Impact of Performance-Based Contracting on Product Reliability: An Empirical Analysis

Jose A. Guajardo, Morris A. Cohen

The Wharton School, University of Pennsylvania, Philadelphia, PA 19104
josegu@wharton.upenn.edu, cohen@wharton.upenn.edu

Sang-Hyun Kim

Yale School of Management, Yale University, New Haven, CT 06511, sang.kim@yale.edu

Serguei Netessine

INSEAD, Boulevard de Constance, 77305 Fontainebleau, France, serguei.netessine@insead.edu

First version: July 2009; this version: September 2011

Abstract

Using a proprietary dataset provided by a major manufacturer of aircraft engines, we empirically investigate how product reliability is impacted by use of two different types of after-sales maintenance support contracts: time and material contracts (T&MC) and performance-based contracts (PBC). We offer a number of competing arguments based on the theory of incentives that establish why product reliability may increase or decrease under PBC. We build a two-stage econometric model that explicitly accounts for the endogeneity of contract choices, and find evidence of a positive and significant effect of PBC on product reliability. The estimation of our model indicates that product reliability is higher by 25-40% under PBC compared to under T&MC, once the endogeneity of contract choice is taken into account. Our results are consistent with two mechanisms for reliability improvement under PBC: more frequent scheduled maintenance and better care performed in each maintenance event.

1. Introduction

The movement towards a service-based economy has led many manufacturing firms to recognize the strategic importance of after-sales product support services that enable the availability of properly functioning products. In many industries, such services are a major source of revenue, profit, and growth and thus act as a source of sustainable competitive advantage (Cohen et al. 2006). This is especially true in those industries where products are complex and the consequences of product downtime can be severe. Moreover, when products have relatively long life cycles (e.g., aircraft, engines, semiconductor fab equipment, medical imaging devices, etc.), they present the firms supplying after-sales support with ample opportunities to provide repair and maintenance services. As many OEM firms in such industries reposition themselves to become service providers, it has become critical for them to evaluate and define contractual relationships with their customers for the provision of after-sales support. Traditionally, after-sales services have been performed under time and material contracts (T&MC), under which the supplier is compensated for the amount of resources consumed (such as spare parts and labor) whenever product maintenance is required. However, a new form of a support contract has emerged in recent years: performance-based contracts (PBC). Under PBC, also referred to as Performance-based Logistics (PBL) in the defense sector and Power by the Hour® in the commercial sector, a supplier is paid based on the realized outcome of customer value. For example, an airline customer pays an engine service provider in proportion to the number of aircraft flying hours, which is affected by engine up-time (i.e., the number of hours the engine was available for use), and which determines the value derived by the customer.

PBC has been popularized in the aerospace industry in particular because it was recognized early on that it brings a potential benefit of aligning incentives among customers and suppliers. PBC compensates the supplier based on the same outcome that the customer cares about (i.e., product utilization), and hence the supplier is motivated to increase product performance, associated with metrics such as product reliability and availability. It has been noted that the risk of supplier moral hazard is high under the traditional T&MC since the provision of services that the customer procures under it, such as the spare parts and repair labor, is typically a high margin and profitable activity for suppliers. PBC can mitigate this problem, since under this contract the supplier is responsible for these costs.

Past theoretical results support the notion that adoption of PBC can result in increased product performance at a lower cost (Kim et al. 2011, Hypko et al. 2010, Randall et al. 2010). Empirical evidence to support this conclusion, however, is not conclusive. Kirk and DePalma (2005) analyze a Navy PBC program and, based on a review of historical repair frequency data for several programs, offer the following mild observation: "there is some evidence that the PBC contract may have helped to improve availability and reliability." To the question "Do you have evidence of higher performance/lower cost based on past or current [PBC] programs?" included in a recent survey conducted in the aerospace industry (Newsome 2008), the responses were decidedly mixed: "Yes" – 33%, "No" – 36%, "Too early to tell" – 31 %. A recent Government Accountability Office (GAO) report on PBC in the defense industry presents a similar view, and points to the limitation of existing studies: "Many DoD [Department of Defense] program offices that implemented [PBC] arrangements have limited cost data, and various other factors – such as the lack of business case analyses – further limit an evaluation of the costs of this support strategy. Available data from the programs GAO reviewed indicated mixed results" (GAO

2008).[1] A 2009 DoD study (DoD 2009), which reviewed over 30 weapon system programs, reported that it is widely accepted that PBC leads to higher material availability, but that conclusions concerning cost reduction and other performance implications such as reliability improvement are less clear.

Our review of industry studies indicate that they are primarily based on comparisons of observed average measures or on simple regression analyses that leave out numerous confounding factors, bias adjustments, and statistical robustness checks. As we discuss below, there are several complexities that make the quantification of the impact of PBC on metrics such as product reliability a challenging task. Indeed, the fact that prior evaluations of PBC have ignored the modeling of such complexities helps to explain why the debate concerning the potential for PBC to provide significant improvement of product performance remains unresolved at this time. The main goal of this paper is to conduct a thorough examination of the competing arguments that associate PBC with improved product performance in a specific context (i.e., aircraft engine reliability) and empirically evaluate them, thereby offering a new perspective to this issue that has captured practitioners' attention in the past few years since PBC became widely adopted.

Our focus on the research question "Does PBC result in reliability improvement?" is motivated by the intense debate currently underway among practitioners and policy makers concerning the same issue. Indeed, product reliability is regarded as a prime performance metric in the aerospace industry, not only because it is an important driver of product utilization, but also because in this industry product failures due to imperfect reliability lead to direct and large financial losses as well as to potentially disastrous consequences (e.g., loss of life). A good illustration of the role of reliability in this industry is the recent engine failure incident that affected the Qantas Airways Ltd. which received extensive media coverage worldwide. The mid-air blowout of one of Qantas' Trent 900 engines frightened all airline passengers, and had direct short-term financial effects on both Qantas and Rolls-Royce, the engine manufacturer; the stock prices of the two companies fell by roughly 10% in the days/weeks after this incident (Clark and Mouawad 2010). In the closely-related defense sector, the DoD considers achieving a high level of product reliability as one of the three essential elements of enabling mission capability, along with availability and maintainability (DoD 2005).

In the academic literature, a few studies have proposed models that link PBC with reliability improvement. Kim et al. (2011), based on a game-theoretic model analysis under a limited set of assumptions, suggest that the answer to the reliability improvement question is an unqualified "yes". However, as we demonstrate in this paper, there are several real-world and theoretical considerations that

[1]Reflecting the intense scrutiny over PBC employed in defense and other government services, House of Representatives recently held a hearing on the benefits and costs of PBC. See "Performance-based Acquisitions: Creating Solutions or Causing Problems?" Full Committee Thursday, May 08, 2008. http://homeland.house.gov/hearings/index.asp?ID=136.

are not captured in Kim et al. (2011) which make the reasoning much more nuanced and the outcome less certain. In fact, as we discuss in more detail in Section 3, there are a number of competing arguments that may lead to a positive, negative, or neutral answer to the question. In this paper, we address this question by presenting the results of an empirical analysis performed on a unique proprietary dataset provided by Rolls-Royce, a leading manufacturer and a service provider for aircraft engines. Rolls-Royce, like many companies in the aerospace industry, offers its customers the two types of contracts we have mentioned: T&MC and PBC.

One of the key challenges in measuring the effect of PBC on product reliability is the presence of heterogeneous customer preferences for contract types that requires us to model customers' contract choice decisions explicitly. Without controlling for this endogenous contract selection process, as we demonstrate, one could erroneously conclude that there is no statistically significant effect of contract type on product reliability. Instead, we propose a two-stage framework that explicitly deals with the endogeneity inherent in contract choice by a customer, and we provide evidence at the 95% confidence level that product reliability under PBC is in fact higher by about 25-40% compared to that under T&MC. The conclusions from this analysis are robust to a large number of alternate specifications and modeling assumptions. Thus, our findings lend support to the view that adoption of PBC results in product performance enhancement. In addition, our results are consistent with two separate mechanisms by which reliability improvement can be achieved under PBC: more frequent scheduled maintenance and better care performed in each maintenance event. The latter impact can be accomplished by such activities as conducting more thorough checks leading to better identification of defects, preemptive parts replacements, and possible product re-design.

The rest of the paper is organized as follows. After a brief literature review in Section 2, in Section 3, we lay out the problem context and the theoretical arguments that lead to the hypothesis we aim to test empirically. We then describe the data and present the analysis in Sections 4-7, concluding in Section 8.

2. Related Literature

While there is an abundance of theoretical papers on the subject of supply chain contracting in the Operations Management (OM) literature (e.g., Cachon 2003 provides an extensive review of more than 200 papers in this area), most of them consider simplified settings in which contract choices made by customers with heterogeneous preferences are not taken into account. In contrast, endogenous contract choice is a central feature in our analysis. Heterogeneity in customer preferences can be explained in many different ways, but especially relevant to our problem context is heterogeneity created by information asymmetry, namely, by adverse selection and moral hazard. In this respect, OM papers such as Corbett and de Groote (2000) and Iyer et al. (2005) that analyze the optimal design of contracts in the presence of information asymmetry are relevant to our analysis. We note, however, that these papers focus mostly on how firms make choices from a menu of contract terms within a single class of contracts (e.g., menu of price-quantity pairs) – as suggested by the standard mechanism design theory – rather than choose one contract type from multiple classes (e.g., T&MC vs. PBC). The latter situation, which we analyze in this paper, is not thoroughly investigated in the literature.

Although numerous papers based on game-theoretic analyses have produced an abundance of predictions on how supply chain contracts should be optimally structured, empirical validations in the OM literature are scarce. Indeed, as Cachon (2003) notes, "the literature contains a considerable amount of theory, but an embarrassingly paltry amount of empiricism." Although a small number of empirical papers do exist under the broad theme of supply chain coordination – for example, Novak and Eppinger (2001) and Novak and Stern (2008, 2009) examine the impact of product characteristics on vertical integration decisions in the automobile industry – the issue of contractual incentives has not received comparable empirical scrutiny. This paper contributes to the literature by providing an empirical analysis of unique transaction data that sheds light on the role of supply chain contracting in affecting product reliability, a key variable of interest in the OM and quality management areas. There are other empirical OM papers that are related to our study, including Ramdas and Randall (2008; product reliability in the automotive industry), Deshpande et al. (2003; after-sales services in the defense industry), and Terwiesch et al. (2005; incentive conflicts in the semiconductor industry). However, none of these papers explicitly consider the incentives created by contract type choices.

In contrast to the current state of empirical research on contracting in the OM literature, progress has been made in other areas. Chiappori and Salanie (2003) detail the development of the contracting literature in economics and point out that, just as in OM, early contracting papers were predominantly theoretical and did not account for the process of contract selection by heterogeneous agents. Only recently have the economists started addressing this shortcoming, focusing mainly on behavior of individuals (e.g., Ackerberg and Botticini 2002). A few very recent exceptions (some still unpublished) have investigated revenue sharing contracts in the video rental industry featuring contract self-selection (Mortimer 2008, Ho et al. 2010, 2011). Not surprisingly, given the different nature of the problems of engine maintenance vs. video rentals, our analysis involves contract types, incentives mechanisms, and performance metrics that are quite distinct from theirs. Except for these examples, there is no empirical paper that considers contract self-selection in supply chains that we are aware of. Similarly, some empirical papers have examined the influence of variants of PBC in applications in service sectors with a focus on the behavior of individuals (e.g., Lazear 2000, Prendergast 2002 in labor economics; Lu et al. 2003, Shen 2003 in health care; Heinrich 2002 in public policy). However, whether the findings from such firm-to-individuals settings extend to firm-to-firm settings is unclear (because of differences in buyer characteristics such as price sensitivity, attitude towards risk, access to capital, number of stakeholders involved, etc.), and comparable studies in the latter setting are sparse. Finally, in a related but different context of offshore software development, Gopal et al. (2003) analyze the impact of fixed-price vs. T&MC on software vendor profits (Gopal and Sivaramakrishnan 2008 perform a similar analysis). As we do, they use a two-stage modeling approach that includes the determinants of contract choice, but in the context of procurement of an intangible product (software). In addition, their research differs from ours because the contracts they study are not based on product performance, which is the central focus of our analysis.

This paper contributes to the operations management (OM) literature by being one of the few studies that empirically examine and test predictions from supply chain contracting models, thereby bridging the gap between theory and empirical evidence in this area. Moreover, we add a new dimension to the literature by showing that customers' product support contract choice decisions are integral in linking a contract type with product performance. This finding counters the arguments found in Kim et al. (2011), Hypko et al. (2010), and Randall et al. (2010) who suggest that it is sufficient to simply count the number of product failures under the two contracts (T&MC and PBC), take averages, and infer that one contract leads to higher reliability (they all suggest PBC does). We show that this is not necessarily the case; without accounting for customers' self-selection of contracts (and thus failing to consider endogeneity), this approach can lead to misleading conclusions.

3. Industry Background and Theoretical Motivation

Our research setting is in the maintenance, repair and overhaul (MRO) market for commercial aircraft. According to Standard and Poor's (2011), the MRO sector generated revenues of $111 billion in 2009, of which $62 billion was attributed to military MRO, $42.7 billion to air transport (commercial aircraft) MRO, and $6.2 billion to business and general aviation MRO. More generally, reported statistics (see Cohen et al. 2006 and the references therein) indicate that sales of spare parts and after-sales services in the U.S. represents 8% of annual domestic product, meaning that customers spend approximately $1 trillion every year to maintain assets they already own.

In recent years, customers in the commercial aerospace industry, i.e., the airlines, have increasingly adopted outsourcing strategies for MRO services in order to focus on their core competencies and to reduce costs. This trend has led to expansion of the range of MRO services offered by suppliers of various types of aircraft subsystems (e.g., hydraulic power systems, engines, avionics systems). A unique feature of this market is that it is quite common for the OEMs that manufacture the systems and subsystems to offer support services for their own products. This is due to the highly customized and complex nature of the products, which can make it difficult for a third party to provide the level of product care that customers require. Consequently, the provision of such services has been very profitable for OEM firms such as Pratt & Whitney, General Electric Co., Rolls-Royce, Boeing and Lockheed Martin.

Not surprisingly, customers, who often end up paying far more for after-sales services than for the products themselves over the life of product usage, have started to demand that suppliers provide contract options that reduce their cost for product support. PBC was introduced as a response to this demand, and it is Rolls-Royce, one of the major aircraft engine manufacturers, that is universally recognized as the company that pioneered this concept. Rolls-Royce offers two different types of contracts to their customers: T&MC and PBC. The main distinction between the two is the basis of compensation to the supplier. Namely, under PBC the customer agrees to pay a fee in proportion to aircraft flying hours, which in turn is affected by the availability of all major aircraft sub-systems (including engines). Flying hours – a key measure of product utilization in the aerospace industry – depends heavily on subsystem reliability as well as other factors such as the stocking levels of spare parts inventory and the speed of repair at maintenance depots. In our dataset, we observe information on product reliability but not on the other factors that may also drive product utilization, explaining our focus on this single metric in our analysis. In defining product reliability, it is important to make a distinction between unplanned (or unscheduled) maintenance and planned (or scheduled) maintenance events. In this paper we associate product reliability with unplanned maintenance events because they represent unforeseen disruptions that can lead to the loss of customer value and to the costly measures that are required to mitigate their impact. Planned maintenance events, on the other hand, reflect managerial interventions that are scheduled in advance, and the time it takes to complete this type of maintenance has much lower impact on product utilization, relative to unplanned maintenance. In fact, the mean repair time for unplanned maintenance events in our sample is in the order of several weeks.

The main focus of this paper is on the question "Does PBC result in reliability improvement?" and in the quantification of such effect (if any). We will hypothesize that the answer to this question is positive, based on the basic theoretical argument that incentive alignment is enhanced under PBC. However, as we will discuss next, the presence of information asymmetry in our setting leads to potential double moral hazard and adverse selection effects, which, besides making the answer to the question not obvious, gives rise to a number of competing mechanisms by which product reliability can be affected throughout the product support process under different contract types.

We start by discussing the supplier side. First, note that the supplier has an opportunity to exert effort to improve reliability of existing products when they are in her possession during repair and maintenance processes. An appropriate framework to study the supplier's incentive to do so under T&MC or PBC is the moral hazard model. Our setting fits this framework because, in most practical cases, the supplier's reliability improvement efforts are discretionary and unobservable to the customer and are influenced by contractual incentives. This is especially true for reliability issues that require engine overhauls, since, in such instances, the customer loses visibility to the supplier's repair capabilities as soon as she sends a defective engine to a repair depot operated by the supplier. Recall that the supplier is compensated in proportion to the product usage under PBC, whereas under T&MC, he is paid each time the product fails and the customer requests repairs. Since product utilization increases with reliability, then, a supplier under PBC is motivated to improve product reliability in order to maximize his revenue. On the other hand, a supplier under T&MC has a skewed incentive that may lead him not to invest in reliability improvement or even degrade reliability, since each failure incident represents an opportunity for him to generate revenue by selling high-margin spare parts and charge for labor and other resources consumed. Thus, the supplier's service revenue could actually grow with lower reliability. This reasoning was formalized in Kim et al. (2011), who reach a similar conclusion in their analysis based on a game-theoretic model. In addition, this intuitive relationship between a contract type and reliability outcome is supported by some industry reports.[2]

If it was the case that reliability is the only variable that can be influenced by a managerial decision, based on the argument above, it is relatively straightforward to hypothesize that PBC is superior to T&MC in incentivizing the supplier to improve reliability. There are, however, two theoretical constructs that can be hypothesized to moderate the aforementioned discussion. The first issue is the multitasking aspect (Holmström and Milgrom 1991): given that the supplier can invest in not only reliability improvement but also in parts inventory, repair time reduction, and other efficiency gains, is it necessarily true that the level of reliability improvement is significantly higher under PBC than under T&MC? It is possible, for example, that improving reliability is too costly and therefore the supplier chooses other means to increase the product utilization. Second, an argument can be made that a reputational concern will prevent the supplier from neglecting reliability improvement, regardless of which contract he is subject to. One may further argue that the supplier is required to deliver the highest level of reliability that he can provide, since the customers in the airline industry are under heavy regulatory mandates. If this reasoning is correct and these effects in fact dominate, we would expect to see no significant relationship between a contract type and reliability. Thus, whereas a simple reasoning that focuses solely on the supplier-side moral hazard would lead us to believe that PBC is more effective in incentivizing the

---

[2] According to Thomas (2005), "… Rolls-Royce had officially won praise from the US Navy for its innovative 'PBtH' support for the F405 engine". According to Captain Win Everett, Program Manager for the US Navy's Undergraduate Flight Training Systems at NATC 'Patuxent River' (Maryland), 'under Rolls-Royce, engine availability has exceeded the current target of 85%, the average time between engine removals has increased from 700 hours to over 900 hours, and expected engine removals have fallen by 15 per cent.'" In addition, Business Wire (2008) cites reliability improvements after introduction of PBC contracts by Rolls-Royce. In PBC environments not directly related to engines or Rolls-Royce, Geary and Vitasek (2008) observe that, "... there were also 90 improvements made to the APU (auxiliary power unit), with 20 of those being reliability improvements", and "the contract with Raytheon... the system design and support concept used in this program have resulted in a 200% improvement in MTBOF (mean time between operations failures) and a 400% improvement in mean time to repair."

supplier to improve reliability than T&MC, arguments based on multitasking and reputation effects make the ultimate outcome of the supplier's actions not obvious.

Next, we turn our attention to the customers. It can be argued that, regardless of the contract type for after-sales services, it is unambiguously in the interest of a customer to make sure that high reliability is maintained given the high opportunity costs of having an aircraft on the ground. However, a careful examination of the customers' incentives under each contract type reveals that the effects of double moral hazard may be significant in this setting. This is because reliability is a function of not only the supplier's actions, but also the level of care that the customer exerts and the pattern of usage that the customer adopts. For customers, PBC can be viewed as a form of insurance that provides them with protection against unforeseen out-of-pocket charges that are incurred when an unexpected product malfunction occurs (unlike T&MC, under which the customer has to pay for spare parts and labor after those events). Therefore, it can be hypothesized that – relative to T&MC customers who are responsible for all costs associated with product failures – PBC customers may tend to operate their product with less care, contributing to wear-and-tear and thus making the product more prone to failures (Padmanabhan 1995 builds a model based on the same intuition, applying it to the case of extended warranties). If such situations are common, then we expect that the impact of a contract type on reliability to be not significant or may even be reversed, i.e., PBC may in fact degrade reliability (which would happen if the customer's moral hazard is dominant).

In summary, although the basic reasoning suggests that adopting PBC will result in higher product reliability based on an alignment of incentives, other theoretical arguments moderate or counter this view. Earlier theoretical studies (Kim 2011, Kim et al. 2011) point to the former conclusion, but they do not capture many of the confounding factors that we have identified. Thus, this question is best answered empirically. We state the hypothesis of our empirical research as follows.

Hypothesis: *Reliability under PBC is significantly higher than reliability under T&MC.*

Finally, and as we will discuss further in the following sections, it is important to account for the customers' contract selection mechanism in order to isolate the effect of PBC on product reliability. As we have described in our discussion so far, the setting we study is characterized by information asymmetry. In particular, a customer may keep his/her product usage profile private. Then, because of the insurance role of PBC, a customer who tends to overuse the product or who possesses products that tend to undergo high levels of stress would prefer PBC. If such an adverse selection motive is significant, then we expect to see a high rate of PBC selection among the customers who tend to use the supplier's maintenance services more often. Since the work of Rothschild and Stiglitz (1976), contract self-selection has been recognized as one of the important determinants of insurance markets, and it is also a feature that we expect to be relevant in our setting given the insurance role of PBC. While selection effects have been studied in the case of insurance markets and empirical testing of the associated theories has grown in the last ten years, empirical evidence indicates that information asymmetry leads to self-selection in some sectors but not in others (Chiappori and Salanie 2003). Additionally, some examples of *advantageous selection* in the insurance markets have been reported in the recent literature (Einav and Finkelstein 2011). Thus, the precise nature of a selection mechanism is far from obvious and is likely to depend on specific settings. Ultimately, whether the presence of contract self-selection is significant in our setting and which outcome the selection leads to are empirical questions, which we examine in our analysis.

4. Data

The dataset which Rolls-Royce made available to us consists of five years of data (July 2002 - July 2007, hereafter the observation period) of maintenance events (engine removals) for different models of aircraft engines produced by Rolls-Royce. A *removal* of the aircraft engine may be necessary due to a part failure (unplanned removal) or for maintenance purposes (planned removal), resulting in a *shop visit* to the service provider. For a better understanding of the data, it is useful to describe more precisely how the data for this research was provided to us. We obtained two different data files: a spreadsheet containing all of the removal events between July 2002 and July 2007 for the engines in our sample (hereafter called the "removals file"), and a spreadsheet containing a list with all engines registered with Rolls-Royce for each customer (hereafter called the "engines file"). For each removal of an engine unit, a list of the relevant information contained in the removals file is as follows: engine unit ID, engine model, date at which the engine entered the repair shop, cumulative flying hours ("time since new", TSN) at the time of a shop visit, cumulative cycles (CSN, defined in table 1) at the time of a shop visit, removal type (planned or unplanned), aircraft tail number in which the engine is installed, aircraft model ID, ID of the customer that owns the product, the contract type (T&MC or PBC) under which the product receives service. The engines file, on the other hand, contains a list with all of the engines registered with Rolls-Royce at the end of the observation period, i.e., July 2007. For each engine in this file we know: engine unit ID, engine model, cumulative aircraft flying hours and cycles at the end of the observation period, and the ID of the customer that owns the aircraft.

After cleaning data by removing inconsistent observations (e.g., for a given unit, reported flying hours at a shop visit in 2006 are less than the flying hours reported in a shop visit in 2005), our sample consisted of 763 engine units for which at least one engine removal is observed in the 5 years observation period. There are essentially two engine models in our sample: for one of these product models there are 3 different versions, and for the other there are 2 different versions. The engine models are installed in three different types of aircraft. For all types of product, we observe engine units covered by either PBC or T&MC. In the sample of 763 engine units with removals, 21.4% are covered by T&MC, 78.6% are covered by PBC. Among the pool of 763 engines, 305 of them (40%) had at least one unplanned removal during the observation period. These 305 units are associated with 48 different customers.

As mentioned in the previous section, our approach for capturing the impact of a contract type on product reliability focuses on unplanned removal events. Unplanned engine removals are highly undesirable events for aircraft owners since an aircraft on-the-ground generally results in high opportunity costs, with estimates as high as hundreds of thousands of dollars per day for an unplanned removal for a fully loaded wide body commercial aircraft. Furthermore, there is always a possibility that unplanned failure may lead to a catastrophic event. In contrast, in the case of planned removals the shop visit is programmed in advance, and the appropriate replacements can be scheduled to be available to avoid having an aircraft on-the-ground. Indeed, we have observed in practice that a release of a product on the scheduled completion date for a planned maintenance event has a high priority. Therefore, whether a removal is planned or unplanned largely determines the downtime of an engine and hence of an aircraft.

Hence, for the purpose of studying product reliability, our main analysis focuses on the sample of 305 engines with at least one unplanned removal. We note that, in the sample of 305 engines with unplanned removals, 21.3% of the engines are covered by T&MC, and 78.7% are covered by PBC, i.e., exactly the same proportions as observed in the full sample: 39.9% of the T&MC engines had unplanned removals, and 40% of the PBC engines had unplanned removals. The observed proportions suggest that focusing the analysis on unplanned removals does not generate, a priori, a sample bias (see Section 7 for further discussion of sampling issues and robustness checks). We note also that reliability is, by definition, a product level variable, i.e., what fails is an individual product. Consequently, the unit of analysis of interest in this research is an engine unit. As will be noted throughout, focusing on an engine unit allows us to capture several details associated with its reliability (initial condition of the product, product type, etc.). From the two data files described earlier, we are able to obtain and calculate several variables of interest which describe both the product and the customer. Table 1 provides definitions and descriptive statistics for the variables used in our analysis.

While there are several possible ways to approach the measurement of product reliability, we have chosen the mean time between unplanned removals (MTBUR) to be the measure employed in our main analysis since (1) it is in fact a key reliability metric that practitioners in the aerospace and other industries constantly monitor and (2) it can be computed from available data. MTBUR represents the average time that a product is used without the need for an unplanned removal for repair and maintenance purposes. Unlike some other metrics (such as mean time between removals, which includes both planned and unplanned removals), MTBUR is a good representation of the physical reliability *inherent* in the product since an unplanned removal event occurs only when an engine fails randomly, free of managerial interventions, unlike planned removals. In the rest of our discussion we use the terms *product reliability* and *mean time between unplanned removals* interchangeably.

Unit	No. obs.	Variable Name	Variable definition	Mean	Median	Std. Dev.	Min	Max
Engine	305	ini_age	Time since new in July 2002 (TSN(T_B))	3,231	2,962	2,760	0	12,916
Engine	305	final_age	Time since new in July 2007 (TSN(T_E))	10,424	11,094	4,819	363	21,600
Engine	305	nyears_em_intro	Years since engine model introduced	3.1	4	1.7	0	5
Engine	305	eng_avgflighttime	Average flight time TSN(T_E)/CSN(T_E)	1.22	1.10	0.27	0.82	2.50
Engine	305	n_unplanned_rem	No. of observed unplanned removals	1.21	1	0.46	1	3
Engine	305	n_planned_rem	No. of observed planned removals	0.52	0	0.64	0	3
Customer	48	fleetsize	No. of engine units registered at Rolls-Royce	40	8	106.2	2	593
Customer	48	fleetmix	No. of engine models registered at Rolls-Royce	1.65	1	1.0	1	5

Table 1: Definition of variables and descriptive statistics. TSN = Time since new, measured in flying hours. CSN = Number of cycles since new (a cycle is defined as the interval between a takeoff and a landing). The variables ini_age, final_age, and eng_avgflighttime are all measured in flying hours.

Although MTBUR is an appropriate metric of product reliability, calculating it still poses nontrivial issues for our analysis because unplanned removals – and, in fact, removals of any kind – are quite rare events in our data. In our observation period of 5 years, the majority of the products in the dataset (81.6%) exhibit only one unplanned removal; the remaining units had either two or three unplanned removals (16.1% and 2.3% of the sample, respectively). Additionally, the data suffers from censoring since information on any unplanned removals that occurred before July 2002 or after July 2007 is excluded. Defining a rule to compute the MTBUR, therefore, is a challenging task. We illustrate the problem and the procedure used to calculate the MTBUR with an example (see Figure 1).

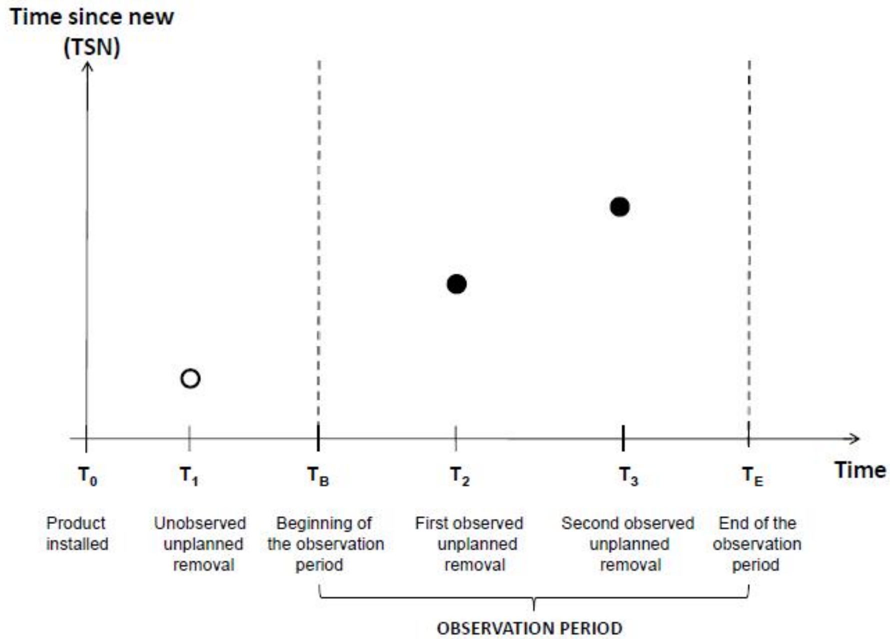

Figure 1: Example of an unplanned removal sequence for a product.

Consider a product that was installed in an aircraft at time T_0, before the beginning of the observation period T_B (July 2002 in our case). Assume that a first unplanned removal occurred at time $T_1 < T_B$, i.e., this event was unobservable to us. Suppose we observe the 2 unplanned removals at times T_2 and T_3, which occurred before the end of the observation period T_E. Let TSN(T) denote the time since new of a product at time T. (Note that TSN is measured in flying hours, i.e., hours of actual activity in the air, which is different from calendar time. In Figure 1, the former is shown on the y-axis and the latter is on the x-axis.) We do not observe the first unplanned removal and we do not even know if it took place or not. In other words, we only know the values of T_B, T_2, T_3, T_E, and the respective measures TSN(T_2), TSN(T_3), and TSN(T_E), but not the values of T_0 (the time at which the product was installed), T_1 (the time the first unplanned removal occurred), the corresponding flying hours TSN(T_1), and the *initial age* of the product at the beginning of the observation period TSN(T_B). We build our main proxy for MTBUR as follows (in section 7 we discuss other proxies for MTBUR and engine reliability):

$$\text{MTBUR}=\frac{\text{TSN}(T_E)-\text{TSN}(T_B)}{\text{Number of observed unplanned removals}}$$

Our proxy for MTBUR is thus defined by the inverse of the observed failure rate. In the example illustrated in Figure 1, it is equal to [TSN(T_E)- TSN(T_B)]/2. However, as we pointed out, the data do not include the value TSN(T_B). We compute an estimate for TSN(T_B), say TSN*(T_B), by assuming that there was a constant rate of usage for the product throughout the observation period. Specifically, we estimate this value as a linear projection of the line defined by the first observed removal and the age of the product measured at the end of the observation period, i.e., we estimate the slope of the line using the first removal and the end of the observation period as the two data points. We then project the line back to T_B in order to obtain an estimate of the initial age of the product defined as max{0, TSN*(T_B)}, which is measured in flying hours. We believe that this approximation provides a reasonable estimate for MTBUR. Note that if we omit subtracting the initial age of the engine, the proxy would overestimate the true MTBUR for all engines that had unplanned removals before the beginning of the observation period. The descriptive statistics for our MTBUR proxy are displayed in Table 2; recall that this variable is measured in flying hours.

	Overall sample		**T&MC only**		**PBC only**	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
MTBUR	6,456	3,278	6,016	3,844	6,575	3,106

Table 2: Descriptive statistics for the MTBUR.

Note that the mean MTBUR is slightly higher for PBC than for T&MC engines (by about 10%). An alternative approach to what is described above, i.e., computing the mean time between unplanned removals for each individual product and using it as a dependent variable in a regression model, would be to infer the mean time between unplanned removals (as an output of the analysis) by estimating the underlying distribution of the time between removals using techniques drawn from duration models (see, for example, Cameron and Trivedi 2005, Ch. 17). This approach – the analysis of which is discussed in Section 7 – has some econometric challenges on its own. In particular, it is difficult to account for the endogeneity of contract choices, a central feature of our model. Thus, in the main part of this paper, we focus our discussion on the results based on the first approach which allows us to use a well-established two-stage econometric framework that explicitly deals with endogeneity of contract choices.

5. Econometric Model

Our goal is to build a model that captures the effect of contract type on product reliability, measured by MTBUR. While the previous literature (Kim 2011, Kim et al. 2011) present separate analyses of PBC and T&MC (but under the same sets of conditions), in practice both contracts are offered to the customers simultaneously and therefore assignment of customers to contracts is not random. Instead, a customer may select a contract based on their private knowledge of engine usage that they possess and/or based on other firm characteristics. Thus, a major challenge associated with isolating the marginal effect of different contract types on product reliability is the inherent endogeneity associated with contract type choice by customers, which has been regarded as a key econometric issue in testing contract design hypotheses (Masten and Saussier 2002). A good illustration of the biases that can be generated by not accounting for self-selection in an application to firm entry and performance can be found in Shaver

(1998). General econometric discussions on the importance of accounting for self-selection and related methods can be found elsewhere (e.g., Heckman 1979, Maddala 1983).

To account for endogeneity in contract selection, we utilize a two-stage treatment effects model (see Maddala 1983, p. 120). This control function approach allows us to estimate the effect of a binary treatment (PBC) on a numeric outcome (product reliability), given that the treatment assignment is not random but rather is determined by an endogenous decision process. The approach utilizes a two-stage structure that involves a first stage to explain contract choice (Equation 2) and a second stage to explain product reliability (Equation 1).

$$\boldsymbol{y_i} = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{\delta z_i} + \boldsymbol{\varepsilon_i} \qquad \text{(Eq. 1)}$$

$$\boldsymbol{z_i} = \mathbf{1}(\boldsymbol{w_i\gamma} + \boldsymbol{v}_i > 0) \qquad \text{(Eq. 2)}$$

The observed reliability of product i, denoted by y_i, is explained by the exogenous covariates x_i and the binary endogenous variable z_i (that in our case is equal to 1 for products covered by PBC contracts and 0 otherwise). As is standard in discrete choice models with latent variable representation, e.g., probit, the binary variable for contract choice (z_i) is modeled as an indicator function, dependent on a set of exogenous covariates w_i, which drive the choice process. The error terms (ε_i, v_i) of the outcome and choice equations, respectively, account for unobservable characteristics which are allowed to be correlated, and are modeled as a bivariate normal random variable with distribution $N_2(0, 0, \sigma^2, 1, \rho)$; where the variance of v_i is normalized to one for identification purposes. If the correlation between both error terms is equal to zero then the outcome and choice equations can be estimated independently (Equation 1 could be estimated by OLS), i.e., the endogeneity of contract type is not relevant for the problem; otherwise, OLS estimation of Eq. 1 will produce a biased estimate for δ. For additional information on two-stage models based on control functions, the reader is referred to Maddala (1983, pp. 117-125).

In order to properly capture the main effect of our interest, i.e., significance and the magnitude of the coefficient δ in Equation 1, we need to specify the observable covariates influencing product reliability (x_i) and contract choice (w_i). There are characteristics of both the product and the customer (the user) that can play roles. Of those, we need to identify the most salient characteristics in order to avoid collinearity issues in our sample of 305 observations. Since we are not aware of other papers that analyze after-sales repair and maintenance contracts in the aerospace industry at the level required for our model, there is no precedent for many of the variables that we use, but we take clues from the reliability and contract theories, and from our in-depth knowledge of the industry.

An obvious factor that can influence the observed MTBUR is the initial condition of the product. Reliability theory (see Rausand and Høyland 2004) argues that very young and very old engines are more likely to have low reliability: younger engine units typically require adjustments at the beginning of their life, while older engines may fail more often due to part wear and tear caused by usage over time. To check these conjectures, we examined the distribution of the MTBUR for different ranges of initial product age and observed that MTBUR is lower for both *new* and *old* products, and is higher for *medium age* products, which is in line with the reasoning proposed above. In order to account for such nonlinearities, we include both linear and quadratic terms for the initial age of the product in our model specification. A polynomial function has been also used by Hubbard (1998), among others, to capture the effect of initial age in a related setting.

Another product characteristic that is related to engine reliability is the average flight time of an engine. For example, more take-off/landing cycles per flight hour may decrease the reliability of the engine since most of the wear and tear happens during these time periods. In our dataset there are records of the time since new (TSN) and the cycles since new (CSN) for each engine, measured at the end of the observation period. The average flight time for each engine is then the ratio TSN/CSN. We include this control variable in the outcome equation.

While the initial age of the engine and the average flight time capture a relevant part of both the initial conditions and usage patterns of the product, engine model characteristics can also affect reliability and should be controlled for. As mentioned in the data description, there are 2 engine models (5 versions in total) and 3 different aircraft models in our sample. Not every engine model can be installed in every aircraft model. In our data, there are three possible variables that help account for engine model effects: engine model dummies, aircraft model dummies, and a continuous variable indicating the number of years since the engine model was introduced into the market. They are strongly correlated and therefore cannot be included simultaneously in the model. Out of the three candidates, we choose to include the last variable in the model specification, as it captures the essence of the differences across products more parsimoniously in a single coefficient. This variable is a reasonable way to approach the effect of engine characteristics on product reliability, e.g., less is known about engine models that were introduced more recently to the market, which could influence reliability and how likely are these engines to be covered by PBC. In Section 7 we discuss models which use engine model dummies and aircraft model dummies, instead, but our results remain robust to these variations.

As we discussed in Section 3, a customer can also affect engine reliability, and thus customer characteristics need to be included in the specification of the outcome equation. For example, geographic location of the owner can be a proxy for closeness and availability of repair shops and for other local market conditions such as weather. We include dummy variables for the geographical region of the owner, categorized as U.S. vs. non-U.S. customers. We collect this information based on the customer ID, which we know from our database. Naturally, we would like greater granularity to capture local effects at the level of countries or even regions within a country. Recall, however, that we only have 48 customers in our sample, and so defining the geographical variable too narrowly would generate an identification problem due to low frequency data.

Further, we include the variables *fleetsize* (number of engine units that a customer has registered with Rolls-Royce) and *fleetmix* (number of different engine models in the portfolio of a customer), which serve as proxies for the customer characteristics that may be correlated with the engine reliability in different ways. For example, one may expect *fleetsize* to be positively correlated with reliability because a customer with a larger fleet is in a better position to request high quality services provided by Rolls-Royce. On the other hand, a negative correlation will be observed if *fleetsize* reflects usage patterns, in which case a customer with a larger fleet may use the engines with less care expecting a slack in engine capacity during low utilization periods. Similarly, both positive and negative correlations between *fleetmix* and reliability are possible. A positive correlation is likely if service priority goes to a customer who demands more attention because of the support needs of his diverse engine portfolio. Alternatively, one can hypothesize a negative correlation on the ground that a less diversified engine portfolio simplifies the engine operation and therefore leads to higher reliability. Whichever may be the outcome, it is clear that the customer characteristics captured in these variables can affect reliability. Hence, it is important to use this information in the model specification to control for customer effects. Note that an alternative way to capture the effect of customer characteristics on reliability is using a random (or fixed) effects model; this approach is discussed in Section 7.

Finally, another factor that could influence the time between unplanned removals is the occurrence of planned maintenance events (planned removals). Planned maintenance events have to follow a schedule that is prescribed by the Federal Aviation Administration, i.e., after every so many flying hours the engine must come in for a planned maintenance check. Of course, there is some discretion in following these rules: the airplane owners will often try to minimize schedule disruptions by arranging planned maintenance at the time when the aircraft is close to the repair facility and when its utilization is relatively low. All else being equal, we expect that the engines that have had more frequent planned maintenance to have lower need for unplanned removals, resulting in a higher MTBUR. Thus, we include in the outcome equation a variable indicating the number of observed planned maintenance events for a given engine.

Regarding the choice equation, we need to include covariates that influence the type of contract selected by a customer. Insights from the contract theory (e.g., Bolton and Dewatripont 2005) suggest that customers who intend to overuse engines and take poor care of them will be incentivized to opt for PBC contracts to begin with (i.e., adverse selection). That is, allocation and sharing of the risk induced by different contract types is one possible reason for self-selection by customers. Clearly, the two contract types of interest here have different implications for risk allocation: under PBC, the risk is shifted entirely towards the supplier, while under T&MC the risk is shifted towards the customer. Thus, PBC can be thought of as an insurance policy that creates predictable cash flows for the customer at a cost. Individual customers would then analyze this trade-off using their internal knowledge about their risk-aversion, and we expect that a customer who is more (less) tolerant to risk will opt for T&MC (PBC).

While it is notoriously difficult to find good proxies for risk-aversion, one of the commonly used proxies is the size of the company which, in the context of our setting, corresponds to *fleetsize*. According to supplier managers we interviewed, *fleetsize* is probably the most relevant variable to explain contract selection, as it is observed anecdotally that customers with a larger fleet are more likely to choose PBC for after-sales product support. This is related to the notion of risk we discussed above: the expected volume of cash flows needed for repair and maintenance services increases with the size of the customer fleet, which may cause larger firms to be more likely to sign on for PBC. This conjecture is in line with the data, which show that the median fleet size of T&MC and PBC customers are 2 and 12, respectively. In addition, we include the variable *fleetmix* as part of the choice equation specification in order to reflect the connection between the complexity of a customer's engine portfolio and his contract choice. For example, a customer may have internal expertise to service some engine types, but ownership of multiple engine types would require a complex mix of internal capabilities and therefore could lead the customer to opt for a comprehensive support arrangement through PBC. Alternatively, a more diversified fleet may be seen as a measure of risk diversification, which may lead *fleetmix* to be negatively correlated with the choice of PBC. Hence, the mix of the customer product portfolio needs to be controlled for.

Similarly, the location of the owner partially captures local characteristics such as prices, availability of repair shops, competition, and marketing efforts by Rolls-Royce that may influence the decision to choose PBC. Therefore, we have included geographical dummies in w_i, as we did for the outcome equation. Finally, another factor that can influence contract choice for a given engine is how complex it is to deal with maintenance for a given engine model. As discussed earlier, the number of years since an engine model was introduced to the market seems to summarize product characteristics that can be related to the type of customers that prefer either type of contract, for example, engine models for which there is little knowledge of maintenance procedures (i.e., newer models) could be perceived as riskier by customers, and therefore they could be more likely to be covered by PBC as a way to mitigate risk. We thus include this variable in the choice equation.

Note that our modeling approach allows for correlation between the unobservable terms of the outcome and contract choice equations. This feature is useful since we do not observe all variables related to the risk profile of the customer, which can influence both product reliability and contract choice. In particular, in our case we expect this correlation to be negative due to adverse selection and customer-side moral hazard, i.e., customers that are more likely to use engines more intensely (increasing failure risk) have more propensity to sign for PBC, which is hypothesized to increase reliability. Similar arguments have been discussed, for instance, in the case of extended warranties for new car buyers (Padmanabhan 1995, Hollis 1999), where it is argued that heavy users have stronger incentives than light users to sign on for extended warranties, since their products are more likely to experience failures. We also note that the control function approach we implement does not involve exclusion restrictions, in contrast to estimation based on instrumental variables methods. That is, all variables in w_i can be included in x_i. This is due to the assumption of the nonlinear selectivity term introduced as control in the outcome equation (see Eq. 3 below), which allows for consistent estimation of the δ coefficient in our problem under the maintained exogeneity assumption for covariates other than z_i (Maddala 1983 p.121). See Petrin and Train (2010) and Cameron and Trivedi (2005), which contain broader discussions on control function approaches for endogeneity correction.

6. Results and Analysis

We now turn to the estimation and results obtained for the two-stage model defined by Equations 1 and 2. The model can be estimated using the usual two-step procedure, defined e.g., in Maddala (1983, pp. 120-122). First, a probit model is estimated for the choice equation, where the probability that observation i receives "the treatment" (engine unit i covered by PBC) is given by $\Pr(z_i = 1 \mid w_i) = \Phi(w_i\gamma)$, where Φ is the cdf of the standard normal distribution Let ϕ be the pdf of the standard normal distribution. From the results of this first stage, a selectivity term is derived as follows:

$$\text{Selectivity term\_}i = \begin{cases} \dfrac{\phi(w_i\gamma)}{\Phi(w_i\gamma)} & \text{if } z_i = 1, \\ \dfrac{-\phi(w_i\gamma)}{1-\Phi(w_i\gamma)} & \text{if } z_i = 0. \end{cases} \qquad \text{(Eq. 3)}$$

The selectivity term calculated from the first stage is used as a regressor in the outcome equation, which allows for consistent estimation of the effect of the endogenous treatment, PBC, on product reliability, by accounting for the endogeneity in contract choice. We estimate the model in STATA using the aforementioned two-step procedure. For a reference, we also report results obtained from estimating the outcome equation using OLS, i.e., ignoring the endogeneity problem of contract choice. Table 3 displays the results obtained from the two-stage model and from the OLS model. We report clustered standard errors at the customer level in all cases, which are robust to correlation among the unobservable terms of observations from a given customer. In our problem, clustered standard errors at the customer level are of comparable magnitude to the ones obtained with the usual two-stage correction. In particular, our conclusions on how PBC affects reliability remain unaffected by the use of the common two-stage standard errors correction.

Our estimate for the effect of PBC on MTBUR is positive and indicates that, on average and all else being equal, PBC significantly increases engine reliability (p-value=0.003). Confidence intervals for the effect of PBC on reliability at the 90% and 95% confidence levels are given by [1154,3895] and [882, 4168], respectively. This provides support for the hypothesis of a positive and significant effect of PBC on product reliability. The model also indicates that the number of planned removals has a positive effect on product reliability: as expected, the more frequent the planned maintenance events, the less frequent the need for unplanned maintenance. Other product attributes with significant effects include the initial age of the engine and the number of years since the engine model was introduced. The effects found indicate that both linear and quadratic terms of the initial engine age are significant, in line with our arguments, and that engine models that were introduced earlier to the market are associated with a larger MTBUR, again in line with our reasoning. Similarly, the model indicates that customer characteristics also affect reliability: once controlled by contract choice, product characteristics, and planned maintenance, engines pertaining to U.S. customers with smaller fleet and greater fleet diversity, have greater MTBUR, relative to their counterparts.

TWO-STAGE MODEL				**OLS MODEL**
Choice equation		Outcome equation		
Variable	Coeff.	Variable	Coeff.	Coeff.
Fleetsize	0.024***	PBC	2525***	30
	(0.008)		(816.7)	(364.1)
Fleetmix	-0.436	n_planned_rem	1084***	1108***
	(0.383)		(257.9)	(250.2)
region_usa	-1.43*	ini_age	0.582**	0.629***
	(0.866)		(0.224)	(0.213)
nyears_em_intro	-0.206	ini_age_square	-0.00007***	-0.00007***
	(0.178)		(0.00002)	(0.00002)
Constant	1.57	eng_avgflighttime	861	555
	(1.088)		(965.3)	(973.1)
		region_usa	2562***	1876***
			(360.9)	(327.5)
		fleetsize	-4.32***	-0.967
			(1.557)	(0.965)
		fleetmix	298*	130
			(157.4)	(136.2)
		nyears_em_intro	1461***	1346***
			(166.7)	(177.4)
		selectivity_term	-1674***	
			(463.6)	
		Constant	-3969**	-1064

	(1732)	(1601)
χ^2(df)=9.75(4)	R^2=0.54, F=162.65	R^2=0.53, F=47.96

Table 3: Two-stage and OLS models results (N=305). Clustered standard errors (customer-level) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. df=degrees of freedom.

As discussed throughout the paper, an important feature of our model is that it controls for the endogenous nature of contract choices. Interestingly, as we show in Table 3, the results obtained by ignoring this endogeneity through using a simple OLS regression (i.e., assuming that the correlation between the error terms in the choice and outcome equations is zero) would suggest that there is no effect at all of PBC on product reliability. To the contrary, once the endogeneity is taken into account, our two-stage model predicts a positive and significant effect of PBC on reliability. In fact, in our two-stage model, the estimate for the correlation between the unobservable terms in the choice and outcome equations is negative (-0.69), and a likelihood ratio test rejects the null hypothesis of the correlation term being zero at the 5% significance level. Further, the selection correction term is significant at the 1% significance level. This evidence confirms the important role of accounting by the selection process in order to estimate the effect of PBC on product reliability in our application, illustrating that the model cannot be estimated by OLS due to the self-selection in the contract choice decision.

With respect to the choice equation, the results indicate that *fleetsize* and the geographical region have explanatory power for the contract type of an engine. In particular, our results show that engines owned by customers with greater fleet sizes are more likely to be covered by PBC than by T&MC, in line with our hypothesis, data, and managerial expectations. Our two-stage model estimates thus indicate that the factors influencing the likelihood of an engine being covered by PBC positively are in turn negatively correlated with engine reliability, which, in addition to the negative estimate of the correlation between the unobservables of the outcome and choice equations, is consistent with the argument of adverse selection. The first stage probit predictions that drive our modeling of the selection process are fairly reasonable. Comparing the proportions of engines covered by PBC vs. T&MC, the actual data in our sample reflects proportions of 78.7%/21.3% respectively (240/65 engine units), almost identical to our model's predictions of 79.3%/20.7% respectively (242/63 engine units). The model replicates the observed proportions remarkably well, and formal tests of equality of predicted and actual proportions cannot be rejected at any relevant significance level.

In summary, the results in this section provide support for our main hypothesis of a positive effect of PBC on reliability, once the endogeneity of contract choice is accounted for. One intriguing question that we have not addressed is: what is the mechanism by which PBC improves reliability? The results of our model are consistent with two possible mechanisms. First, our estimation results indicate that the frequency of scheduled maintenance is positively correlated with reliability. More frequent planned maintenance could be one mechanism by which PBC improves reliability. Summary statistics in Table 4 indicate that the frequency of scheduled removals is indeed higher under PBC.

	Overall sample		**T&M only**		**PBC only**	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
MTBPR	6,675	3,114	8,034	3,709	6,332	2,849
n_planed_rem	0.518	0.644	0.354	0.543	0.563	0.663

Table 4: Descriptive statistics for MTBPR[3] and number of planned removals.

Further, we note that without including the planned removals as a control in our empirical model (not reported in Table 3), we obtain a bigger and still highly significant coefficient on PBC, 2968, which suggests that part of the PBC effect is related to higher frequency of planned maintenance. Based on both observations, i.e., the descriptive statistics noted above and the increase in the magnitude of the PBC coefficient when planned removals are excluded from the model specification, our results suggest that more frequent scheduled maintenance under PBC is one mechanism by which reliability improvements may be achieved. However, as the estimation of our main model illustrates (Table 3), there is an effect of PBC on reliability even when we control for the number of scheduled maintenance events, which suggests that more frequent planned removals is not the only mechanism by which PBC influences reliability. Indeed, this observation is consistent with an alternative mechanism by which reliability is improved under PBC, namely, reliability improvement resulting from the supplier's provision of better care in each maintenance event. This could be achieved by various actions such as preemptive replacements of parts, more thorough identification of defects, and possible product re-design.

7. Robustness

In this section we examine the robustness of our main findings with respect to variations in a number of relevant model constructs, such as model specification, reliability proxies, modeling techniques, and sampling issues. Table 6 reports the results of some of these tests. The remaining tests, mentioned in this section but not reported in the table due to space constraints, are available upon request.

Model specification: The first issue we address is related to the variables used to control for product type. In our main model, we included the number of years since the introduction of an engine model to the market as the variable capturing product effects (nyears_em_intro), as it captures the product effects more parsimoniously while giving more accurate first stage predictions. As we mentioned in earlier

[3] The MTBPR (mean time between planned removals) is computed similarly to the MTBUR: the only change is the use of the number of observed *planned* removals in the denominator. The relevant sample for this calculation consists of the engines that had at least one planned removal. The sample for the statistics for the variable n_planed_rem consists of the 305 engines we analyze in our models.

sections, there are different ways to control for product characteristics, e.g., dummy variables to characterize aircraft model, engine model, or engine type. They cannot be included simultaneously in the model, as different proxies are almost perfectly collinear. As a robustness check, we estimated our model using alternate aircraft model dummies, dummies for the 5 engine model versions and dummies for the 2 main engine types. Our results remain qualitatively the same with these variations.

A second issue is the use of the initial age of the engine in the outcome regression. As we noted, this variable is not available to us, and we used extrapolation to obtain an estimate of it. A concern that can be raised is that we are using this variable as an input to calculate our dependent variable, as well as to calculate an independent variable. While we do not believe that this is critical in our case – our proxies for product reliability would be less meaningful if we did not take into account the initial age of the engine, to start with – we try a different proxy for capturing the initial conditions of the engine. We obtain the year of production of each aircraft from the Federal Aviation Administration (FAA) and related websites, matching the aircraft serial numbers in our dataset.[4] Using this information, we construct a variable reflecting the number of years since the aircraft was manufactured with respect to 2002 (the beginning of the observation period). While we believe that the initial age of the engine (measured in flying hours) is a more precise proxy for the initial conditions of the engine, the age of the aircraft can capture some relevant initial conditions of the engine in a similar fashion. We use age of the aircraft instead of initial age in Equation 2 to estimate our model. The results obtained, displayed in column 1 of Table 6, show that our results and conclusions are robust to this variation.

Another issue related to the model specification is the potential role of uncontrolled unobserved heterogeneity in driving our results. Given the clustered nature of our data (i.e., engines pertaining to different customers), it could be hypothesized that unobserved customer effects could have an influence in our estimates. As a way to explore the robustness of our results with respect to this issue, we included an unobserved heterogeneity term at the customer level in the contract choice and outcome equations, which is modeled as a draw from a normal distribution with mean zero and unknown variance, i.e., a two-stage random effects model. We find that our main results remain qualitatively the same under this variation (see column 2 in Table 6), and that the variance contributed by the unobserved heterogeneity term at the customer level explains only 2.7% of the total variance attributed to unobservable factors. In other words, most of the variance is already captured in the idiosyncratic error term at the engine unit

[4] We obtained data on this variable from aeronautic institutions from different countries. Websites consulted include: USA (FAA database, http://www.faa.gov), France (Direction Générale de l'Aviation Civile, http://www.immat.aviation-civile.gouv.fr/immat/servlet/aeronef_liste.html#), UK (Civil Aviation Authority http://www.caa.co.uk/application.aspx?catid=60&pagetype=65&appid=1), Sweden (Transport Styrelsen, http://www.luftfartsstyrelsen.se/templates/LS_LuftFartyg_Sok____39453.aspx), . For other countries like Brazil, Italy, Belgium, South Africa, Canada, Thailand, etc., we used different industry websites, e.g. http://www.airport-data.com, www.airframes.org, http://www.aerotransport.org/php/querybuilder.php?tab=regn, http://www.planespotters.net/Production_List/.

level, which was already considered in Equation 2. Furthermore, we also examine whether the PBC effect holds at the customer level. For that purpose, we estimate a two-stage between effects model, i.e., a regression on group means where each costumer defines a different group, which is an extension of the two-stage random effects model described above. We find that the PBC effect remains significant and is of comparable magnitude to the one reported in our main model, which suggests that the PBC effect is also present at the customer level. Altogether, these robustness checks suggest that unobserved heterogeneity is not a major concern for our conclusions.

Overall, for the robustness tests considered here, the PBC effect is significant at the 95% confidence level in all cases, with estimates in the range of [1655,2521] flying hours, which is in line with our main model. This represents a reliability improvement under PBC in the 25-40% range relative to T&MC.

Measurement of product reliability: The first issue here is the proxy we use to characterize product reliability. Consider again the example Figure 1 in Section 4. Given the nature of our data, there are several ways in which a proxy for product reliability could be defined, depending on how the right tail of the distribution is accounted in the calculations. In addition to the MTBUR variable defined in Section 4 and used in our models – which considers the right tail of the distribution in its entirety and therefore is a metric that is less sensitive to our estimation of the initial age of the engine – we explore different definitions of product reliability that deals with the right tail of the distribution in different ways. Based on the example in Figure 1, we define the following alternative proxies for MTBUR:

- MTBUR_proxy2 = $\frac{\text{TSN(Latest observed unplanned removal)} - \text{TSN}(T_{\text{B}})}{\text{Number of observed unplanned removals}}$
- MTBUR_proxy3 = average{TSN(T_2) – TSN(T_B), TSN(T_3) – TSN(T_2), TSN(T_E) – TSN(T_3)},
- MTBUR_proxy4 = max{MTBUR_proxy2, MTBUR_proxy3}.

Proxy 2 ignores the right tail of the distribution where unplanned removals did not occur and it is therefore also more sensitive to the estimated initial age of the engine, while proxies 3 and 4 incorporate that information in different ways than our main proxy. The descriptive statistics for these alternative MTBUR definitions are displayed in Table 5; all four proxies are measured in flying hours.

	Overall sample		**T&MC only**		**PBC only**	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
MTBUR_proxy2	3,328	2,458	2,872	2,632	3,451	2,400
MTBUR_proxy3	3,507	1,837	3,147	2,029	3,605	1,773
MTBUR_proxy4	4,079	2,227	3,609	2,536	4,206	2,123

Table 5: Mean and standard deviation of alternative reliability proxies.

While the values in Table 5 differ across the proxies (reflecting the distinct ways in which they account for the right tail of the distribution), in all cases we see that MTBUR of PBC engines is slightly higher than MTBUR of T&MC engines. We run our two-stage models using these MTBUR proxies as dependent variables, and find that the results that are consistent with the ones from our main MTBUR proxy. In particular, the PBC effect remains significant at 95% and 99% confidence levels in all cases (see columns 3, 4 and 5 in Table 6).

A second consideration related to the measurement of product reliability is the cross-sectional modeling approach we have used. Thus far, our analysis has relied on constructing proxies for MTBUR of each engine unit and performing two-stage estimations. As mentioned earlier, a duration model offers an alternative way to analyze our data. In fact, the sample in our dataset represents *multiple spells* data since some products have more than one unplanned removal. The main advantage of this modeling approach is that it inherently accounts for some data censoring issues, which are common in duration data such as ours. Unfortunately, dealing with the endogeneity issue in the context of duration models is difficult, in the sense there are no sufficiently established techniques that we could employ in our problem to convincingly account for the endogeneity of contract choices (research in this area is ongoing, see Bijwaard 2007 for a recent contribution to this research stream).

VARIABLES	(1)	(2)	(3)	(4)	(5)
PBC	2521.2***	1817.6***	1476.7**	1257.7***	1391.6***
	(865.3)	(484.9)	(605.0)	(429.2)	(514.6)
n_planned_rem	1002.9***	1070.6***	484	700***	885***
	(256.7)	(260.7)	(295.8)	(150.4)	(216.8)
ini_age		0.567**	0.222	0.224	0.324*
		(0.225)	(0.154)	(0.154)	(0.163)
ini_age_square		-0.00007***	-0.00001	-0.00002	-0.00003
		(0.00002)	(0.00002)	(0.00002)	(0.00002)
aircraft_age	575**				
	(234.1)				
aircraft_age_square	-26**				
	(11.91)				
eng_avgflighttime	1137.8	636.7	1367**	716	1111
	(1,163)	(932.6)	(679.5)	(583.9)	(670.2)
region_usa	2781.5***	2286.3***	958***	1156.7***	1394.3***
	(331.7)	(386.4)	(355.9)	(182.7)	(221.3)
fleetsize	-4.7**	-2.5**	0.5	-1.1	-0.4
	(1.896)	(1.242)	(1.382)	(1.046)	(1.105)
fleetmix	168	212	-22	63	41
	(191.2)	(158.9)	(148.0)	(98.15)	(116.4)
nyears_em_intro	1565.3***	1230.8***	681***	775***	863***
	(163.1)	(172.2)	(113.1)	(100.2)	(112.7)
selectivity_term	-1601.7***	-1054.9***	-1016**	-839***	-953***
	(494.7)	(204.4)	(411.7)	(241.9)	(337.1)
Constant	-6154.1***	-2483.1	-3025.4*	-2219.7**	-2953.1**
	(1,824)	(1,543.4)	(1,547)	(1,077)	(1,351)
	R^2=0.54, F=158.75	R^2=0.54, χ^2(df)=1128(10)	R^2=0.25, F=42.98	R^2=0.54, F= 145.88	R^2=0.49, F=166.54

Table 6: Selected robustness checks. All columns display results for the outcome equation of the two-stage model considered in each case. (1): Dependent variable=MTBUR; (2): Dependent variable=MTBUR, the model includes a random effect at customer level; (3): Dependent variable=MTBUR_proxy2; (4): Dependent variable=MTBUR_proxy3; (5): Dependent variable=MTBUR_proxy4. Clustered standard errors (customer-level) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

However, the following informal approach has been used by some researchers: (1) run the probit model (Equation 2) to predict contract choices, (2) calculate the selectivity term from this analysis (Equation 3), and (3) perform duration analysis using the selectivity term as one of the regressors. A similar approach in the context of sample selection for duration models was used by Rao et al. (2001), based on the generalization of the Heckman selection model proposed in Lee (1983). Although consistency of this approach is not, to our knowledge, fully established, we use this procedure in this subsection as a robustness check, which can be taken as complementary evidence with the caveat on endogeneity as described above. In order to proceed with this approach, we analyze the data at the removal level (instead of at the engine unit level), and we must examine the influence of contract type on the respective unplanned removal rates. As is standard in duration analysis, we conduct experiments using both semi-parametric (Cox) and parametric (exponential, log-logistic, and log-normal) transition rate models. The explanatory variables are the same as the ones used in the two-stage regression models; the only modifications in explanatory variables we incorporate – in order to take advantage of the multiple spell nature of the data – are to replace the initial age of the product with the age of the product at the beginning of each spell in the outcome equation and to use the number of observed planned removals in each spell, instead of in the full observation period. For example, for the spell associated with the occurrence of the 3rd unplanned removal of a product unit, instead of the initial age (age of the product at the beginning of the sample period) we include the age of the product unit at the moment of the 2nd removal (for both the linear and quadratic terms). Note that this also makes these models much less sensitive to our estimation of the variable ini_age, which may serve as additional evidence of the robustness of our results to that issue. We estimate the models using clustered standard errors at the customer level. We obtain largely the same findings obtained with the two-stage models, in particular, for the effects of PBC on product reliability, regardless of the modeling technique employed (semi-parametric and parametric effects defined above).

Sampling issues: The first issue is the low frequency of unplanned removals in our data. As we indicated earlier, for a significant part of our sample we observe only one unplanned removal. As a robustness check, we run our model for the sample of engines which had at least two unplanned removals (ignoring engines with only one unplanned removal). We obtain coefficients for the PBC effect similar to the ones obtained for the whole sample. Some significance is lost, however; under clustered standard errors at the customer level we obtain a p-value for the PBC effect of 0.124, and a p-value of 0.058 under non-clustered standard errors. It is reasonable to expect some reduction in the significance levels in this case, as the number of engine units with at least two removals is very small (56 engine units). We repeat the exercise for duration models, which have the advantage of accounting for each spell separately (thus alleviating the problem of dealing with such a small sample size to some extent), and again obtain similar PBC effects which are now significant at the 95% confidence level. While we do not attempt to make inferences based on this small sample of 56 engine units, these exercises alleviate potential concerns regarding the role of low frequency of unplanned removals in our analysis.

Perhaps a more fundamental sampling concern is that our definition of MTBUR for the two-stage model is conditional on the occurrence of at least one unplanned removal. This means that engines without observed unplanned removals are not considered in the estimation of our model. Naturally, this is driven by the nature of the dependent variable, mean time between unplanned removals, which is undefined if unplanned removals did not occur at all. As we pointed out, the concern of a potential bias in our sample due to this issue is alleviated by the fact that the observed proportions of PBC and T&MC engines in the sample of 305 engine units for which we observe unplanned removals is the same as the proportion observed in the full sample of 763 engines. Duration models, however, allow for the possibility of estimating the model including those engine units for which there are no observed events (unplanned removals), by incorporating the associated spells into the likelihood function. We estimate duration models for the full sample of 763 engines, using the same procedure and modeling assumptions described earlier, and obtain largely the same findings as in the case of the 305 engine unit sample. This alleviates concerns regarding the potential role of the conditional nature of the sample used in our analysis.

8. Conclusions and Discussion of Limitations

Our analysis suggests that product reliability improvement is achieved under PBC. Our estimates indicate that, in comparison to traditional T&MC, there is a positive and statistically significant effect of PBC on the mean time between unplanned removals (MTBUR) of a product and that its magnitude is in the 25-40% range. We also show that endogeneity of contract choice, which has not been either modeled or discussed previously in this context, is clearly an issue here. Indeed, without explicitly accounting for this endogeneity, the significance of the PBC effect disappears. These findings are supported by numerous robustness checks under a number of alternative model specifications and modeling approaches. The results obtained from our analysis provide a first step towards understanding the overall impact of PBC on product reliability.

Although our analysis shows firm evidence that supports these conclusions, it is not free of limitations. One of them is our model specification that allows us to explicitly deal with the endogeneity of contract choices at the expense of treating the rest of the covariates as exogenous. As a side effect, this approach precludes us from exploring a potential correlation between the frequency of planned removals and the unobservables in the outcome equation. This, however, is not unreasonable since planned removals are subject to strict regulation in the aerospace industry, (although there may be some discretion by aircraft owners in this regard). In addition, there are several issues stemming from the nature of the data. First, our ability to accurately measure engine reliability is limited by the low frequency of failures in our sample. Failures of aircraft engines are rare events, even though we observe the system for the relatively long period of 5 years. This problem creates some imprecision in the definition of the dependent variable, MTBUR. We deal with this issue by conducting analysis based on several proxies for MTBUR and using alternative modeling approaches (two-stage models, duration model analysis), and find that the results are consistent overall. Second, while our dataset is rich in terms of characterizing the removal incidents for a given product, we have only limited data to characterize a customer (such as the customer risk profile) that may impact the results. To alleviate this concern, we included in our analysis such variables as fleet size, fleet mix, and the region of the owner, as a way to account for some relevant customer characteristics. Third, our dataset does not provide the detailed characteristics of the contracts; partly because of this, we only distinguish between T&MC and PBC. This limitation prevents us from exploring, for example, the influence of price and contract length on the customer's contract choice. We also do not observe data before and after the adoption of PBC, which would have made it possible to study how incentives evolve over time. However, we tested whether the influence of unobserved heterogeneity could influence the results, and found that the PBC effect remains remarkably robust, which suggests that the influence of these unobserved factors should not affect our results significantly.

Lastly, it is important to note that the goal of our research is driven in large part by data availability. For instance, our analysis does not address cost implications of PBC because cost data were not made available to us. Undoubtedly, an access to a richer set of data including financial and managerial information, other performance metrics such as product availability, and contract term details would enable a more complete analysis with a larger scope than the current study offers, leading to a deeper understanding of the benefits and downsides of PBC. Despite the limitation, we believe that our study provides valuable insights to practitioners who are continuously striving to achieve the highest level of product reliability.

In this paper we present one of the few studies that empirically estimate the impact of contracting on supply chain outcomes. While we cannot claim that the conclusions obtained in this study are applicable to all supply chain settings, our findings are of interest, not only to the aircraft repair and maintenance industry but also to all industries in which the quality of manufactured products is an important driver of firm performance. We also believe that this study is practically relevant and timely because it is the first empirical investigation that rigorously tests the hypothesis that reliability improvement is achieved under PBC, an issue of considerable interest among practitioners who are currently weighing the costs and benefits of adopting PBC-based relationships for after-sales product support.

Acknowledgement

The authors are grateful for the suggestions made by the anonymous reviewers, associate editor, and department editor, and by the participants at the numerous conferences and universities where versions of this paper have been presented. We would also like to acknowledge the following individuals for their input: Vibhanshu Abhishek, Richard Lai, Donald Lee, Marcelo Olivares, Justin Ren, Rachna Shah, Enno Siemsen, Manuel Sosa, and Owen Wu. This research was partially funded by the Fishman-Davidson Center for Service and Operations Management, The Wharton School, University of Pennsylvania and by a grant from the US Air Force (in collaboration with the University of Tennessee). Support for Guajardo was also provided by the Wharton Risk Center's Russell Ackoff Doctoral Student Fellowship and by the Joseph M. Juran Center for Leadership in Quality. Finally the authors would like to acknowledge the active participation in the project by the Rolls-Royce Corporation, with special thanks to Miles Cowdry, W. Bunker Snyder, James J. Olsen, Lindsey E. Shackelford, and Jarod Downing for their assistance in collecting data and providing input to our understanding of their PBH program.

References

Ackerberg, D., M. Botticini. 2002. Endogenous matching and the empirical determinants of contract form. *Journal of Political Economy*, 110(3), 564-591.

Bolton, P., M. Dewatripont. 2005. *Contract Theory*. The MIT Press, Cambridge, MA.

Cachon, G. 2003. Supply chain coordination with contracts. In: S. Graves and T. de Kok, Editors, *Handbooks in Operations Research and Management Science: Supply Chain Management*, North Holland, Amsterdam, The Netherlands.

Cameron, A. C., P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Chiappori, P.-A., B. Salanie. 2003. Testing contract theory: a survey of some recent work. In *Advances in Economics and Econometrics,* M. Dewatripont, L.P. Hansen and S.J. Turnovsky, editors. Volume 1, Cambridge University Press.

Clark, N., J. Mouawad. 2010. In Qantas blowout, concern for jet engine maker. *The New York Times* (November 5). http://www.nytimes.com/2010/11/06/business/global/06engine.html. (Retrieved: July 2011.)

Cohen, M. A., N. Agrawal, V. Agrawal. 2006. Winning in the Aftermarket. *Harvard Business Review*, 84(5), 129-38.

Corbett, C.J. and X. de Groote. 2000. A supplier's optimal quantity discount policy under asymmetric Information. *Management Science*, 46(3), 444-450.

Department of Defense. 2005. DoD guide for achieving reliability, availability, and maintainability. http://www.acq.osd.mil/dte/docs/RAM_Guide_080305.pdf.

Department of Defense. 2009. DoD weapon system acquisition reform: Product support assessment. https://acc.dau.mil/CommunityBrowser.aspx?id=328610.

Deshpande, V., M. A. Cohen, K. Donohue, 2003. An empirical study of service differentiation for weapon system service parts. *Operations Research*, 51(4), 518–530.

Einav, L., A. Finkelstein. 2011. Selection in insurance markets: theory and empirics in pictures. *Journal of Economic Perspectives*, 25(1), 115-138.

Geary, S., K. Vitasek. 2008. *A Contractor's Guide to Life Cycle Product Support Management*. Supply Chain Visions.

Government Accountability Office. 2008. Improved analysis and cost data needed to evaluate the cost-effectiveness of Performance Based Logistics. GAO-09-41.

Gopal, A., K. Sivaramakrishnan, M. S. Krishnan, T. Mukhopadhyay. 2003. Contracts in offshore software development: An empirical analysis. *Management Science*, 49(12), 1671–1683.

Gopal, A., K. Sivaramakrishnan. 2008. On vendor preferences for contract types in offshore software projects: the case of fixed price vs. time and materials contracts. *Information Systems Research*, 19(2), 202-220.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.

Ho, J., K. Ho, J. Mortimer. 2011. Welfare effects of full-line forcing contracts in the video rental industry. Working paper.

Ho, K., J. Ho, J. Mortimer. 2011. The use of full-line forcing contracts in the video rental industry. Working paper.

Hollis, A. 1999. Extended warranties, adverse selection, and aftermarkets. *The Journal of Risk and Insurance*, 66(3), 321-343.

Holmström, B., P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization,* 7, 24–52.

Hubbard, T. N. 1998. An empirical examination of moral hazard in the vehicle inspection market. *The RAND Journal of Economics*, 29(2), 406-426.

Hypko, P., M. Tilebein and R. Gleich. 2010. Benefits and uncertainties of performance-based contracting in manufacturing industries. *Journal of Service Management,* 21(4), 460-489.

Iyer, A.V., L.B. Schwarz, S.A. Zenios. 2005. A principal-agent model for product specification and production. *Management Science,* 51(1), 106-119.

Kim, S.-H. 2011. Strategic reliability investments in multi-indenture supply chains. Working paper, Yale University.

Kim, S.-H., M. A. Cohen, S. Netessine. 2007. Performance contracting in after-sales service supply chains. *Management Science*, 53(12), 1843-1858.

Kim, S.-H., M. A. Cohen, S. Netessine. 2011. Reliability or inventory? Analysis of product support contracts in the defense industry. Working paper, Yale University.

Kirk, R. L., T. J. DePalma. 2005. Performance-based logistics contracts: a basic overview. CNA Corporation. November 2005. Available at http://www.cna.org/documents/D0012881.A2.PDF (accessed April 23rd 2009).

Lazear, E. P. 2000. Performance pay and productivity. *American Economic Review*, 90(5), 1346-1361.

Lee, L. F. 1983. Generalized economic models with selectivity. *Econometrica*, 51(2), 507-512.

Lu, M., C. Ma, L. Yuan. 2003. Risk selection and matching in performance-based contracting. *Health Economics*, 12(5), 339-354.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press. Cambridge, UK.

Masten, S., S. Saussier. 2002. Econometrics of contracts: An assessment of developments in the empirical Literature on contracting. Ch.16, In: *The Economics of Contracts-Theories and Applications* (E. Brousseau and J-M Glachant, eds.), Cambridge University Press.

Mortimer, J. H. 2008. Vertical contracts in the video rental industry. *Review of Economic Studies*, 75 165-199.

Newsome, Jon. 2008. Aviation Week Performance-Based Logistics Survey. Available at http://thecenter.utk.edu/images/Users/1/PBL/WhitePapers/AviationWeekPBLSurveyConductedforSAPAmericas.pdf. (Retrieved: November 2010.)

Novak, S., S. D. Eppinger. 2001. Sourcing by design: Product complexity and the supply chain. *Management Science*, 47(1) 189-204.

Novak, S., S. Stern. 2008. How does outsourcing affect performance dynamics? Evidence from the automobile industry. *Management Science*, 54(12), 1963-1979.

Novak, S., S. Stern. 2009. Complementarity among vertical integration decisions: evidence from automobile product development. *Management Science*, 55(2), 311-332.

Padmanabhan, V. 1995. Usage heterogeneity and extended warranties. *Journal of Economics & Management Strategy,* 4(1), 33-53.

Petrin, A., K. Train. 2010. A Control Function Approach to Endogeneity in Consumer Choice Models. *Journal of Marketing Research*, 47(2), 3–13.

Prendergast, C. 2002. Uncertainty and incentives. *Journal of Labor Economics*, 20(2), 1071-1102.

Ramdas, K., T. Randall. 2008. Does component sharing help or hurt reliability? An empirical study in the automotive industry. *Management Science*, 54(5), 922-938.

Rao, H., H. R. Greve, G.F. Davis. 2001. Fool's gold: Social proof in the initiation and abandonment of coverage by Wall Street analysts. *Administrative Science Quarterly*, 46(3), 502-526.

Randall, W.S., T.L. Pohlen, J.B. Hanna. 2010. Evolving a theory of performance-based logistics using insights from service dominant logic. *Journal of Business Logistics,* 31(2), 35-61.

Rausand, M., A. Høyland. 2004. *System Reliability Theory: Models, Statistical Methods, and Applications*. Wiley, New Jersey.

Rothschild, M., J. Stiglitz. 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90(4), 629-649.

Shaver, J. M. 1998. Accounting for endogeneity when assessing strategy performance: does entry mode choice affect FDI survival? *Management Science*, 44(4), 571–585.

Shen, Y. 2003. Selection incentives in a performance-based contracting system. *Health Services Research*, 38(2), 535-552.

Standard & Poor’s. 2011. Industry Surveys: Aerospace and Defense (February 10th).

Terwiesch, C., J. Ren, T. H. Ho, M. A. Cohen. 2005. An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Science*, 51(2), 208-220.

Thomas, T. J. 2005. Performance logistics now coming to Australia. *Australian Defense Business Review*, 24(7), 23-24.