5-2012

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

Jose Guajardo
University of Pennsylvania

Morris. A. Cohen
University of Pennsylvania

Serguie Netessine

Recommended Citation

Guajardo, J., Cohen, M. A., & Netessine, S. (2012). Impact of Performance - B ased Contracting on Product Reliability : An Empirical Analysis. *Management Science, 58* (5), 961-979. http://dx.doi.org/10.1287/mnsc.1110.1465

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

Abstract
Using a proprietary data set provided by a major manufacturer of aircraft engines, we empirically investigate how product reliability is impacted by the 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.

Keywords
reliability, maintenance repairs, empirical operations management, supply chain contracting, aerospace industry

Disciplines
Other Engineering

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

Jose A. Guajardo ● Morris A. Cohen ● Serguei Netessine

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

Sang-Hyun Kim

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

July 15th, 2009

Abstract

Using a proprietary dataset provided by a major manufacturer of aircraft engines, we empirically investigate the impact of incentives present in after-sales repair and maintenance support contracts on product reliability. In particular, we compare the reliability of products under time and material (T&M) contracts, which have been used traditionally in the airline industry, with the reliability under performance-based contracts (PBC), which are gaining wide acceptance. We discover that there is inherent endogeneity in contract choices by the customers. To account for this endogeneity, we estimate a two-stage econometric model, and find that larger customers and users of certain equipment types are more likely to select PBC over T&M contracts. After controlling for this selection process, we find evidence for the positive and significant effect of performance incentives created by PBC on product reliability. Our estimates indicate an improvement of product reliability in the 10-25% range under PBC, compared to the reliability observed under T&M contracts. This finding is robust to numerous alternative specifications, modeling assumptions and estimation methods. Thus our research provides a valuable input into the ongoing policy debate about the effectiveness of performance contracts which are currently being introduced extensively in both the government and the private sectors.

1. Introduction

The move to performance-based contracting (PBC) represents a fundamental shift in customer-supplier relationships for after-sales product support. A PBC contract will pay the service provider on the basis of product use or "up-time" [1], whereas under traditional Time and Material (T&M) contracts the customer pays the supplier for resources consumed due to the occurrence of product failures and maintenance events. The former approach, which in some industries is referred to as "servicization" (see, for example, Toffel 2002), essentially converts the sale of service products (such as spare parts and repair labor) to the customer into the sale of a service that enables the customer to generate value through the use of the product.

While the provision of after-sales support is a major driver of revenue and profit in many industries, the movement to PBC is especially relevant for after-sales product support in the aerospace and defense industry where products are "mission critical", since their unavailability due to either scheduled or unscheduled maintenance can be very costly. The movement toward performance-based servicization is motivated by the premise that PBC aligns incentives between the customer and the supplier, so that both benefit when the product's use generates value to the customer. As a consequence, most observers expect that the adoption of PBC will lead to more effective value creation, i.e., with a higher level of performance and a lower cost to the customer, as well as a higher level of profit to the supplier of support services. A growing body of literature (e.g., Kim et al. 2007, 2008, 2009) analyzes this issue from an economic and operations modeling perspective. The results of that research indicate that it is possible to design coordinating contracts based on performance and that under such contracts, suppliers and customers have a strong incentive to increase product availability through various means that include improving support capabilities, investing in an appropriate level of support resources, and improving the underlying reliability of the products. While the managerial and analytical arguments for PBC are pervasive, empirical research to support these conclusions is currently non-existent.

The evidence that PBC for aftermarket support and system sustainment has a positive impact on performance primarily originates from industry reports that are not based on rigorous analysis. These reports and conventional wisdom suggest that PBC improves service outcomes,

---

[1] A typical performance-based contract will include terms that lead to payment to suppliers based on the number of hours the product has been used and/or the level of availability of key resources.

i.e., it can lead to higher product availability, lower cost of ownership, and reduced customer wait times. Kirk and DePalma (2006), for example, analyze a PBL[2] program in the Navy and, based on a review of historical repair frequency data for several programs, concluded that "there is some evidence that the PBL contract may have helped to improve availability and reliability". Not surprisingly, questions have been raised about the quality of the data and the associated analysis presented in such reports. For example, in a recent report the Government Accountability Office (GAO) states that, "Many DOD program offices that implemented PBL 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).

The need for empirical research to better understand the effects of PBC is particularly relevant considering not only the paucity of scientific evidence, but also the economic impact associated with after-sales support services. For example, according to Standard and Poor's (2009), the global maintenance, repair and overhaul sector generated revenues of $117 billion in 2007, of which $45 billion relates to commercial aircraft. 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 on assets they already own. The investment in resources required to enable the delivery of services to support products is also huge, e.g. spare parts inventory investment has been observed to be 5% of sales in computer and high technology industries (Cohen et al. 1997).

The findings of this paper are especially timely as performance-based contracts for after-sales support have become increasingly popular in industries such as aerospace and defense, automobile, semiconductors, information technology and software development (e.g., Software as a Service). The adoption of contractual relationships based on performance for after-sales support and other services also spans the public sector. In the U.S., the federal government, and the Defense Department in particular, has mandated this form of contracting for services on a wide-spread basis. Nonetheless, there is an ongoing debate between suppliers of support services and the various federal agencies who are engaged in implementing performance-based

[2] Performance-based contracting is also referred to as Performance-based Logistics (PBL) and Power by the Hour® in the defense and commercial aerospace industries, respectively. The latter is a registered Rolls-Royce trademark.

programs, regarding the value of contractual relationships based on performance. In addition, the GAO, as noted above, has questioned the accuracy of predictions of the positive impact of PBC and the House of Representatives recently held a hearing on the benefits and costs of PBC.[3]

In this paper, we focus on repair and maintenance services for commercial aircraft in the aerospace and defense industry where PBC for after-sales support has been in place for many years. Our proprietary dataset comes from Rolls-Royce, which, as a major supplier of aircraft engines, provides repair and maintenance services to its customers under two different types of contracts: T&M and PBC. The main question we analyze is the following: Does the use of PBC have a positive effect on product reliability over the use of T&M? Our focus on reliability is grounded on a theoretical prediction that reliability is likely to be impacted by the choice of contracts (Kim et al. 2008). Intuitively, T&M does not create a strong incentive to improve reliability, as the supplier's compensation under T&M is proportional to the amount of consumed service resources, which decreases as products fail or are serviced less often (i.e., products become more reliable). PBC, on the other hand, promotes investing in reliability improvement as doing so leads to higher product up-time, which in turn brings larger financial gains to the supplier. While a linkage between product reliability and performance-based incentives is seemingly intuitive, prior to this paper it has never been verified empirically, and the extent of this relationship has never been estimated. Our analysis, which uses a two-stage framework that explicitly deals with the endogeneity inherent in contract choice by a customer, provides evidence that PBC, indeed, improves product reliability. Our results quantify the observed benefits – about a 10-25% reliability improvement – that these contracts have generated in our sample. These findings are robust to a number of specifications and modeling assumptions. While the focus of this paper is on PBC for commercial aircraft after-sales services, our findings are also relevant to other industries that provide after-sales support for mission critical products.

The paper is organized as follows. In Section 2, we review relevant literature on supply chain contracting. In Sections 3 and 4 we describe the industrial context and the data, and continue with presentation of the econometric analysis and the model specification in Section 5. In Section 6 we discuss estimation results, robustness checks, and limitations of our analysis. Section 7 concludes the paper.

---

[3]"Performance-based Acquisitions: Creating Solutions or Causing Problems?" Full Committee Thursday, May 08, 2008. http://homeland.house.gov/hearings/index.asp?ID=136.

2. Literature Review

There has been substantial interest in the operations management literature concerning the role of contracts in supply chains. Cachon (2003) provides an extensive and relatively recent review of more than 200 papers in this area. As noted there, “the literature contains a considerable amount of theory, but an embarrassingly paltry amount of empiricism.” Although papers such as Novak and Eppinger (2001) and Novak and Stern (2008, 2009), for example, empirically examined the impact of product characteristics on vertical integration decisions in the automobile industry, the numerous theoretical predictions in the OM literature related to contractual incentives, for the most part, have not received comparable empirical scrutiny. The findings of this paper, contribute to filling this gap in the literature by providing an empirical validation of the hypothesis generated by an earlier theoretical study (Kim et al. 2008).

In contrast to the current state of empirical OM research on supply chain contracting, researchers in other areas have been quite active in studying similar topics. In economics, several papers have studied revenue sharing contracts in the video rental industry. Mortimer (2008), for example, analyzes fixed-price vs. revenue sharing contracts between distributors and retailers using a structural equation econometric approach. Her results indicate that both downstream and upstream firms’ profits for popular titles increase by 10% under revenue sharing contracts and consumers are also generally better off. Ho et al. (2008) and Ioannou et al. (2009) investigate other aspects of revenue sharing contracts in this industry. In addition, several empirical papers examine franchising contracts (e.g., Lafontaine and Shaw 1999, Lafontaine and Slade 2001). Finally, there is a stream of literature that studies incentive alignment between firms and their employees by means of different variants of pay-for-performance contracts. Examples include Prendergast (1999, 2002a, 2002b), Lazear (2000) and Banker et al. (2001). Although this research on performance-based contracts is distinct from ours because of its focus on a labor setting, we note two results from this literature that are relevant to our study: (i) employees respond to incentives which usually improve firm performance, and (ii) there is self-selection of better employees through pay for performance schemes. The findings of our paper offer similar insights in the firm-to-firm aftermarket support setting.

PBC has also received attention in application areas such as health care, public policy, and software development. The empirical results in Lu et al. (2003) show that PBC leads to more referrals and a better match between illness and treatment intensity, suggesting that PBC induces incentive alignment. Shen (2003) finds that PBC provides incentives for nonprofit providers of substance abuse treatment to select less severe patients into treatment. Other examples from the public sector include Heinrich (2002), who analyzes outcome-based performance management using data from federal job-training programs and a case study by Heinrich and Choi (2007), based on social welfare programs in Wisconsin, showing that the service provider responded to the incentives afforded by PBC.

In the context of offshore software development, Gopal et al. (2003) analyze the impact of fixed-price vs. T&M contracts on software vendor profits. As we do, they use a two-stage approach that includes modeling of the determinants of contract choice in the vendor-developer relationship. Their results indicate that vendor profits are higher under T&M contracts, controlling for variables such as project type and effort. Gopal and Sivaramakrishnan (2008) extend this analysis further by examining the impact that factors such as project size and duration, team size, and risk of employee attrition, have on contract choice.

The papers noted above study the influence of contracts on different performance outcomes. In our research, an equally important issue is determining the factors that influence contract choice. A number of papers examine contract choice in different contexts. Examples include Slade (1996), who studies contracts between oil companies and their service stations and tests various multitask agency hypotheses as drivers of contract choice. Ackerberg and Botticini (2002) examine contract choices using archival data on agricultural contracts between a landlord and tenants, and test hypotheses related to the role of risk sharing and transaction costs on contract selection. Using data from U.S. Air Force engine procurement, Crocker and Reynolds (1993) analyze how different variables affect the degree of contract completeness (i.e., how precisely the contract specifies future duties and contingencies) that parties choose.

Our paper also contributes to the growing stream of empirical research in operations management. Papers that are related to ours include Ramdas and Randall (2008), who find that component sharing in the automotive industry can hurt product reliability. We also study reliability empirically but in a different industry and in the contracting choice context. In an empirical study that focuses on after-sales support for defense systems, Deshpande et al. (2003) analyze how the interaction of attributes such as the criticality and cost of service parts and the nature of inventory policies to manage them affect performance. As in our case, this is an empirical study in an aftermarket context, although it is not focused on reliability or contracting.

Terwiesch et al. (2005) empirically analyze demand forecast sharing in the buyer/supplier contractual relationship in an application to the semiconductor industry. Their analysis indicates that non-optimal gaming behavior among all parties occurs as a consequence of conflicting incentive schemes. While this paper is not about contracting, it is related to our research since it empirically studies the role of incentives in an operational context. Finally, in a study related to PBC, Lee and Zenios (2007) develops a performance-based payment system for Medicare by empirically estimating data from patients needing kidney dialysis.

To summarize, we note that while fields such as economics, public policy, information systems, and healthcare have generated many examples of empirical research on the role of performance contracting, this has not been the case in operations management, despite significant attention that has been given to supply chain contracting research. Thus, in addition to providing a scientific input to the ongoing policy debate on performance incentives for government and defense service procurement, our paper contributes to the existing OM literature as it represents one of the few empirical studies of supply chain contracting. As a result, we believe that our paper contributes to closing the gap between theory and practice in this important area of OM research. Finally, to the best of our knowledge, our paper provides the first empirical comparison between performance contracts and non-performance contracts that are used for aftermarket customer support.

3. Industry Background

In this paper we focus on the maintenance, repair and overhaul (MRO) industry for commercial aircraft. According to the previously noted Standard and Poor's industry report (2009), this sector generated revenues of $117 billion in 2007, of which $60 billion was related to military MRO, $45 billion to air transport (commercial aircraft) MRO, and $12 billion to business and general aviation MRO. Aircraft owners (e.g., airlines) face the problem of properly managing maintenance and repair of aircraft equipment, including managing the risk of infrequent equipment failures and the disruption of scheduled maintenance checks, so as to preserve aircraft availability, thus avoiding the high opportunity cost of having an aircraft on-the-ground. Customers typically purchase after-sales service support from the OEM and/or other service suppliers on either a transaction basis (i.e., through T&M contracts), or negotiate a contract for support where payment is based on the number of flying hours (i.e., PBC).

Aircraft fleet availability therefore is the most important performance metric for airlines and other customers. We note that availability is influenced by several factors that include subsystem and part reliability, spare parts inventory, repair capacity and repair lead time. For example, if a critical part of an aircraft subsystem stops functioning, it must be replaced by a working unit drawn from the spares inventory (if it is available) as soon as possible to minimize disruption to flight schedules. All broken units are ultimately returned to a support depot where they are either repaired or scrapped. A key challenge for both customers and suppliers in this industry is to reduce the cost of failures while maintaining an acceptable level of fleet uptime (availability) through the management of resources (i.e., spare parts inventory and repair capacity) and through interventions that could affect the reliability of the product and/or the performance of the support processes.

Recently, the airlines have adopted outsourcing strategies for MRO services in order to focus on their core activities and 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, engine, landing gear, avionic system). Typically, it is the OEMs themselves that offer support services for their subsystem products because the highly customized and complex nature of their products makes it difficult for a third party to provide similar product care. The provision of such services is also very profitable with margins that often exceed those associated with the sale of the product. Examples of such major (OEM) suppliers in the MRO industry include Pratt & Whitney, General Electric Co., Rolls Royce, Honeywell Aerospace, Lockheed Martin and Boeing. There are also many smaller MRO companies that provide a wide range of services that include scheduled maintenance checks and parts repair. Such providers include the Triumph Group Inc., AAR Corp and Heico Corp. MRO service providers usually offer different types of contracts under which their customers can receive support services, including time and material (T&M), fixed-price (FP) and performance-based contracts (PBC). Different versions of PBC have become widely adopted, evolving and transforming the relationship between manufacturers and customers in the industry. For example, recently Boeing aggressively started pursuing contracts for its 787 GoldCare program. As noted in 2006 by a company vice-president, “before we were just selling parts, now we are selling airlines a power-by-the-hour service and we are guaranteeing availability".[4] Although PBC is an important factor in the industry, there are no published estimates of its actual effect on performance metrics or on supplier profit.

In this paper, we study the performance implications of the contractual relationship between Rolls-Royce, a major supplier of aircraft engines and services to support them, and its customers. The dataset we analyze was provided by Rolls-Royce. Rolls-Royce delivers after-sales repair and maintenance services for its customers under two different types of contracts: T&M and PBC. The main difference between the two types of contracts is the mechanism under which the customer pays for the support services. Under a T&M contract the customer pays for the materials and resources that are consumed each time a maintenance event occurs. Under PBC, the customer agrees to pay a fee that is proportional to the actual flying hours the customer generates from their fleet of aircraft. (For example, a customer pays $*x* per flying hour to the supplier with a guaranteed minimum number of hours flown per quarter;[5] note that flying hours can be converted into aircraft fleet availability.) In other words, compensation to Rolls-Royce under PBC is directly tied to the performance outcome that the customer values.

4. Data

The dataset consists of five years of data (July 2002 - July 2007, hereafter the observation period) of maintenance events (product 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 or for maintenance purposes, resulting in a *shop visit* to the service supplier. Removals are undesirable events for aircraft owners since an aircraft on-the-ground generally results in high opportunity costs, with estimates being as high as hundreds of thousands of dollars per day for a non-scheduled failure for a fully loaded wide body commercial aircraft.

For each product unit, the available variables in our data include:

- exact time of each product removal in the observation period,
- cumulative aircraft flying hours at the time of each shop visit,
- product model,
- the contract type (T&M or PBC) under which the product receives service,

[4] Lou Mancini, in *Airline Business*, "Maintenance special report: service culture" (09/26/2006).
[5] Variations of this payment scheme are observed in other settings, especially in the defense industry. There, many PBC contracts are based on a weighted average of different performance metrics (including "soft" metrics like customer satisfaction). However, flying hours is the only metric used in Rolls-Royce's PBC ("Power by the Hour®").

- aircraft tail number associated with the product,
- ID of the customer that owns the product, and
- the list of all products belonging to each customer at the end of the observation period.

The original data include 9 different product models; among them, there are 4 models where the installed base is covered exclusively by PBC contracts. We therefore limited our sample to the remaining 5 product models for which we have data on the removals that occurred under both PBC and T&M contracts. We also discarded some product units that have inconsistent data (e.g., reported flying hours at a shop visit in 2006 are less than the flying hours reported in a shop visit in 2005). Our final sample consists of 1,076 removals associated with 763 different product units where 78.6% of the product units are covered by PBC. This sample includes 5 different product models and 62 different customers. The unit of analysis in our main analysis is a product unit but we will discuss alternative approaches in Section 6.

While there are a number of performance metrics that would be of interest in our study (e.g. availability, reliability, and cost of ownership), our analysis focuses on product reliability since data on other metrics were unavailable in our dataset. As mentioned in the Introduction, theory guides us to believe that different incentive structures under either T&M or PBC contracts should be reflected in variations of product reliability. Among several possible measures of product reliability, mean time between removals (MTBR) is the measure that we employ in our analysis since it is a key reliability metric that practitioners in the aerospace and defense industry constantly monitor. It represents the average time (flying hours) that a product is used without the need for a removal for repair and maintenance purposes. MTBR is also a key input for the estimation of the demand rate for parts consumption and the arrival rate of repair events in stochastic models of repairable parts supply chains (see Sherbrooke 1992) which are used to compute fleet availability.

It is important to note that, by definition, MTBR is only a partial representation of the physical reliability *inherent* in the product. To be precise, MTBR is a function of both physical reliability and managerial efforts to avoid future failures (e.g., via more frequent scheduled maintenance checks). The natural question is then: is it obvious that higher product reliability corresponds to larger MTBR? Indeed, it seems possible that the supplier who wishes to have a more reliable product may actually lower MTBR since he may perform more frequent inspections. Although the answer to this question should ultimately be obtained empirically, it is true that most practitioners view MTBR as a leading indicator of product reliability. This is so because a removal is typically a very costly and time-consuming process which occurs only when an engine actually fails randomly in the field or when it is necessary to perform a removal to execute a scheduled maintenance check which is mandated at specified time periods. Removals, for whatever cause, lead to lack of product availability which leads to reduced value generation, and, thus, they are associated with a reduction in reliability. Consistent with this widespread convention in practice, we use the terms *product reliability* and *MTBR* interchangeably.

Although use of MTBR as a metric of product reliability is well justified, it still poses nontrivial issues for our analysis because removals are quite rare. In our observation period of 5 years, the majority of products in the dataset (67%) are removed only one time and another 27% are removed only two times[6] which somewhat limits our ability to compute the true MTBR of a product. Additionally, the data suffers from censoring as information on any removals that occurred before July 2002 or after July 2007 are excluded. Therefore, we build a proxy for MTBR that *adjusts* for the unobserved data and use it as a dependent variable. There are several candidates for such a proxy, but as we show later on, the conclusion of our analysis is quite insensitive to proxy choices.

We illustrate the procedure used to calculate the proxy with an example (see Figure 1). 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 removal occurred at time $T_1 < T_B$, i.e., this removal was unobservable to us. We observe the 2 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 (flying hours) of a product at time T.[7] In the example, the (true) MTBR is given by TSN(T_3)/3, but we do not observe the first removal and we do not even know if this removal took place. 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 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).

[6] It should be noted that the products that were never removed during the observation period are not included in the sample, since no removal record exists for them. Therefore, we are in fact considering a conditional MTBR in our analysis: i.e. MTBR of a product given that a removal has occurred at least once in the sample period.

[7] 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 in the y-axis and the latter is in the x-axis.

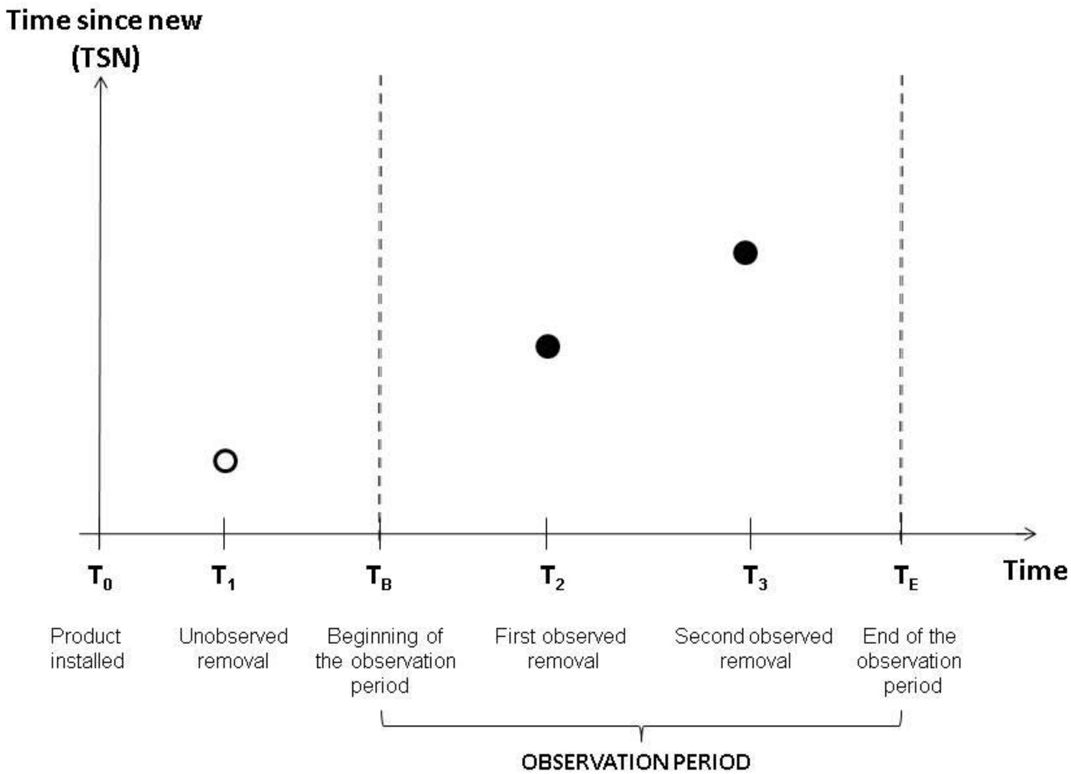

Figure 1: Example of a removal sequence for a product.

We built a proxy for MTBR to account for the effect of this unobserved data. The proxy is defined in the following equation[8]:

$$\text{MTBR} = \frac{\text{TSN(lastet observed removal)} - \text{TSN}(T_B)}{\text{\# observed removals}}$$

In the example, our proxy for the MTBR is given by [TSN(T_3)- 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

[8] In the robustness section we discuss alternative proxies that treat the data censoring issue differently, and show that the results of the paper are not affected by alternative definitions of the MTBR variable.

period. In our example, we estimate the slope of the constant usage line using T_2, T_E, TSN(T_2), and TSN(T_E). 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)}. We believe that this approximation provides a reasonable estimate for MTBR. Note that our adjusted measure gives the correct value for MTBR if the product was not installed before July 2002. The bias introduced by this metric will vary since it depends on the number of unobserved removals. We attempt to reduce the potential bias by subtracting the initial product age. We have confirmed with managers of the company that the magnitude of the estimates based on this approach is in line with actual measures that the company tracks.

Table 1 displays summary statistics for the variables MTBR, the initial age of the product (ini_age) and the number of removals (nremovals) in our final sample. The MTBR and the initial age of the product are measured in flying hours. Based on sample averages, products covered by T&M contracts have a slightly greater MTBR than PBC products do. Also note that there is more variability in MTBR for products covered by T&M in comparison to PBC products.

	Overall sample		**T&M only**		**PBC only**	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
MTBR	3441.9	1768.8	3711.7	2336.8	3368.6	1574.2
ini_age	3381.8	2788.8	4324.9	3314.2	3125.6	2572.1
nremovals	1.41	0.65	1.25	0.50	1.45	0.68

Table 1: Summary statistics

An alternative approach would be to infer MTBR (as an output of the analysis) by estimating the underlying distribution of the time between removals using techniques drawn from duration modeling (see, for example, Cameron and Trivedi 2005), instead of computing the MTBR for each individual product and using it as a dependent variable in the linear regression analysis. Duration modeling is a nonlinear regression method that requires estimating the distribution of the time between removals. In fact, the sample in our dataset represents *multiple-spell durations*, since many products have more than one removal. The main advantage of this estimation technique is that it has a built-in mechanism to deal with some of the censoring issues, which are common in duration data such as ours. Unfortunately, there is currently no widely accepted econometric procedure to account for endogeneity within the framework of duration modeling.

As we noted previously, uncovering endogenous contract choices by customers is a central feature of our problem, and thus our purpose is better served by using an alternative two stage estimation approach using MTBR as a dependent variable. Unlike duration modeling, this approach allows us to explicitly and consistently deal with the endogeneity that is inherent in the data and has important implications for testing our key hypothesis concerning the impact of contract choice on reliability. For this reason, the main part of analysis in the subsequent sections builds on this proxy-based approach. For completeness, we present in Section 6.4 the results of an analysis based on one method that combines duration modeling with instrumental variables, which has been used by some researchers to account for endogeneity.

5. Econometric Model

We first introduce the econometric framework we use in the analysis, and then elaborate on the model specification.

5.1 Econometric Framework

Our goal is to build a model that properly captures the effect of contract type on product reliability, i.e., MTBR. A major challenge associated with isolating the true marginal effect of different contract types on product reliability is the inherent endogeneity associated with contract type choice by customers. Indeed, customers do not sign on for either a T&M or PBC contract randomly, but rather respond to several factors that influence this decision, i.e., self-selection is expected in our setting. A number of empirical studies have considered the issue of contract choice decisions by firms in different contexts (see, for example, Slade (1996) for an application in the oil industry and Crocker and Reynolds (1993) for an application to the Air Force engine procurement process). In these and other studies, endogeneity of contract choice has been regarded as a key econometric issue in testing contract design hypotheses (Masten and Saussier 2002). As our results will indicate, there is evidence that endogeneity of contract choice by a customer is present in our data as well. As a consequence, usual ordinary least squares (OLS) estimation would lead to biased estimates of the marginal effect of contract choice on product reliability. General econometric discussions on the importance of accounting for self-selection and related methods can be found elsewhere (e.g., Heckman 1979, Maddala 1983, Greene 2008).

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

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

$$\boldsymbol{z_i} = \mathbb{1}[\boldsymbol{w_i\gamma + \upsilon_i} > 0] \qquad \text{(Eq. 2)}$$

The observed MTBR 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 in standard 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 $(\varepsilon_i, \upsilon_i)$ 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 υ_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 (Eq. 1 could be estimated by OLS), i.e., endogeneity is not relevant for the problem. For additional information on this model, including a discussion of identification, the reader is referred to Maddala (1983, p.117-125). 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).

5.2 Model Specification

We postulate that contract type can affect a product's MTBR. As we have discussed, under PBC the supplier of the service gets paid in proportion to the flying hours generated by the customer, creating an incentive to maintain or improve product reliability. Theoretical support for this hypothesis has been discussed extensively in a previous analytical research (see Kim et al. 2008 for details). What then are some of the actions that the supplier and customer can take to improve product reliability under PBC? First, on the supplier side, given that the supplier does not receive any benefit for having product on-the-ground under PBC, he can give priority to products covered by PBC through more rapid support service, which in turn increases product availability for the PBC customer. The supplier may also provide a higher quality of service (fewer errors, more careful testing, etc.) to a PBC customer. While this behavior primarily impacts service time, it also may indirectly affect product reliability as measured by MTBR which captures the frequency of support events. More directly, the supplier can invest in pre-emptive maintenance in order to avoid future product failures. This is particularly important if we take into account the fact that there is an important difference between minor product revisions / scheduled maintenance checks, for which repair service can take several days, and major product failures and maintenance checks, for which the average shop visit takes about 5 weeks. Pre-emptive maintenance actions can reduce the frequency and duration of future checks as well as lead to fewer failures. Perhaps the most important impact that the supplier can have on reliability, however, is through product re-design and engineering change. Investment in such activity results in a more reliable product and improved versions of parts used for product support which supersede existing, less reliable versions.

On the customer side, the intensity of product utilization and the procedures customers use to maintain the product could also impact product reliability. For example, under a PBC arrangement the customer may request removals more frequently and engage in more pre-emptive maintenance. They are encouraged to do so under PBC since, beyond opportunity costs, there are no direct expenses to the customer associated with a maintenance shop visit. In contrast, T&M customers need to pay for all support services consumed regardless of the impact of such services on performance improvement. This conjecture is in fact in line with the statistics displayed in Table 1, that show that on average, the number of removals during the observation period is greater for PBC than for T&M products. The arguments above motivate the main hypothesis we test in this research, i.e. *product reliability increases under PBC*. Although we have provided arguments that support this hypothesis, note that while an increased number of removals will reduce the observed values of MTBR, pre-emptive maintenance may reduce the occurrence of random failures and thus could actually increase the time between removals which would increase observed values for MTBR. The net effect of customer response to a PBC is thus, a priori, not obvious and is ultimately an empirical question.

Another factor that can influence the observed MTBR is the initial condition of the product. In fact, it is expected that older products are more likely to fail than new ones, particularly in the case of very old products. However, it can also be argued that at the very beginning of a product's life cycle, the product is likely to need more adjustments than mature products do, resulting in more frequent shop visits, which can reduce the value of the observed MTBR. To check these conjectures, we plot the distribution of MTBR for different ranges of initial product age in Figure 2. The graph suggests that MTBR 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 fact, there appears to be a concave relationship between MTBR and initial age. In order to account for such effects, we include both linear and quadratic terms for the initial age of the product in our model specification. The linear term should take care of the initial increasing trend in product reliability, while the quadratic term should reflect the decreased MTBR for *old* products. Since our observation period is five years, it is important to note that we do not necessarily observe five years of behavior for all products, as some of them (15%) were new at the beginning of the observation period. This measurement issue also makes it important to control for initial age of the product in order to isolate the effect of PBC on MTBR.

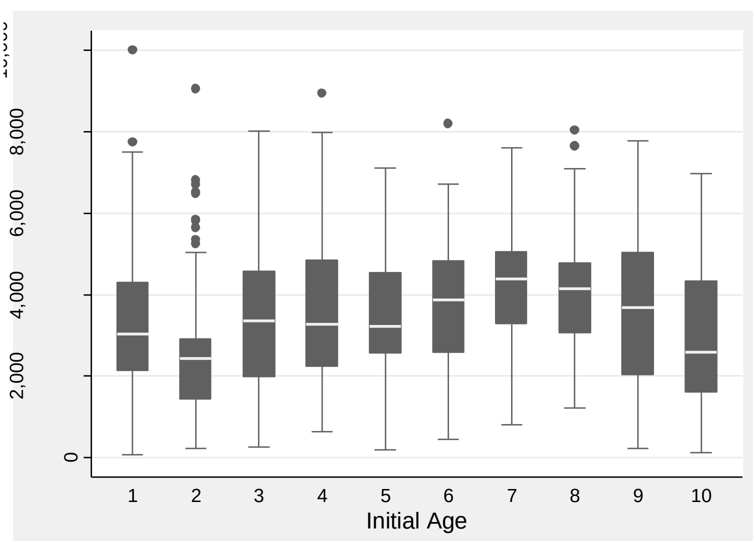

Figure 2: MTBR vs. Initial age (flying hours). Variable initial age is categorized as follows: '1'→ 0<initial age<=1000, '2' → 1000<initial age<=2000, ..., '9'→ 8000<initial age<=9000, '10'→ 9000<initial age

Finally, product model identity can also be correlated with the MTBR, especially since different products in our sample were not launched simultaneously and have different designs. We include product model dummies as control variables in the outcome equation of our model specification, to account for these factors. This completes our specification of the outcome equation.

In order to specify the choice equation, we need to include covariates that influence the type of contract selected by a customer. According to supplier managers who we interviewed, a factor that greatly influences contract selection is customer size, i.e., customers with a larger fleet are more likely to choose a PBC contract for after-sales support. This conjecture is in line with the data: the median fleet size of T&M and PBC customers are 2 and 9, respectively. We measure fleet size as the number of products that a customer has registered with the service provider. A larger fleet size is expected to be associated with greater total fleet-flying-hours (at the customer level), e.g., a customer with a fleet of 50 products is likely to have, as a customer, more flying hours per year than a customer with 10 products. This may cause larger firms to expect to use the MRO service more frequently, which may influence the likelihood to sign on for PBC.[9] Alternatively, a smaller fleet may lead to more intense use per aircraft, which would also lead to more frequent maintenance events, providing an incentive for PBC. Finally, larger customers may have internal capabilities to deal with maintenance and therefore may not desire comprehensive coverage through PBC. Thus, although we do not have a clear theoretical prediction with respect to the impact of customer size on the contract choice, we include fleet size as a control variable. Our main model uses the logarithm of fleet size of the customer in the choice equation in order to smooth out the distribution of this variable. Measures of size have been widely used to explain contract choice in different contexts; see Lafontaine and Slade (2001), and the references therein, for examples.

Other factors that can influence contract choice for a given product are the value of the aircraft equipment, the customer's risk profile, and contract terms. These variables, however, were not available in our sample. We note that the product models owned by a customer also may be an indicator for some of these factors. For example, product models are expected to partially reflect customer fleet valuation. It can also be the case that customers' propensity to

[9] Of course, another major variable that is likely to affect contract choice is pricing. Unfortunately, we do not possess pricing data.

repair certain models of products under a particular contract type with the supplier varies depending on the product model, due to, e.g., the customer's ability to perform certain maintenance or repair tasks in-house for particular product models. In fact, our data show that some product models are more likely to be covered by PBC contracts than others: the proportion of units covered by PBC contracts is equal to 59.6%, 81.2%, 50.9%, 81.1% and 95.7%, for product models 1, 2, 3, 4 and 5, respectively. This suggests existence of correlation between product model and contract type so we include the product model as a control variable in the choice equation.

Naturally, other variables could have explanatory power for contract choice. For example, a factor that has been recognized to play a role in theoretical models is the level of risk aversion of the customer. Unfortunately, and as has been recognized in previous research, from an empirical point of view, it is virtually impossible to measure this factor (Lafontaine and Slade 2001). Thus, our specification of the choice equation is somewhat constrained by data availability.[10] It is, however, plausible, that customer fleet size variable is a partial proxy for risk aversion.

Summarizing, our model specification postulates that the MTBR of a product can be explained by the contract type under which the product is serviced (PBC vs. T&M), the product age at the beginning of the observation period (linear and quadratic terms), and the product model. Note that these variables are representative of some of the key factors that could drive the decision to remove a product for maintenance, i.e., contract terms, nature of the product, and the usage environment. The contract type under which a product is serviced is explained by the customer fleet size and the product model.

Finally, note that our modeling approach allows for correlation between the unobservable effects of the outcome and choice equations. This feature is especially useful since we do not observe variables related to the risk profile of the customer, which can influence both product reliability and contract choice. For example, the choice of PBC may be more likely for more risky customers, if they expect to use their fleet and MRO services more frequently. In addition, products owned by more risky customers are expected to show a lower MTBR in comparison to those owned by low-risk customers, due to the higher risk of failures. Similar arguments have been discussed in the case of extended warranties for new car buyers (Padmanabhan 1995), where it is hypothesized that heavy users have stronger incentives than light users to sign on for

[10] We discuss alternative specifications in the robustness section.

extended warranties, since their products are more likely to experience failures. In our case, we expect that the correlation between both error terms will be different from zero due to self-selection of the customers, and in particular, for the reasons outlined above, we expect the correlation between both the unobserved effects to be negative.

6. Analysis and Results

6.1 OLS Analysis

To illustrate the relevance of the endogeneity problem in the isolation of the effect of contract choice on MTBR, we begin by estimating the model defined by Eq. 1 using a regular OLS regression. Of course, this does not take into account the self-selection problem by customers over contract types, as defined by Eq. 2. Table 2 presents the results obtained with OLS with regular standard errors (column 1) and clustered standard errors at the customer level (column 2) to account for possible correlations across engines owned by the same customer.

VARIABLE	SE	Cluster SE (customer)
prodmodel2	-435.4*	-435.4
	(248.1)	(376.7)
prodmodel3	-660.6*	-660.6*
	(310.1)	(345.7)
prodmodel4	-2620***	-2620***
	(275.1)	(440.1)
prodmodel5	-1866***	-1866***
	(345.7)	(503.9)
ini_age	0.122**	0.122*
	(0.047)	(0.072)
ini_age_sq	-0.00002***	-0.00002**
	(0.000004)	(0.000006)
PBC	-255.3*	-255.3
	(138.6)	(338.5)
Constant	4533***	4533***
	-287.7	(391.7)
Observations	763	763
R-squared	0.277	0.277
Adj. R-squared	0.27	0.27

Table 2: OLS estimation. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

OLS results suggest that there is a negative effect of PBC on product reliability, counter to our hypothesis. This effect is partially significant when regular standard errors are used, and largely insignificant under clustered standard errors. We have argued that contract type is endogenous, as it is expected to be correlated with the error term of the regression in Eq. 1 because of unobserved customer characteristics that are embedded in the error term (e.g. customer risk profile) that are also correlated with the contract choice. If that is the case, OLS regression generates biased estimates. In particular, we have argued that the unobserved term in the choice equation is expected to be negatively correlated with the unobserved term in the outcome equation, which implies that the OLS coefficient for PBC would be biased downwards.

6.2 Two-stage Approach: Estimation and Results

We now turn to the estimation and results obtained for the two-stage model defined by Eqs.1 and 2. The likelihood for product *i* is given by:

$$L^i(\beta,\delta,\gamma,\sigma,\rho) = [f(y_i|z_i = 1)P(z_i = 1)]^{z_i}[f(y_i|z_i = 0)P(z_i = 0)]^{1-z_i}$$

Given the bivariate normality assumption, $\boldsymbol{v_i}$ can be expressed as $\boldsymbol{v_i} = \frac{\boldsymbol{\rho}}{\boldsymbol{\sigma}}\boldsymbol{\varepsilon_i} + \boldsymbol{\eta_i}$, where $\boldsymbol{\eta_i} \sim \mathbf{N}(\mathbf{0}, \mathbf{1} - \boldsymbol{\rho}^{\mathbf{2}})$, independent of $\boldsymbol{\varepsilon_i}$. The log-likelihood of the model can then be written as:

$$\begin{aligned} &LL(\beta,\delta,\gamma,\sigma,\rho) \\ &= \sum_{i=1}^{N}\left\{z_i\left[-\frac{1}{2}\frac{(y_i - x_i\beta - \delta)^2}{\sigma^2} - \ln\left(\sqrt{2\pi}\sigma\right) + \ln\Phi\left(\frac{w_i\gamma + \frac{\rho}{\sigma}(y_i - x_i\beta - \delta)}{\sqrt{\mathbf{1} - \boldsymbol{\rho}^{\mathbf{2}}}}\right)\right]\right. \\ &\left. + (1 - z_i)\left[-\frac{1}{2}\frac{(y_i - x_i\beta)^2}{\sigma^2} - \ln\left(\sqrt{2\pi}\sigma\right) + \ln\Phi\left(-\frac{w_i\gamma + \frac{\rho}{\sigma}(y_i - x_i\beta)}{\sqrt{\mathbf{1} - \boldsymbol{\rho}^{\mathbf{2}}}}\right)\right]\right\}, \end{aligned}$$

where Φ is the cdf of the standard normal distribution.

We estimate the model in STATA by fully maximizing likelihood. Table 3 displays the results obtained from the two-stage model, including estimates for both the outcome and choice equations, considering two types of standard errors. Columns (1) and (2) display results with regular standard errors for the maximum likelihood estimation, while columns (3) and (4) display results with clustered standard errors at the customer level. Clustered standard errors at the customer level allow for correlation between products of the same customer, while maintaining the independence assumption for products of different customers. This is particularly important in our case, as we have argued that the error terms of the outcome and choice equations involve customer unobservables. In the case of cluster standard errors at the customer level, the variance-covariance matrix of the estimates involves the computation of the sum of the interactions between the residuals and the covariates for each of the products of a given customer, which is repeated for each customer (see e.g. Greene 2008 pp.188-190, Wooldridge 2002 pp.328-331, for further details).

	SE		Cluster SE (customer)	
VARIABLE	Out. Eq	Ch. Eq.	Out. Eq	Ch. Eq.
prodmodel2	-667.4***	0.305	-667.4	0.305
	(258.5)	(0.212)	(597.3)	(0.659)
prodmodel3	-568.7*	-0.464*	-568.7	-0.464
	(319.9)	(0.264)	(492.9)	(0.564)
prodmodel4	-2853***	0.735***	-2853***	0.735
	(285.2)	(0.239)	(566.3)	(0.647)
prodmodel5	-2266***	1.159***	-2266***	1.159
	(362.7)	(0.391)	(678.6)	(0.805)
ini_age	0.112**		0.112	
	(0.046)		(0.071)	
ini_age_sq	-0.00002***		-0.00002***	
	(0.000004)		(0.000007)	
PBC	**790.3*** **		**790.3***	
	(248.4)		(460.7)	
lfleetsize		0.468***		0.468***
		(0.0436)		(0.146)
Constant	3936***	-1.403***	3936***	-1.403*
	(314.8)	(0.248)	(615.8)	(0.762)
# Obs.	763	763	763	763

Table 3: The two-stage model. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Our estimate for the effect of PBC on MTBR is positive and indicates that, on average and all else being equal, PBC increases the MTBR of a product by 790.3 flying hours in our five year observation period. This effect is significant at all relevant significance levels when regular standard errors are used, and remains significant when cluster standard errors at the customer level are employed, though only at the 90% confidence level in this case (p-value=0.086)[11]. For the two different standard error measures outlined above, a 90% confidence interval for the effect of PBC on MTBR is given by (382, 1199), (32, 1548), respectively. Results of a likelihood ratio test (see e.g. Wooldridge 2002 p. 397) for the null hypothesis of no correlation between the error terms of Equations 1 and 2 provide strong statistical evidence to reject the null hypothesis of independent equations: p-value<0.001 for SE and p-value=0.0026 for cluster SE (customer level). This is in line with our hypothesis of the endogeneity for the variable PBC, confirming that the model cannot be estimated by OLS due to the endogeneity in the contract selection decision. In fact, the estimate of the correlation between both error terms is -0.455, also in line with our expectation that they are negatively correlated. Overall, these results provide important support for our main hypothesis that employing PBC increases product reliability. It is important to note that this result is consistent with the results of previous economic modeling of performance-based contracts that predict that suppliers will invest more in product reliability improvement when they are under a performance contract as compared to their behavior under a T&M type of contract (Kim et al. 2008).

Results in Table 3 also provide support for the hypothesis of a non-linear relationship between MTBR and the initial age of the product. The linear term is positive while the quadratic term is negative, which suggest a concave functional form, also in line with our data and hypothesis, indicating that medium-age products have greater MTBR than very new and very old products. Both coefficients remain significant under the two standard error measures, with the exception of the linear term for cluster standard errors at the customer level. The mean of the residuals for the predicted MTBRs is -0.0000545, which suggests no general bias in the predictions.

With respect to the choice equation, the results indicate that both fleet size and product model dummies have explanatory power, although the latter lose significance under clustered standard errors. In particular, our results show that customers with greater fleet sizes are more likely to choose PBC than T&M contracts, in line with our hypothesis, data, and managerial expectations.

[11] We have alternatively considered clustered standard errors at the aircraft level, as it is possible to argue that products of the same aircraft type may exhibit some correlation, e.g., given that an aircraft is on-the-ground due to a product failure, a customer may be more likely to include other products of the same aircraft type in a shop visit. Results under clustered standard errors at the aircraft level are more significant than clustered standard errors at the customer level, i.e. we are reporting the most conservative case. In particular, under clustered standard errors at the aircraft level, the p-value for the PBC coefficient is 0.003.

This result holds no matter what standard errors are used. The choice model correctly predicts 81.1% of the observations in the sample, in line with accuracy levels reported in related applications in the literature (e.g. Terwiesch et al. 2005).

6.3 Robustness Checks

We have shown that our main result – the positive effect of PBC on MTBR – is robust to different standard error measures, given our specification. In this section we perform further robustness checks of our results to alternative model specifications. For simplicity, we focus the analysis on our main hypothesis that PBC increases product MTBR.

First, we consider several specifications that vary between original scale and logarithmic forms for our numeric variables. Our main model uses the logarithm of fleet size of the customer in the choice equation in order to smooth out the distribution of this variable. If we use the original scale variable (fleet size) instead of the logarithm form – so that all the numeric explanatory variables are in their original form – our estimate for PBC is 699 (p-value=0.012; 0.362, for the two standard error measures, respectively). Similarly, if we use logarithms instead of the original scale values for the linear and quadratic terms of initial age in the outcome equation – so that all the numeric explanatory variables are in logarithm form – our estimate for PBC is 887 (p-value=0.000; 0.049).

Second, we consider the inclusion of other variables in the choice equation. For example, if we view the initial age of the product as a proxy for the risk of specific customer equipment, it can be argued that the initial age of the product can influence the decision of a customer to choose PBC instead of T&M contracts. When a linear term for the initial age of the product is included in the choice equation, the estimate for PBC is 790 (p-value=0.002; 0.092); when both a linear and a quadratic term are included, the estimate for PBC is 787 (p-value=0.002; 0.088).

Similarly, a factor that can influence the decision of the type of contract is the expected usage of the service. It can be hypothesized that customers that expect to use their fleet more frequently are more likely to choose PBC contracts. Such customers may give more weight to having access to future maintenance services without having to make a payment for the resources consumed each time the service is needed. We consider different measures that attempt to capture aggregate customer usage: average usage per year and several ranking indicators across customers based on average usage per year. Our estimates for PBC are in the range 391 to 842, with p-values ranging from 0.001 to 0.1, depending on the proxy used to capture this effect. Ideally, we would calculate a proxy for the expected frequency of usage based on, for example, pre-sample data, to derive an exogenous measure not directly related with our measure of MTBR thus avoiding circularity problems. In the absence of this exogenous source though, we built the proxies for this variable (as indicated above) based on our in-sample data.

Our estimates for the effect of PBC on MTBR of a product are remarkably robust to many model variations. The main finding that PBC has a positive and significant marginal effect on product MTBR remains robust in virtually all cases, with coefficients ranging from 391 to 887 flying hours. Equally importantly, our tests reject the null hypothesis of uncorrelated error terms (outcome and choice equations) in all cases, providing strong evidence to support the self-selection hypothesis, which justifies our econometric modeling approach.

We also test the robustness of the results with respect to the definition of the variable MTBR. We have discussed the assumptions and approximations used to calculate the MTBR variable in section 3, in particular, we noted the censoring issue in our data. We explore three alternative proxies that involve a different treatment of the censored data, which we define using the example in Figure 1 (we use different names for clarity):

- MTBE = average{TSN(T_2) – TSN(T_B), TSN(T_3) – TSN(T_2), TSN(T_E) – TSN(T_3)}.
- MAXMTBRMTBE = max{MTBR, MTBE}
- InvFRate = [TSN(T_E) – TSN(T_B)]/[number of observed removals]

Thus, the MTBE (mean time between events) includes the most recent portion of the time in which an engine was not removed, i.e., it considers that as an 'event'. The MAXMTBRMTBE also includes that information by contrasting the MTBE with the MTBR. Finally, the InvFRate represents the inverse of the failure rate, the latter being calculated as the number of observed removals over the flying hours of an engine during the observation period. Our main findings remain robust to these three variations. When OLS is used for estimation, PBC has a negative effect on reliability. When our two-stage approach is used to estimate the model, we find a positive and significant effect of PBC on product reliability.[12] In all cases, likelihood ratio tests confirm the endogeneity of the contract choice variable. The coefficient of PBC remains essentially in the same range: 470.9 (MTBE), 744.9 (MAXMTBRMTBE), 751.2 (InvFRate).

[12] Significance is lost when clustered standard errors at the customer level are used though.

These experiments confirm the main findings obtained with our base model, i.e. when the endogeneity inherent to contract choice is accounted for, PBC have a positive effect on product reliability.

6.4 Duration Model Analysis

Thus far, our analysis has relied on constructing proxies for the MTBR of each product and performing a two-stage estimation. As mentioned earlier, the duration model offers an alternative way to analyze our data, despite its aforementioned deficiency related to incorporating endogeneity into the model. As research on how to deal with this issue is ongoing, currently there is no agreed upon method or statistical package that we can adopt for our purposes (see Bijwaard (2007) for a recent contribution to this research stream). However, the following informal approach has been used by some researchers: (1) run the probit model (Eq. 2) to predict contract choices, (2) calculate the selectivity term from this analysis, and (3) perform duration analysis using the computed residuals 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, yet fully established, we use this procedure in this subsection as a final robustness check.

In order to proceed with this approach, we must analyze the data at the removal level (instead of at the product level), and we must examine the influence of contract type on the respective 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. Beyond the change in the dependent variable, the explanatory variables remain the same; the only modification in explanatory variables we incorporated is to replace the initial age of the product with the age of the product at the time of the previous removal (if any) in Eq. 1.[13] We estimate the models in STATA, using clustered standard errors at the customer level. When the models are estimated disregarding the endogeneity of PBC, we find that PBC has a positive effect on the removal rate, i.e., analogously to the OLS analysis of the MTBR, we find that PBC decreases reliability. This effect is non-significant no matter which hazard rate model is used.

[13] For example, for the 3rd 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).

To account for endogeneity in duration models, we employ the two-stage procedure described above: first, we run a probit regression to estimate the contract choice equation (Eq. 2). From the results of this first stage, a selectivity term can be derived as follows (see e.g. Maddala 1983 p.121):

$$selectivity\, term\_i = \begin{cases} \frac{\phi(w_i\gamma)}{\Phi(w_i\gamma)}\ ,\ if\ z_i = 1 \\ \\ \frac{-\phi(w_i\gamma)}{1-\Phi(w_i\gamma)}\ ,\ if\ z_i = 0 \end{cases}$$

Here, ϕ and Φ are the density and the cumulative distribution function of the standard normal variable. This selectivity term is then used as an explanatory variable in the hazard rate model, as a way to capture the true effect of PBC on the removal rate taking into account the self-selection involved in the PBC variable. Using this procedure, we find a negative and significant effect of PBC on the removal rate, in all of the four different semi-parametric and parametric transition rate models under consideration. Thus, the conclusion that PBC increases reliability remains the same under this alternative modeling approach, which also reaffirms the relevance of accounting for endogeneity to estimate the effect of PBC. We suppress detailed output from the duration model analysis for the sake of brevity but we conclude that this analysis is consistent with the two-stage approach that we utilize in the rest of the paper.

6.5 Limitations

Although we have shown that there is strong evidence to support our conclusions, our analysis is not free of limitations. There are several issues regarding the nature of the data. First, we analyze product removals from an aircraft, and calculate a proxy for product reliability for the products we observed, which are those products that were removed from an aircraft at least once during the observation period. Thus, our analysis is driven by the nature of our dependent variable, the mean time between removals. Our analysis is silent, however, with respect to those products that were never removed. Second, in our sample we observe a small proportion of products covered by T&M contracts (21.4%), which may suggest the risk of sample selection bias. Recall that we use data from only one supplier in the market, so it is possible that customers that prefer a T&M contract scheme choose a different supplier for their repair and maintenance service. Naturally, we do not observe such customers, and we do not have any basis to construct a model to explain a potential sample selection of suppliers by customers.[14] Although our study is limited in this regard, we nonetheless believe that the insights from our analysis are relevant to wide range of firms, especially those in the aerospace and defense industry which are undergoing a major shift toward the PBC approach. Moreover, to make sure that the uneven proportion of PBC and T&M products in our sample is not playing a role in our results, we conduct further experiments estimating our model with balanced samples (50% PBC, 50% T&M) by randomly selecting units from the pool of PBC observations; our results and main findings remain qualitatively the same under this variation. Third, while our dataset is rich in terms for characterizing the removal incidents for a given product, we have only limited data to characterize a customer. In particular, unobservables related to customer risk profile and behavior might have an impact on our results. It is also true that our dataset is not rich in describing the specific terms of the contracts in each case; we only distinguish between T&M contracts and PBC. This does not allow us to explore the influence of price and other contract conditions on the customer's contract choice although the supplier suggested to us that there are no major differences in contract parameters: most of them are signed at common list prices. Finally, we do not observe data before/after adoption of PBC, which would have made possible to study the dynamics, e.g., by using difference-in-differences estimators.

With respect to our modeling approach, the main assumption we impose is the exogeneity of the independent variables (other than PBC) in our model specification. In essence, we assume that the unobservables related to customer risk profile and behavior are uncorrelated with the initial age of the product, the product model and the fleet size. This assumption rules out the possibility that customer behavior might vary depending on the initial age of the product, or that customer risk exposure will change for different product models. Thus, while our approach explicitly deals with the endogeneity of contract choice, our results need to be understood in the context of the exogeneity assumption for the rest of the covariates.

7. Conclusion

[14] The supplier has informed us that the majority of their products are covered by PBC contracts while maintenance for the majority of T&M removals is conducted by independent contractors. Such contractors generally have a good reputation for both cost and turnaround time, but their impact on reliability is unknown. These observations are consistent with our sample.

We have examined the impact of performance-based contracts on product reliability (as measured by the time between removals), in an application to the aerospace and defense maintenance and repair services industry. Using a proprietary dataset from Rolls-Royce, a major supplier of engines, we propose a two-stage approach that allows us to explicitly account for the endogeneity inherent in contract choice by a customer. The first stage of the econometric model describes the customer decision with respect to selecting a contract and the second stage analyzes the impact of contract type on product reliability. Our analysis shows that there is a positive and statistically significant effect of PBC on the MTBR of a product, i.e., performance-based contracts induce improvements in product reliability in our sample. Our estimates indicate reliability improvements under PBC in the 10-25% range, in comparison to traditional T&M contracts. We have also shown that there is strong statistical evidence to support the hypothesis of correlation between the error terms of the outcome and choice equations, confirming that the coefficient of PBC cannot be identified without taking into account the inherent endogeneity of contract type selection by customers. These findings are supported by several robustness checks under a number of alternative model specifications, which also allowed us to measure the impact of other covariates on product reliability and contract choice.

Our analysis focuses on the marginal effect of performance-based contracts on product reliability. Our results provide a first step towards understanding the overall impact of performance-based contracts, and our approach was largely driven by data availability. The availability of richer data about customers, financial and managerial information, and the specific contract terms between them and the supplier, would enable a more complete analysis to cover a number of open questions. Such analysis, which is underway by the authors, could lead to a deeper understanding of the benefits of PBC contracts, e.g., what drives reliability improvement? Is this reliability improvement profitable to the supplier? Does the cost of pre-emptive maintenance exceed the benefits due to reliability improvement? Is the price charged to the customer appropriate? How do specific contract terms moderate the impact on reliability? These questions remain open for future research.

This paper is one of the few studies that empirically estimate the impact of a performance vs. non-performance contract type and other causal factors on supply chain outcomes. Our findings are relevant not only to the aircraft repair and maintenance service industry but also to all industries that provide after-sales support for mission critical products. The results are especially relevant for practitioners since this is the first attempt to test the reliability improvement hypothesis for performance contracting based on transactional data. While there are numerous papers that model various supply chain contracts, there is little empirical evidence of the impact of such contracts on supply chain outcomes. Our paper thus makes a step in closing the gap between theoretic modeling and empirical evidence.

Acknowledgement

This research was partially funded by the Fishman Davidson Center for Service and Operations Management, The Wharton School, University of Pennsylvania. The authors would also like to acknowledge the financial support and active participation in the project by Rolls-Royce, who provided the data used in this study. We would, in particular, like to acknowledge the assistance of Miles Cowdry and W. Bunker Snyder who supported the project as well as James J. Olsen, Lindsey E. Shackelford and Jarod Downing who assisted us in collecting the data. Finally, we acknowledge the financial support and input from the US Air Force, (in collaboration with the University of Tennessee), who is funding our ongoing research in this area. We are also thankful to the following individuals for their extensive comments on the earlier version of this paper: Donald Lee, Richard Lai, Marcelo Olivares, Justin Ren, Enno Siemsen, Rachna Shah, Manuel Sosa and Owen Wu.

References

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

Banker, R., S.Y. Lee, G. Potter, D. Srinivasan. 2001. An empirical analysis of continuing improvements following the implementation of a performance-based compensation plan. *Journal of Accounting and Economics*, 30(3), 315-350.

Brijwaard, G. E. 2007. Instrumental variable estimation of treatment effects for duration outcomes. Working paper.

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.

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

Cohen, M. A., V. Agrawal, Y-S Zheng. 1997. Service parts logistics: a benchmark analysis. *IIE Transactions*, 29(8), 627-639.

Crocker, K., K. Reynolds. 1993. The efficiency of incomplete contracts: an empirical analysis of Air Force engine procurement. *RAND Journal of Economics*, 24(1), 126-146.

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.

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.

Greene, W. H. 2008. *Econometric Analysis*. Sixth edition, Prentice-Hall, Upper Saddle River, NJ.

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

Ho, J., K. Ho, J. Mortimer. 2008. Welfare effects of full-line forcing contracts in the video rental industry. Working paper.

Ioannou, I., J. Mortimer, R. Mortimer. 2009. The effects of capacity on sales under alternative vertical contracts. Working paper.

Kesavan, S., V. Gaur, A. Raman. 2008. Incorporating price and inventory endogeneity in firm-level sales forecasting. Working paper, University of North Carolina.

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

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. 2008. Reliability or inventory? analysis of product support contracts in the defense industry. Working paper.

Kim, S.-H.., M. A. Cohen, S. Netessine, S. Veeraraghavan. 2009. Contracting for infrequent restoration and recovery of mission-critical systems. Working paper.

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

Lafontaine, F., M. Slade. 2001. Incentive contracting and the franchise decision. In *Advances in Business Applications of Game Theory*, K. Chatterjee and W. Samuelson (eds.), Kluwer Academic Press.

Lafontaine, F., K. L. Shaw. 1999. The dynamics of franchise contracting: evidence from panel data. *The Journal of Political Economy*, 107(5), 1041-1080.

Lee, D. K. K. and S. A. Zenios. 2007. Evidence-based incentive systems with an application in health care delivery. Working paper.

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, UK, Cambridge University Press.

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.

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 effect 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.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature*, 37(1), 7-63.

Prendergast, C. 2002a. The tenuous tradeoff between risk and incentives. *Journal of Political Economy*, 110(5), 1071-1102.

Prendergast, C. 2002b. 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, T., S. Netessine, N. Rudi. 2006. An empirical examination of the decision to invest in fulfillment capabilities: A study of Internet retailers. *Management Science*, 52(4), 567–580.

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

Slade, M. E. 1996. Multitask agency and contract choice: an empirical exploration. *International Economic Review*, 37(2), 465-486.

Shen, Y. 2003. Selection Incentives in a Performance-Based Contracting System. *Health Services Research*, 38(2), 535-552.

Sherbrooke, C. C. 1992. *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques*. Wiley.

Standard & Poor’s. 2009. Industry Surveys: Aerospace and Defense (February 12th).

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.

Toffel, M. W. 2002. Contracting for servicizing. Working paper, Harvard University.

Ton, Z., A. Raman. 2009. The effect of product variety and inventory levels on sales: a longitudinal study. *Production and Operations Management Journal*, forthcoming.

U.S. Government Accountability Office. 2008. Defense logistics: improved analysis and cost data needed to evaluate the cost-effectiveness of Performance-Based Logistics, GAO-09-41.

Wooldrigde, J. M. 2002. *Econometric Analysis of Cross-Section and Panel Data*. MIT Press, Cambridge, MA.