_i^G}{\partial d_p} = \frac{2M}{(4 + (1-d_p)M)^2} > 0, \tag{7.47}$$

shows that b_i^G is strictly increasing in d_p. b_i^x is decreasing in d_p. There exists a $\overline{d}_p^i \in [\underline{d}_p^i; 1]$ for which $b_i^x = b_i^G$: because $\lim_{d_p \to 1^-} b_i^G = q_i$ and $\lim_{d_p \to 1^-} b_i^x = \frac{q_i + c_i}{2}$, it holds that $b_i^G > b_i^x$ for $d_p = 1$. Additionally, $b_i^T = q_i$ for $d_p < \underline{d}_p^i$, therefore there exists a $\overline{d}_p^i \in [\underline{d}_p^i; 1]$ such that $b_i^T = b_i^G$. Therefore

$$b_i^T = \begin{cases} q_i & \text{for} \quad 0 \le d_p \le \underline{d}_p^i \\ b_i^x & \text{for} \quad \underline{d}_p^i < d_p \le \overline{d}_p^i \\ b_i^G & \text{for} \quad \overline{d}_p^i < d_p \le 1 \end{cases}. \tag{7.48}$$

The proof for $v_e > v_i = 0$ follows the same logic, however, $b_e^G = \overline{q}_e^{post}(\hat{q}_e = 0, d_p)$. It is simple to show that $\lim_{d_p \to 1^-} b_e^G = \overline{q}_e$ and b_e^G is strictly increasing in d_p, which results in the same argument as above. □

b) In the proof of Proposition 7a we have shown that $\overline{d}_p^j$ is characterized as the intersection point of b_j^T and b_j^G. The proof for Proposition 7b follows the same path as the proof of Proposition 6c. We show how b_j^T and b_j^G depend on v_j and M. First, v_j depends on q_j and c_j, therefore we differentiate w.r.t. to q_j and c_j:

$$\frac{\partial b_j^T}{\partial q_j} = \frac{1}{2} > 0, \qquad \frac{\partial b_j^G}{\partial q_j} = 1 > 0, \tag{7.49}$$
$$\frac{\partial b_j^T}{\partial c_j} = \frac{1}{2} > 0, \qquad \frac{\partial b_j^G}{\partial c_j} = 0. \tag{7.50}$$

So b_j^T and b_j^G are both strictly increasing in q_j, and b_i^G is increasing at a strictly higher rate, therefore, the intersection point $\overline{d}_p^j$ is strictly decreasing in q_j, i.e. strictly decreasing in v_j (v_j is strictly decreasing in q_j). Second, b_j^T is strictly increasing in c_j and b_j^G is constant, hence, the intersection point is strictly increasing in c_j and therefore strictly decreasing in v_j (v_j is strictly decreasing in c_j). Next we differentiate b_j^T and b_j^G w.r.t. M:

$$\frac{\partial b_j^T}{\partial M} = \frac{1}{d_p}\frac{1-d_p}{(4+M(1-d_p))^2} > 0 \tag{7.51}$$
$$\frac{\partial b_j^G}{\partial M} = -2\frac{1-d_p}{(4+M(1-d_p))^2} < 0. \tag{7.52}$$

So b_j^T is strictly increasing in M, while b_j^G is strictly decreasing in M, therefore the intersection point $\overline{d}_p^j$ is strictly increasing in M. □

Proof of Proposition 8

Proposition 8 directly follows from Propositions 6 and 7. For $v_i = v_e$ bids are strictly decreasing for $d_p > \underline{d}_p$ and therefore procurement costs for the buyer are at a minimum for $d_p = 1$. For $v_j > v_{-j} = 0$ supplier $-j$'s bid is constant and supplier j's bid is minimal for $d_p = \overline{d}_p^j$, hence, the buyer minimizes procurement costs for $d_p = \overline{d}_p^j$.

Proof of Proposition 9

Assume that $Q_e \sim Beta(1,1)$. With the assumption of $v_i = v_e$ both bids are decreasing for $d_p > \underline{d}_p$ (see Proposition 6a). The terminal bids are (assuming that $\delta \to 0$)

$$b_j^T = \begin{cases} \mathbb{E}[q_j] & d_p < \underline{d}_p \\ c_j + \frac{1-d_p}{2d_p}\frac{M}{4+M(1-d_p)} & d_p \geq \underline{d}_p \end{cases} \tag{7.53}$$

With this we can characterize the optimization problem (3.1). For $d_p \leq \underline{d}_p$ we get:

$$\mathbb{E}[U(d_p)] = M\left((1-d_p)\left(\tfrac{q_i-b_i^T}{2} + \tfrac{\bar{q}_e-b_e^T}{2}\right) + d_p \int_0^1 f(\hat{q}_e)V(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e\right) \tag{7.54}$$

$$= M\left((1-d_p)\left(\tfrac{q_i-q_i}{2} + \tfrac{\bar{q}_e-\bar{q}_e}{2}\right) + d_p \int_0^1 f(\hat{q}_e)V(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e\right) \tag{7.55}$$

where $V(\hat{q}_e,d_p) = \max\left\{q_i - q_i, \bar{q}_e^{post}(\hat{q}_e,d_p) - \bar{q}_e\right\}$.

With $\bar{q}_e^{post}(\hat{q}_e,d_p)$ from Proposition 2 simplification yields

$$M\left((1-d_p)\left(\tfrac{q_i-q_i}{2} + \tfrac{\bar{q}_e-\bar{q}_e}{2}\right) + d_p \int_0^1 f(\hat{q}_e)V(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e\right) \tag{7.56}$$

$$= M\left(d_p\Big(\int_0^{\frac{1}{2}} f(\hat{q}_e)(q_i-q_i)\mathrm{d}\hat{q}_e + \int_{\frac{1}{2}}^1 f(\hat{q}_e)(\bar{q}_e^{post}(\hat{q}_e,d_p)-\bar{q}_e)\mathrm{d}\hat{q}_e\Big)\right) \tag{7.57}$$

$$= M\left(d_p \int_{\frac{1}{2}}^1 f(\hat{q}_e)(\bar{q}_e^{post}(\hat{q}_e,d_p)-\bar{q}_e)\mathrm{d}\hat{q}_e\right) \tag{7.58}$$

$$= M\left(d_p\Big(\int_{\frac{1}{2}}^1 \bar{q}_e^{post}(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e - \bar{q}_e\int_{\frac{1}{2}}^1 1\mathrm{d}\hat{q}_e\Big)\right) \tag{7.59}$$

$$= M\left(d_p\Big(\int_{\frac{1}{2}}^1 \frac{1+\hat{q}_e\frac{1}{2}(1-d_p)M}{2+\frac{1}{2}(1-d_p)M}\mathrm{d}\hat{q}_e - \bar{q}_e(1-\frac{1}{2})\Big)\right) \tag{7.60}$$

$$= M\left(d_p\Big(\Big[\frac{\hat{q}_e+\hat{q}_e^2\frac{1}{2}(1-d_p)M}{2+\frac{1}{2}(1-d_p)M}\Big]_{\frac{1}{2}}^1 - \bar{q}_e(1-\tfrac{1}{2})\Big)\right) \tag{7.61}$$

$$= M\left(d_p\Big(\Big(\frac{1+1^2\frac{1}{2}(1-d_p)M}{2+\frac{1}{2}(1-d_p)M} - \frac{\frac{1}{2}+\frac{1}{2}^2\frac{1}{2}(1-d_p)M}{2+\frac{1}{2}(1-d_p)M}\Big) - \bar{q}_e(1-\tfrac{1}{2})\Big)\right) \tag{7.62}$$

$$= M\left(d_p\Big(\Big(\frac{\frac{1}{2}+\frac{3}{8}(1-d_p)M}{2+\frac{1}{2}(1-d_p)M}\Big) - \tfrac{1}{2}\tfrac{1}{2}\Big)\right) \tag{7.63}$$

$$= M\left(d_p\Big(\Big(\frac{\frac{1}{2}+\frac{3}{8}(1-d_p)M}{2+\frac{1}{2}(1-d_p)M}\Big) - \tfrac{1}{4}\Big)\right) \tag{7.64}$$

$$= \tfrac{(1-d_p)d_pM^2}{32-8(1-d_p)M} \tag{7.65}$$

which is concave because

$$\tfrac{\partial^2\mathbb{E}[U(d_p)]}{\partial^2 d_p} = -\tfrac{M^2(4+M)}{(4+(1-d_p)M)^3} < 0 \tag{7.66}$$

and is maximized at

$$\frac{\partial \mathbb{E}[U(d_p)]}{\partial d_p} = \frac{M^2(4+M-2(4+M)d_p+Md_p^2)}{8(4+M(1-d_p))^2} \stackrel{!}{=} 0 \Rightarrow d_p = \frac{1}{1+\frac{2}{\sqrt{4+M}}} = d_p^Q. \tag{7.67}$$

For $d_p \geq \underline{d}_p$ we obtain

$$\begin{aligned}\mathbb{E}[U(d_p)] =& M\Bigg((1-d_p)\left(\frac{q_i-b_i^T}{2}+\frac{\frac{1}{2}-b_e^T}{2}\right)\\ &+ d_p\int_0^1 f(\hat{q}_e)V(\hat{q}_e,d_p)\mathrm{d}\hat{q}_e\Bigg)\end{aligned} \tag{7.68}$$

where $V(\hat{q}_e, d_p) = \max\left\{q_i - b_i^T, \bar{q}_e^{post}(\hat{q}_e, d_p) - b_e^T\right\}$.

First, we need to look at $V(\hat{q}_e, d_p)$ and find a $\hat{q}_e$ such that $q_i - b_i^T > \bar{q}_e^{post}(\hat{q}_e, d_p) - b_e^T$ (which means we find cases for which the incumbent wins the postponed volume (and vice versa)). Because of $v_i = v_e$ it is easy to see that $\hat{q}_e < \frac{1}{2}$. With this simplification of

Eq. (7.68) yields

$$\mathbb{E}[U(d_p)] = M\Bigg((1-d_p)\Big(\tfrac{q_i-b_i^T}{2}+\tfrac{\frac{1}{2}-b_e^T}{2}\Big)+ \tag{7.69}$$

$$d_p\Big(\int_0^{\frac{1}{2}} f(\hat{q}_e)(q_i-b_i^T)\mathrm{d}\hat{q}_e+\int_{\frac{1}{2}}^1 f(\hat{q}_e)(\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T)\mathrm{d}\hat{q}_e\Big)\Bigg) \tag{7.70}$$

$$= M\Bigg((1-d_p)\Big(\tfrac{q_i-b_i^T}{2}+\tfrac{\frac{1}{2}-b_e^T}{2}\Big)+ \tag{7.71}$$

$$d_p\Big((\tfrac{1}{2}-0)(q_i-b_i^T)+\int_{\frac{1}{2}}^1 f(\hat{q}_e)(\bar{q}_e^{post}(\hat{q}_e,d_p)-b_e^T)\mathrm{d}\hat{q}_e\Big)\Bigg) \tag{7.72}$$

$$= M\Bigg((1-d_p)\Big(\tfrac{q_i-b_i^T}{2}+\tfrac{\frac{1}{2}-b_e^T}{2}\Big)+ \tag{7.73}$$

$$d_p\Big((\tfrac{1}{2}-0)(q_i-b_i^T)+\int_{\frac{1}{2}}^1 \Big(\frac{1+\frac{1-d_p}{2}\hat{q}_e}{2+\frac{1-d_p}{2}}-b_e^T\Big)\mathrm{d}\hat{q}_e\Big)\Bigg) \tag{7.74}$$

$$= M\Bigg((1-d_p)\Big(\tfrac{q_i-b_i^T}{2}+\tfrac{\frac{1}{2}-b_e^T}{2}\Big)+ \tag{7.75}$$

$$d_p\Big((\tfrac{1}{2}-0)(q_i-b_i^T)-b_e^T\int_{\frac{1}{2}}^1 1\mathrm{d}\hat{q}_e+\frac{\frac{1-d_p}{2}M}{2+\frac{1-d_p}{2}M}\int_{\frac{1}{2}}^1 \hat{q}_e\mathrm{d}\hat{q}_e\Big)\Bigg) \tag{7.76}$$

$$= M\Bigg((1-d_p)\Big(\tfrac{q_i-b_i^T}{2}+\tfrac{\frac{1}{2}-b_e^T}{2}\Big)+d_p\big(\tfrac{1}{2}(q_i-b_i^T) \tag{7.77}$$

$$-(1-\tfrac{1}{2})b_e^T+(1^2-(\tfrac{1}{2})^2)\Big(\tfrac{1}{2}\frac{\frac{1-d_p}{2}M}{2+\frac{1-d_p}{2}M}\Big)\Bigg). \tag{7.78}$$

With the definition of b_j^T from (7.53) in the appendix, simplification results in

$$\mathbb{E}[U(d_p)] = \tfrac{M(16(2c_j-1)+(1-d_p)(4-d_p(4-8c_j+d_p))M)}{8d_p(4+M(1-d_p))}. \tag{7.79}$$

Differentiation w.r.t. d_p yields

$$\tfrac{\partial\mathbb{E}[U(d_p)]}{\partial d_p} = \tfrac{M^2(4(4+d_p^2(1-d_p))+(1-d_p)^2(4+d_p^2)M)}{8d_p^2(4+M(1-d_p))^2} > 0, \tag{7.80}$$

hence, $\mathbb{E}[U(d_p)]$ is strictly increasing in $d_p > \underline{d}_p$ and therefore there exist a local maximum for $d_p \geq \underline{d}_p$ at $d_p = 1$. In summary, for $v_i = v_e$ $\mathbb{E}[U(d_p)]$ is concave for $0 \leq d_p \leq \underline{d}_p$ with a local maximum at $d_p = \frac{1}{1+\frac{2}{\sqrt{4+M}}} = d_p^Q$, and strictly increasing for $d_p > \underline{d}_p$ with a local maximum at $d_p = 1 = d_p^C$.

To calculate the global maximum we need to compare the expected utilities of the local maxima $\mathbb{E}[U(d_p^Q)]$ and $\mathbb{E}[U(d_p^C)]$

$$\mathbb{E}[U(d_p^Q)] = \frac{1}{8}(8 + M - 4\sqrt{4+M}) \tag{7.81}$$
$$\mathbb{E}[U(d_p^C)] = \frac{1}{2}(M - 2c_j M) \tag{7.82}$$

Comparison of both utilities yields

$$\mathbb{E}[U(d_p^Q)] \geq \mathbb{E}[U(d_p^C)] \tag{7.83}$$
$$\Leftrightarrow \quad c_j \geq \frac{-8+3M+4\sqrt{4+M}}{8M} \tag{7.84}$$
$$\Leftrightarrow \quad \frac{1}{2} - v_j \geq \frac{-8+3M+4\sqrt{4+M}}{8M} \tag{7.85}$$
$$\Leftrightarrow \quad v_j \leq \frac{1}{2} - \frac{-8+3M+4\sqrt{4+M}}{8M} \tag{7.86}$$

□

Proof of Proposition 10

This proof directly follows from the results of Proposition 7. Consider the incumbent's expected terminal bid $\mathbb{E}[b_i^T]$: with probability p we are in the case $v_i > v_e = 0$ and in this case b_i^T is strictly decreasing in d_p for $\underline{d}_p^i < d_p \leq \overline{d}_p^i$ and strictly increasing in d_p for $\overline{d}_p^i < d_p \leq 1$. With probability $1 - p$ we have $v_i = 0$ and $b_i^T = q_i$, hence, b_i^T is constant in d_p. Because $\mathbb{E}[b_i^T] = p b_{i,v_i>v_e=0}^T + (1-p) b_{i,v_i=0}^T$, for $0 < p < 1$ the expected terminal bid $\mathbb{E}[b_i^T]$ is strictly decreasing in d_p for $\underline{d}_p^i < d_p \leq \overline{d}_p^i$ and strictly increasing in d_p for $\overline{d}_p^i < d_p \leq 1$. The proof for the entrant's expected terminal bid can be reproduced by using equivalent arguments. □

7.3 Appendix III

In this section we explain how we compute the numerical results presented in Section 3.3.3 and 3.4. We evaluate Eq. (3.1) for a given set of parameters $v_i = q_i - c_i$, $v_e = \bar{q}_e - c_e$, M and entrant quality distribution $Q_e \sim Beta(\alpha, \beta)$ $(\alpha = \beta > 0)$, and a given postponed volume d_p. Because the terminal bids derived in Section 3.3.2 can not be directly computed, the numerical evaluation is defined by a multiple step process: 1) compute terminal bids, 2) compute expected utility, quality and costs, 3) identify the global maxima and minima for expected utility, quality and costs:

1. For each $d_p \in [0, 0.01, \ldots, 0.99, 1]$:

Initialize supplier bids $b_j^0 = \mathbb{E}[q_j]$ and compute the terminal bids b_j^T by following the definition given in Proposition 5.

2. Compute *Expected Utility* $\mathbb{E}[U(d_p)]$, *Expected Quality* ($:= \mathbb{E}[Q(d_p)]$) and *Expected Costs* ($:= \mathbb{E}[C(d_p)]$) as defined by Eq. (3.1), for the terminal bids b_j^T for each $d_p \in [0, 0.01, \ldots, 0.99, 1]$.

3. Calculate $d_p^* = \text{argmax}\{\mathbb{E}[U(d_p)]\}$, $d_p^Q = \text{argmax}\{\mathbb{E}[Q(d_p)]\}$ and $d_p^C = \text{argmin}\{\mathbb{E}[C(d_p)]\}$.

Chapter 8

Appendix to Chapter 4

The mixed integer linear program for the single-product setting as defined in Section 4.4 can be formulated as follows:

$$\min \sum_{t=1}^{T} I_t h + B_t b + O_t k + Q_t p \tag{8.1}$$

$$\text{s.t.} \quad I_0 = 0 \tag{8.2}$$

$$I_t = I_{t-1} + Q_t - \delta_t, \quad t \in \{1, \ldots, T\} \tag{8.3}$$

$$B_t \geq -I_t, \quad t \in \{1, \ldots, T\} \tag{8.4}$$

$$B_t \geq 0, \quad t \in \{1, \ldots, T\} \tag{8.5}$$

$$O_t \in 0, 1, \quad t \in \{1, \ldots, T\} \tag{8.6}$$

$$O_t M - Q_t \geq 0, \quad t \in \{1, \ldots, T\} \tag{8.7}$$

I_t is the inventory level in period t, B_t are the backorders in period t, O_t is a binary variable indicating whether or not an order in period t is placed and Q_t is the size of the order in period t. Constraint (8.2) sets the starting inventory as zero, and (8.3) establishes the logic how inventory is carried over from period $t-1$ to t. Constraints (8.4) and (8.5) are a linear representation of the constraint $B_t = \max[0, -I_t]$ for the backorders in period t. The binary variable for an order in period t is defined in (8.6), and constraint (8.7) makes sure that an order in period t is only positive if $O_t = 1$.

The mixed integer linear program for the multi-product setting as defined in Section

4.4 can be formulated as follows:

$$\min \sum_{t=1}^{T} O_t^l k + O_t^h k + \sum_{i=1}^{I} I_t^i h + B_t^i b + Q_t^{i,h} p + Q_t^{i,l}(1-\epsilon)p \tag{8.8}$$

$$\text{s.t.} \quad I_0^i = 0, \quad i \in \{1,\dots,I\} \tag{8.9}$$

$$I_t^i = I_{t-1}^i + Q_t^{i,h} + Q_t^{i,l} - d_t^i, \quad t \in \{1,\dots,T\},\ i \in \{1,\dots,I\} \tag{8.10}$$

$$B_t^i \geq -I_t^i, \quad t \in \{1,\dots,T\},\ i \in \{1,\dots,I\} \tag{8.11}$$

$$B_t^i \geq 0, \quad t \in \{1,\dots,T\},\ i \in \{1,\dots,I\} \tag{8.12}$$

$$O_t^h \in \{0,1\}, \quad t \in \{1,\dots,T\} \tag{8.13}$$

$$O_t^l \in \{0,1\}, \quad t \in \{1,\dots,T\} \tag{8.14}$$

$$O_t^h M - Q_t^{i,h} \geq 0, \quad t \in \{1,\dots,T\},\ i \in \{1,\dots,I\} \tag{8.15}$$

$$O_t^l M - Q_t^{i,l} \geq 0, \quad t \in \{1,\dots,T\},\ i \in \{1,\dots,I\} \tag{8.16}$$

$$\textstyle\sum_{i=1}^{I} Q_t^{i,l} - \theta O_t^l \geq 0, \quad t \in \{1,\dots,T\},\ i \in \{1,\dots,I\} \tag{8.17}$$

I_t^i is the inventory level of product i in period t, B_t^i are the backorders in period t, O_t^h is a binary variable indicating whether or not a high cost order in period t is placed and $Q_t^{i,h}$ is the size of the order in period t for product i. O_t^l is a binary variable indicating whether or not a low cost order in period t is placed and $Q_t^{i,l}$ is the size of the order in period t for product i. Constraint (8.9) sets the starting inventory as zero, and (8.10) establishes the logic of how inventory is carried over from period $t-1$ to t. Constraints (8.11) and (8.12) are a linear representation of the constraint $B_t^i = \max[0, -I_t]$ for the backorders in period t. The binary variable for a high cost order in period t is defined in (8.14), and constraint (8.15) makes sure that an order in period t is only positive if $O_t^h = 1$. The binary variable for a low cost order in period t is defined in (8.13), and constraints (8.16) and (8.17) make sure that this order in period t is only positive if $O_t^l = 1$ and is equal to or larger than the threshold θ.