

2013/40



Pooling in manufacturing:
do opposites attract?

Tanja B. Mlinar and Philippe Chevalier



CORE

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium

<http://www.uclouvain.be/core>

CORE DISCUSSION PAPER
2013/40

**Pooling in manufacturing:
do opposites attract?**

Tanja B. MLINAR¹ and Philippe CHEVALIER²

July 2013

Abstract

In a stochastic environment pooling naturally leads to economies of scale, but heterogeneity can also create variability. In this article, we investigate this trade-off in the case of a manufacturing environment. Pooling for queueing systems has been widely investigated while much less attention has been given to manufacturing systems where jobs are given a due date upon arrival. In such system it is not the elapsed time until the actual completion of the job that counts, but rather the lead time that can be promised to the customer in order to guarantee a high service level. In this paper, we study the benefits of pooling stochastic systems in such manufacturing setting with multiple customer types. Our results demonstrate that, in stark contrast with what was previously observed in service environments, heterogeneity is generally not deteriorating performance. Furthermore, our analytical and simulation studies reveal that the benefits of pooling in terms of the expected sojourn time can serve as a good prediction for the benefits of pooling on the average due-date lead time in a wide range of situations.

Keywords: manufacturing, resource pooling, queueing, due dates

¹ Université catholique de Louvain, Louvain School of Management and CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: tanja.mlinar@uclouvain.be

² Université catholique de Louvain, Louvain School of Management and CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: philippe.chevalier@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.

The authors are thankful to Jean-Sébastien Tancrez for useful comments and suggestions that helped to improve this paper.

This work was supported by the ARC (Action de Recherche Concertée – Concerted Research Action) project on Managing Shared Resources in Supply Chains of the Communauté Française de Belgique. The scientific responsibility is assumed by the authors.

1 Introduction

There is a clear trend in manufacturing towards product variety and heterogeneity. For instance, a recent study about the food industry in several European countries showed that the number of different products increased by 53 percent from 2000 to 2010 (Swinnen and Herck, 2011). Moreover, the number of companies declined while the size of the survivors increased. As a consequence, manufacturers must find ways to increase the diversity of their production without losing competitiveness. One way to deal with the expansion of product variety is to leverage economies of scale across different products. The advantages of economies of scale can be viewed in terms of cost savings or better operational performance such as a short delivery time and a faster service. Pooling operations with uncertain demand can be an important source of economies of scale. However, a company must be aware of the variability that may arise when pooling product lines with heterogeneous products. This variability can potentially diminish benefits achieved by pooling. Thus, it is necessary to get deeper understanding about the interplay between economies of scale and variability in manufacturing.

In this article we shed a new light on this dilemma by taking into account the fact that in practice most firms operate with due dates. Short due-date lead times (DDLTs) and on-time delivery are important measures of performance of manufacturing companies. They enable the manufacturer to obtain a price premium (So and Song, 1998), attract customers and ensure long term profitability (Keskinocak and Tayur, 2004). Therefore, we study a manufacturer whose objective is to be able to quote due dates to orders of different demand streams which minimize the average due-date lead time (DDLT) while being able to respect those due dates with a high probability. The purpose of this paper is to analyze whether pooling two or more such type of systems allows manufacturers to obtain economies of scale.

The benefits of pooling are often evaluated by comparing the performance of the pooled system with the traditional "stand-alone" situation in which each order type is assigned to its own facility. We follow the same idea by considering that a manufacturer caters orders from multiple demand streams which differ in arrival rates and processing requirements. We start with a system where each facility is dedicated to the production of one order type. We model this dedicated system as a collection of single-server queues with a single order type where due dates are assigned to jobs upon their arrival. In the case of pooling, all order types are merged into a single priority queue and jobs are treated by a joint flexible facility with a pooled capacity. We consider that the pooled production system gives a non-preemptive priority to orders waiting to be processed based on their due dates and/or processing times in order to reach the desired objective.

A production facility in any of the two models could involve a series of processes but we model it as a single-server queueing system. This is done in order to isolate the effect of heterogeneity in demand patterns and processing times from other effects that can occur in such complex environments. Similar modeling assumptions about these two systems are made in Buzacott (1996), Mandelbaum and Reiman (1998) and Iyer and Jain (2004). In the case of the pooled system, we also study an alternative model where there are parallel facilities (one for each dedicated facility that is pooled). We show that the choice made regarding to the pooled system does not have a significant impact on our findings.

Our main evaluation criteria are the pooling gains for the DDLT (defined as the relative reduction in the average DDLT if we pool dedicated capacities). There are two common practices to assign due dates in production systems. Given that the choice of using either practice is often imposed to the manufacturer by the context in which he/she is operating, we analyze the pooling gains in both cases. The first practice assigns a fixed DDLT to all orders from the same demand stream. Because of its simplicity for both the manufacturer and the client, this practice is com-

monly applied. The second one is more sophisticated because due dates are assigned dynamically based on the state of the system. This leads to higher efficiency but the variability of the resulting lead times tends to complicate the planning process for the clients.

As an example of this type of situation motivating our work, we can cite the joint venture Behr–Hella Service which is formed in order to supply the aftermarket for vehicle air conditioning and engine cooling. One of the company’s major objectives is pooling resources and expertise devoted to the after-market sales and service of their products (Hella Australia Pty Ltd, 2007). Another way of pooling is through a merger or an acquisition which aims to produce a larger value than the sum of the values of each individual facility (see e.g. Gupta and Gerchak, 2002). Thus, ”it is important to be able to quantify the value of pooling capacities” (Jain, 2007).

There exists a large literature on economies of scale associated with pooling queueing systems in service environments. For such operations it is often desirable to maintain a First-Come-First-Served (FCFS) discipline e.g., banks, shops, restaurants etc. The most important measure of performance of such systems is the time a customer spends waiting to be served. The conclusion of these studies is that when customers are heterogeneous the economies of scale tend to disappear because of the additional variability brought by the heterogeneity (in terms of different service times and arrival rates). Here, the aim is to analyze how the due-date setting and scheduling policies can be used to deal with the variability resulting from pooling production capacities with multiple order types.

We start our analysis by performing a simulation study. Our results show that in manufacturing it is possible to create economies of scale even in circumstances where such economies would not arise if restricted to the FCFS rule. In particular, the pooling gains for the DDLT are much less sensitive to heterogeneous order types compared to what can be observed in service environments with FCFS policy. Moreover, the pooling gains for the DDLT increase with the utilization rate. These interesting insights can be explained by the flexibility provided by the due-date mechanism combined with appropriate scheduling rules. We also find that asymmetry in traffic loads for the different product types has a positive impact on the overall performance of the pooled system (when compared to the dedicated system).

In order to provide deeper insights into pooling for manufacturing environments with heterogeneous order types, we develop analytical models to calculate the pooling gains for the simplified setting when the expected sojourn time is considered instead of the due-date performance measure. Interestingly, it turns out to be a very good indicator of the pooling gains for the DDLT in a wide range of situations. We then show that increased product variety improves the pooling gains. This can be explained by the decrease in the variance of the processing times of the pooled facility when the number of pooled operations increases.

In a broader sense, our findings bring both theoretical insights and practical applications. From a theoretical point of view, our contribution lies in providing the first approach to bridge the gap between the following literature streams: the operation management literature based on pooling queueing systems, the literature on flexibility in manufacturing and the literature based on due-date management problems. From a practical point of view, we show for a wide range of instances that manufacturers should not be afraid of diversity when pooling production resources.

The paper is organized as follows. In Section 2 we review the related literature. In Section 3, we introduce our model and provide a detailed description of the production configurations. In Section 4, we describe the due-date setting and scheduling policies we consider in this paper (Section 4.1), perform a simulation study (Section 4.2), define performance measures of interest (Section 4.3) and discuss the results obtained for a pooled single-facility configuration (Section 4.4). In Section 5 we develop different analytical models (Section 5.1) and analyze the impact of the main assumptions we make in order to provide a better understanding of our results: we discuss how

increasing the product variety impacts the pooling decision (Section 5.2), and investigate the due-date performance in a pooled multi-facility configuration (Section 5.3). Finally, Section 6 concludes the study and presents suggestions for future research.

2 Literature review

In this section, we give an outline of the existing works on related topics and our contribution to each of them.

2.1 Pooling queueing systems

The benefits of pooling for queueing systems have been studied in the operation management literature (see Stidham, 1970; Rothkopf and Rech, 1987; Whitt, 1992). These benefits are assessed mainly by comparing the expected waiting time (Smith and Whitt, 1981; Benjaafar, 1995; van Dijk and van der Sluis, 2008) or the mean sojourn time (Mandelbaum and Reiman, 1998; Enns and Grewal, 2010) for pooled systems and dedicated ones. These studies mainly concern service environments restricted to the FCFS priority. Some applications are in call centers (van Dijk and van der Sluis, 2008), communication networks (Smith and Whitt, 1981) and computer systems (Kelly et al., 2009). However, research on these issues for manufacturing environments is still scarce. In particular, Iyer and Jain (2004) and Benjaafar et al. (2005) consider pooling in production and inventory systems, and Enns and Grewal (2010) investigate pooling in batch manufacturing systems.

There are two papers closely related with our work. van Dijk and van der Sluis (2008) investigate the benefits of pooling on the expected waiting time for a multi-server call center which serves two job types according to the FCFS policy. The authors point out that the resource pooling can be counter-productive under high utilization, customer heterogeneity and a small number of servers. In order to reduce the impact of uncertainty, the authors deal with an alternative multi-server flexible system with an overflow and threshold policy. Their results show that such configuration can outperform two extreme cases: a strictly dedicated and pooled configuration. In Tekin et al. (2009) the authors further investigate the potential benefits of partial pooling on the expected waiting time in call centers under either the FCFS or a non-preemptive priority. While their analysis is limited to equal utilization rates across servers which can be a common practice in call centers, in manufacturing environments this is often not the case. Given the complexity of the considered problem, the authors use queueing approximations for the expected sojourn time in a multi-server system under the non-preemptive priority, which may limit their model to be accurate only for low levels of heterogeneity.

In contrast to the two previous works, we investigate the benefits of pooling in a manufacturing environment with due dates and heterogeneous order types. In such environment the objective of the company differs and one of the most important measures of performance is due-date related rather than the time that customers spend waiting before being served. In order to minimize the average DDLT a policy that prioritizes orders according to their due dates and not on their arrival times is often employed. Moreover, we investigate how heterogeneity in both service times and traffic loads influences the pooling gains for the DDLT.

The benefits of a pooled single-server system compared to dedicated systems have been investigated in the literature. For example, Iyer and Jain (2004) investigate these benefits in terms of inventory costs when a pooled single capacity is owned by two companies serving demand from the two warehouses with separate ownerships. In a more general setting, the benefits of pooling multiple capacities into a single flexible capacity in queueing networks is investigated in terms of

the mean sojourn time in Mandelbaum and Reiman (1998). We build upon these works to investigate how the due-date performance is affected in a pooled single-facility system in the presence of different due-date setting and scheduling policies.

2.2 Flexibility in manufacturing

In the literature on flexible manufacturing, various flexible configurations have been investigated. Most of the research on this topic develops strategic models to determine the amount of flexibility needed while minimizing shortages, capital investments or maximizing sales revenues. A seminal article by Jordan and Graves (1995) studies partial manufacturing flexibility for a multi-plant and multi-product configuration. The authors show that for homogeneous processing times a chain configuration with each production resource capable of processing two product types achieves most of the benefits of total flexibility. The work was extended by Graves and Tomlin (2003) for multi-stage supply chains. Sheikhzadeh et al. (1998) show that for the homogeneous case and setup times up to 50% of the processing time, fully flexible configurations can outperform the chained ones in terms of the average throughput and mean flowtime if an appropriate scheduling rule is implemented. Bassamboo et al. (2010b) show that it is optimal to invest in at most two adjacent levels of flexibility in systems with homogeneous demand patterns (so-called pairing) under the general framework of a multidimensional newsvendor model. Furthermore, Bassamboo et al. (2010a) show that for parallel queueing systems in a heavy-traffic and homogeneous processing times it is optimal to invest in dedicated resources in order to process the base demand and flexible resources which can process only two product types in order to handle variability. Our paper also contributes to understanding the flexibility and design of production systems but from the operational point of view. Our aim is to examine and understand the effects of flexibility on the operational performance of a firm for which the DDLTs are one of the fundamental factors in the production planning.

Gurumurthi and Benjaafar (2004) consider more general queueing systems with multiple customer classes that vary in demand rates and routing flexibility, and multiple servers that vary in service rates and service flexibility. The authors consider a set of control policies. They show that higher flexibility does not necessarily improve throughput. In practice heterogeneity in processing times is often imposed by specific requirements of different customer classes. In this paper, we investigate how this heterogeneity can impact the decision to use flexible resources by developing rather simpler models while focusing on their due-date performance measure.

The research on flexibility provides valuable insights on how much flexibility is needed. Unfortunately, very little is known in the literature about how to use this flexibility to improve a due-date performance when facing customers with heterogeneous demand. In this paper, we aim to provide a framework to guide a manager to exploit economies of scale in manufacturing by leveraging the additional flexibility brought by the due-date setting and scheduling policies.

2.3 Due-date management

The due-date management literature distinguishes two types of due-date setting practices, namely: static and dynamic (see e.g. Cheng and Gupta, 1989; Keskinocak and Tayur, 2004). The first one assigns a constant DDLT to each incoming customer (Cheng and Gupta, 1989) while the second one assigns a DDLT dynamically based on the state of the system (as in e.g. Wein, 1991; Sridharan and Li, 2008; Koulamas and Kyparisis, 2008). The quotation of due dates is mostly combined with different scheduling policies for a single-machine or job-shop production systems (see Keskinocak and Tayur, 2004, for a survey of due-date setting and scheduling policies).

While in practice it is common to quote fixed due dates, a vast body of research finds that dynamic due-date setting methods provide better performance for the measures based on the flow time and tardiness (see e.g. Wein, 1991; Wein and Chevalier, 1992; Vinod and Sridharan, 2011). Wein (1991); Wein and Chevalier (1992); Hopp and Sturgis (2000) develop models which quote due dates dynamically in order to satisfy a global service level such as the percent of orders satisfied on time, the fraction of tardy jobs or the average job tardiness. The interaction between due-date setting methods and scheduling rules in a dynamic job shop production system is investigated in Vinod and Sridharan (2011). This paper represents a direct extension of these works in due-date management in which we investigate the potential value of having a more flexible manufacturing environment. Such environment makes it possible to pool the demand streams for different job classes. Despite the large body of research on due-date management problems, economies of scale for such systems have, to our knowledge, not been explored.

3 A model for manufacturing operations with due dates

In this section, we introduce our model for manufacturing operations with due dates and present our two benchmark configurations, namely the dedicated and pooled configurations. We kept our model quite simple in order to better emphasize the effects of the pooling decision.

In the dedicated configuration (configuration d), each demand stream is handled by an independent facility M_i^d (we use superscript d to indicate the dedicated configuration). We denote q_i^d the queue for orders of type i ($i = 1, 2, \dots, n$) that wait for service at M_i^d . In the pooled configuration (configuration p) all demand is treated together in a joint facility M^p (we use superscript p to indicate the pooled configuration). This configuration results from the aggregation of the different production facilities in order to handle all demand types. We suppose that the production and demand characteristics are not affected by pooling. The operations of the pooled configuration can potentially be even more complex than in the dedicated case, here again we model the resulting facility as a single server queueing system with N order types in order to isolate the effect of the pooling decision. We denote q^p the queue for the pooled demand streams that wait for service at M^p . In Section 5.3 we examine the potential impact of this modeling choice.

An order type is characterized by its arrival process and processing time distribution. We assume that order type i has its own exponential processing time distribution with mean τ_i and its Poisson arrival process with rate λ_i . The traffic load of order type i is then $\rho_i = \lambda_i \cdot \tau_i$, where $\rho_i < 1$ in order to ensure the stability of the system. Therefore, configuration d is modeled as n independent M/M/1 queueing systems. Similarly, configuration p is modeled as a multi-class M/M/1 queueing system. In order to make a comparison between the two configurations we consider that the global capacity of the pooled configuration is equivalent to the sum of the production capacities of the n dedicated production systems. Consequently the processing times are n times faster in p . The assumption about the equal capacities is made in a large number of previous articles about pooling (Smith and Whitt, 1981; Mandelbaum and Reiman, 1998; van Dijk and van der Sluis, 2008; Benjaafar, 1995; Benjaafar et al., 2005).

As soon as an order arrives, a due date must be given according to a due-date setting policy with the objective to quote due dates with the shortest possible average DDLT while being able to provide a high service level. In order to assign a due date with a short lead time, the cycle time should be as short as possible. The question thus becomes to determine the scheduling rule such as to create economies of scale for the cycle times when pooling different demand streams. We assume that preemption (i.e. interruption of the production of an order) is not allowed, and setup times are negligible. We explain in detail in Section 4.1 the due-date setting and scheduling policies used

in this paper.

We define $D_i^s(t)$ as the due date assigned to the order of type i arriving at time t to configuration s ($s \in \{d, p\}$). In configuration d , due date $D_i^d(t)$ is assigned to the incoming order so as to minimize the average $DDLT$ of type i , denoted by \overline{DDLT}_i^d , with a service level SL_i (the percentage of orders of type i finished on time). In configuration p , due dates are assigned in order to minimize the average $DDLT$ of all order types while respecting the service level requirements of each order type. The due-date management problem for the dedicated and pooled configuration with two order types is illustrated in Figures 1a and 1b, respectively.

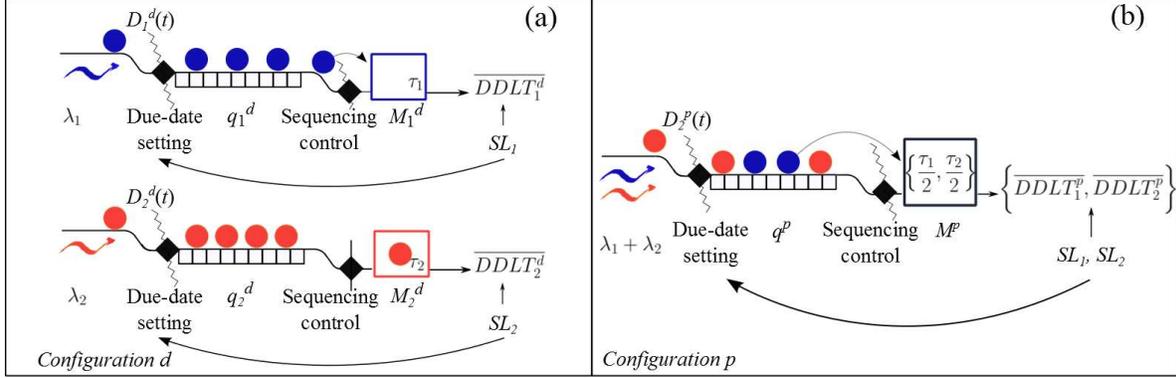


Figure 1: Due-date management problem for the dedicated (a) and pooled (b) production configurations

4 Simulation study

In order to compare the average DDLTs of the two configurations we perform a simulation study. Simulation has been used extensively for due-date setting and scheduling problems in previous studies (e.g. Vinod and Sridharan, 2011; Veral, 2001; Wein and Chevalier, 1992; Bookbinder and Noor, 1985; Baker and Bertrand, 1982). One of the reasons is the analytical complexity of such problems which obviate the possibility of a theoretical treatment as explained in Barman (1998).

Our simulation study has two purposes: the first purpose of the simulation model is to explore and evaluate how heterogeneity in processing times, arrival rates, and capacity utilization affects the due-date performance of the pooled configuration compared to the dedicated configuration; the second purpose is to demonstrate how the due-date setting practices and scheduling policies influence the decision to pool production resources.

4.1 Operating policies

In this section we describe in more detail the two due-date setting policies (DDSP) and the scheduling policies (SP) considered in this paper.

4.1.1 Due-date setting policies

The *static policy* (CON) assigns a constant DDLT to orders irrespective of the state of the facilities. $D_i^s(t)$ can be expressed as: $D_i^s(t) = t + f_i^s$, where f_i^s represents the lead time of an order of type i in configuration s . This parameter is obtained via simulation so as to satisfy the service level of order type i .

The *parametric dynamic due-date policy* (DYN) proposed in Wein (1991) assigns due dates based on the conditional expected sojourn time i.e. the total time the order spends in the queueing system assuming that the shortest processing time policy is applied. We choose this dynamic policy because it represents one of the most efficient existing dynamic policies (Wein, 1991).

Based on Wein (1991) the due date $D_i^s(t)$ is expressed as: $D_i^s(t) = t + r_i \cdot E[S_i^s(t)]$, where r_i represents a parameter obtained via simulation in order to satisfy the service level of order type i , and $E[S_i^s(t)]$ represents the conditional expected sojourn time of the order of type i that arrived at time t conditional to the state of the system in configuration s . Let $E[S_i^p(t)]$ and $E[S_i^d(t)]$ be the conditional expected sojourn time of type i in configuration p and d , respectively. Then, $E[S_i^p(t)]$ is expressed as (Wein, 1991):

$$E[S_i(t)^p] = \frac{\tau_i}{n} + \frac{\sum_{j=1}^n m_j^p(t) \cdot \frac{\tau_j}{n} + E[R^p(t)]}{1 - \sigma_{i-1}}, \quad (1)$$

where $m_j^p(t)$ represents the number of orders of type j in queue q^p at time t , $E[R^p(t)]$ is the expected residual processing time of the order being processed at time t by M^p , and parameter $\sigma_i = \frac{1}{n} \sum_{j=1}^i \rho_j$. For the exponentially distributed processing times, the value of $E[R^p(t)]$ is defined by expression (2).

$$E[R^p(t)] = \begin{cases} \frac{\tau_i}{n}, & \text{if at time } t \text{ an order of type } i \ (i = 1, 2, \dots, n) \text{ is being processed,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

In configuration d , the dedicated facility M_i^d can process only orders of type i and $E[S_i^d(t)]$ is given by:

$$E[S_i^d(t)] = \tau_i + m_i^d(t) \cdot \tau_i + E[R_i^d(t)], \quad (3)$$

where $m_i^d(t)$ represents the number of orders of type i in queue q_i^d at time t and $E[R_i^d(t)]$ is the expected residual processing time of the order of type i being processed by M_i^d at time t . For the exponential processing times, the value of $E[R_i^d(t)]$ is defined as follows:

$$E[R_i^d(t)] = \begin{cases} \tau_i, & \text{if } M_i^d \text{ is busy,} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

4.1.2 Scheduling policies

In the dedicated configuration the orders are processed according to the FCFS policy. In the pooled configuration there is more freedom in scheduling given that the facility can choose a job of a specific type to process from the queue in order to improve the due-date performance. We tested three scheduling policies that facility M^p can follow: Shortest Expected Processing Time policy (SEPT) which assigns the priority to the order type with the shortest expected processing time (type 1), Earliest Due Date policy (EDD) which assigns the priority to the job with the earliest due date, and the least remaining SLACK policy which assigns the priority to the job with the smallest slack (where the slack is the difference between the due date of the order, its mean processing time and time at which the scheduling decision is being made). We choose these policies because they are the most studied policies in the literature and generally preferred in practice (see e.g. Keskinocak and Tayur, 2004).

We denote DDSP-SP a combination of a due-date setting policy, denoted by DDSP, and a scheduling policy, denoted by SP. Therefore, there are six combinations of due-date setting and

scheduling policies in configuration p : CON-SEPT, CON-EDD, CON-SLACK, DYN-SEPT, DYN-EDD and DYN-SLACK; and two combinations in configuration d : CON-FCFS and DYN-FCFS. Note that, we compare the due-date performance of the two configurations under the same DDSP policy.

4.2 Experimental design

In our numerical experiment, we consider that the manufacturer processes orders of two types. We discuss this choice in Section 5.2. The performance of the system depends on the characteristics of order types. Therefore, an instance is determined by traffic loads of each type (ρ_1 and ρ_2) and the coefficient of heterogeneity c , defined as a ratio between the two order types ($c = \tau_2/\tau_1$). We fix the service level to 95% and τ_1 at unity. Note that for a fixed traffic load there exists a direct link between the processing times and the arrival rates i.e. $\lambda_i = \rho_i/\tau_i$.

Utilization. To isolate the effect of heterogeneity in processing times or arrival rates from the effect of asymmetry in workloads, we initially consider the case where the workloads for the two order types are equal (balanced case). In the balanced case we try four different values for the utilization rate: $\rho = \{0.6, 0.7, 0.8, 0.9\}$. In the unbalanced case the asymmetry in workloads can be represented by the coefficient δ , which we define as $\delta = \rho_1/\rho_2$. We consider two values for the utilization rate: $\rho = \{0.6, 0.75\}$ and two cases of asymmetry in traffic loads: the traffic load of one order type is 50% higher/lower than the traffic load of another type i.e. $\delta = \{2/3, 3/2\}$. This gives us four different combinations of ρ_1 and ρ_2 i.e. $(\rho_1, \rho_2) = \{(0.48, 0.72), (0.72, 0.48), (0.6, 0.9), (0.9, 0.6)\}$.

Coefficient of heterogeneity. In order to investigate the impact of heterogeneity in processing times on the decision to pool production systems, we set the coefficient of heterogeneity to take the following values $c = \{4, 8, 12\}$. Therefore, in our experiment, for a given traffic load, the order type will either have a fast service and frequent arrivals or a long service with infrequent arrivals. Such situation often occur in practice when pooling the operations of two firms where one would be targeting clients with small orders while the other is targeting clients with larger orders. Examples are in a custom plastic injection molding shop or a printing shop for printing brochures, books, catalogues, calendars etc.

In total, there are 144 and 48 different instances for the pooled and dedicated configurations, respectively. We construct two simulation models each for one production configuration in order to study the pooling gains for these instances. Each experiment was simulated during a period $T = 200,000$ time units and was replicated 40 times using unique random numbers. The system was started empty and without warm-up period. The simulation parameters were chosen such as to ensure that for the most variable settings (long service times with infrequent arrivals) the 95% confidence interval does not exceed 5% of the average DDLT.

4.3 Performance measures

As our main performance measures we consider the average DDLT of each order type and the global average DDLT (the average $DDLT$ over all order types). We will denote the average DDLT of type i orders by \overline{DDLT}_i and the global average DDLT by \overline{DDLT}_g . Furthermore, in order to gain additional intuition into the pooling gains we will consider the expected sojourn time of order type i and the global expected sojourn time, denoted by $E[S_i]$ and $E[S_g]$, respectively.

To evaluate the gains that arise from pooling n dedicated production systems for a specific performance measure for order type i we define $\varepsilon(PM_i) = \frac{PM_i^d - PM_i^p}{PM_i^d}$, where PM_i^s is the value of the performance measure for order type i in configuration s . The pooling gains for a specific

global performance measure is evaluated in a similar way, i.e. as $\varepsilon(PM_g) = \frac{PM_g^d - PM_g^p}{PM_g^d}$, where PM_g^s is the global performance measure for configuration s .

4.4 Pooling gains for the due-date performance measures

The complete results are presented in Appendix A in Table 1 and Table 2 for the balanced and unbalanced case, respectively. Each line presents the results for a combination of due-date setting policy and scheduling policy. For example, the DYN-EDD combination means that the pooled and dedicated configurations operate under the DYN due-date setting policy and the scheduling policy is the EDD policy and the FCFS in the pooled and dedicated configuration respectively.

4.4.1 Pooling gains for the global due-date performance measure

Here, we contrast our results concerning pooled manufacturing operations with due dates with results concerning pooled service operations restricted to the FCFS policy studied in the literature so far. In Figure 2 we show how the global pooling gain for the DDLT under the DYN-SEPT and CON-SEPT policy combination evolve compared to the global pooling gain for the expected sojourn time under the FCFS policy¹. Based on our results we make several observations related to the characteristics of order types and the operating policies considered.

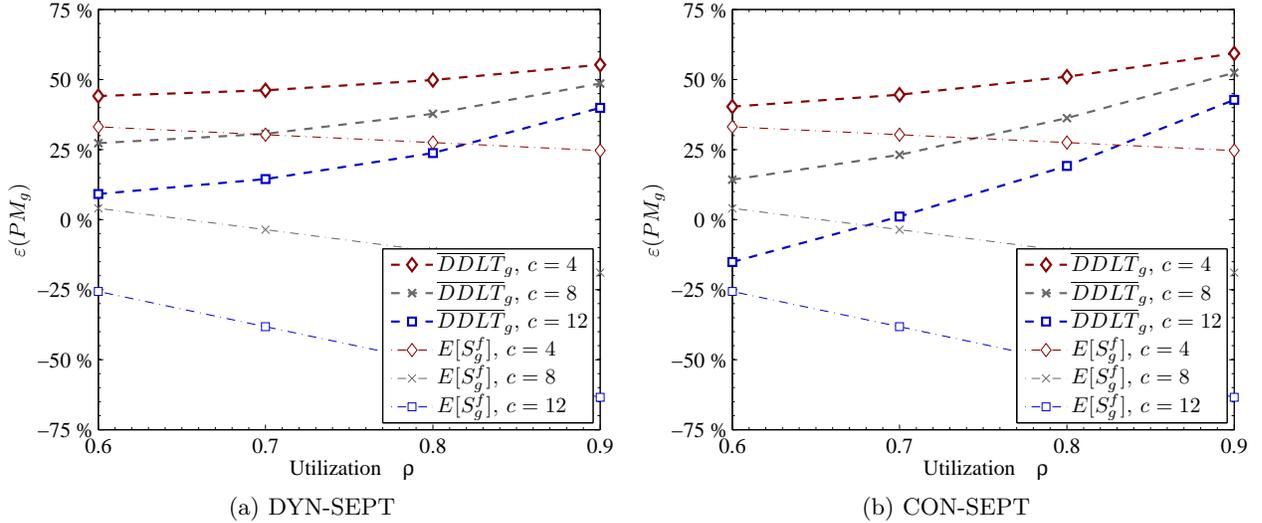


Figure 2: The pooling gains (in %) on $\overline{DDL T}_g$ and $E[S_g^f]$ for the balanced case and different values of ρ and c

¹The pooling gain for the global expected sojourn time under the balanced case and the FCFS (we use superscript f to denote the FCFS policy) are computed by the following expression: $\varepsilon(E[S_g^f]) = \frac{1}{2} - \frac{\rho(c-1)^2}{8c}$. The expression is obtained from $E[S_g^{d,f}]$ and $E[S_g^{p,f}]$, where $E[S_g^{d,f}]$ is calculated based on the expressions (5) and (8). $E[S_g^{p,f}]$ is calculated as follows: $E[S_g^{p,f}] = E[W_g^{p,f}] + \bar{\tau}^p$ where $E[W_g^{p,f}] = \frac{1}{2}(1+v^2)\frac{\rho \cdot \bar{\tau}^p}{1-\rho}$, $\bar{\tau}^p = \frac{\sum_i^n \lambda_i \tau_i^p}{\sum_i^n \lambda_i}$ and $v^2 = \frac{(\sum_i^n \lambda_i) \sum_i^n \lambda_i (\tau_i^p)^2 (v_i^2 + 1)}{(\sum_i^n \lambda_i \tau_i^p)^2} - 1$ (for the exponential processing times $v_i^2 = 1 \forall i = 1, \dots, n$). Refer to van Dijk and van der Sluis (2008); Tekin et al. (2009) for more details about the pooling gains under the FCFS policy.

Impact of Heterogeneity. Figure 2 shows that the global gain achieved by pooling manufacturing operations with due dates is less sensitive to heterogeneity in order types compared to the service operations restricted to the FCFS policy. In particular, when $c = 12$ and $\rho = 0.6$, $\varepsilon(\overline{DDLT}_g)$ is equal to 9.15% under the DYN DDSP while $\varepsilon(E[S_g^f])$ is equal to -25.63%. When pooling is beneficial in the context of services it is also beneficial in the context of manufacturing. However, the opposite does not necessarily hold. In manufacturing, it is indeed different: in many cases pooling is advantageous when with identical parameters in service operations restricted to the FCFS it is not the case.

Impact of the Global Utilization Rate. Pooling brings drastically different gains in the two environments studied. While the pooling gains increase with utilization in the context of manufacturing these gains decrease in the context of services. It is important to note that as the utilization rate decreases the pooled configuration faces fewer arrivals and there are less opportunities to prioritize between orders i.e., jobs are more often processed in the order in which they arrive to the system which reduces the impact of the analyzed scheduling policies.

Impact of the Due-date Setting Practice. From our results we observe that pooling is advantageous under the DYN due-date setting policy for all instances. Moreover, we can note that both due-date setting practices have a similar tendency under any scheduling policy considered (see Table 1 and Table 2). With the DYN due-date setting policy the pooling gains are more stable with respect to changes in the scheduling policy or heterogeneity factor c . Therefore, companies which operate under the constant due-date setting practice should pay an additional attention to the choice of scheduling policy applied while pooling their operations.

Impact of the Scheduling Policy. We observe that as the coefficient of heterogeneity increases, the performance improvement brought by the SEPT policy increases compared to the two due-date scheduling policies. This can be attributed to the higher priority given to orders with shorter expected processing times that are the orders most easily hurt by pooling. Similarly, as heterogeneity decreases the two order types become more similar, and the scheduling policies which discriminate between orders based on their due dates will lead to a higher improvement of the overall performance of the pooled configuration compared to the SEPT policy.

4.4.2 Pooling gain for the due-date performance measure of each order type

We now analyze the pooling gain for the DDLT of each order type. Figure 3a and Figure 3b illustrate these gains for order type 1 and 2, respectively, in case of the balanced workloads under the DYN-SEPT combination, the different values of utilization rate ρ and coefficient of heterogeneity c .

The pooling gain is almost constant and generally more beneficial for order type 2 (long processing time), while heterogeneity and utilization have a more pronounced impact for order type 1 (short processing time).

In particular, the pooling gains for the DDLT of type 1 increase as the coefficient c decreases and the utilization ρ increases. If the incoming order of type 1 finds the facility processing a type 2 order then its sojourn time will increase with the coefficient c and, therefore, its DDLT. For example, when $c = 4$ the value of $\varepsilon(\overline{DDLT}_1)$ is equal to 34.02%, while when $c = 8$ then $\varepsilon(\overline{DDLT}_1) = -2.22\%$. The pooling gains for type 1 increase with its traffic load because the efficiency of the scheduling policy which prioritizes more orders of that type increases with the utilization rate. This will consequently improve the due date performance of type 1 in the pooled configuration. For example, when $c = 8$ the value of $\varepsilon(\overline{DDLT}_1)$ increases from -2.22% to 48.63% when ρ increases from 0.6 to 0.9.

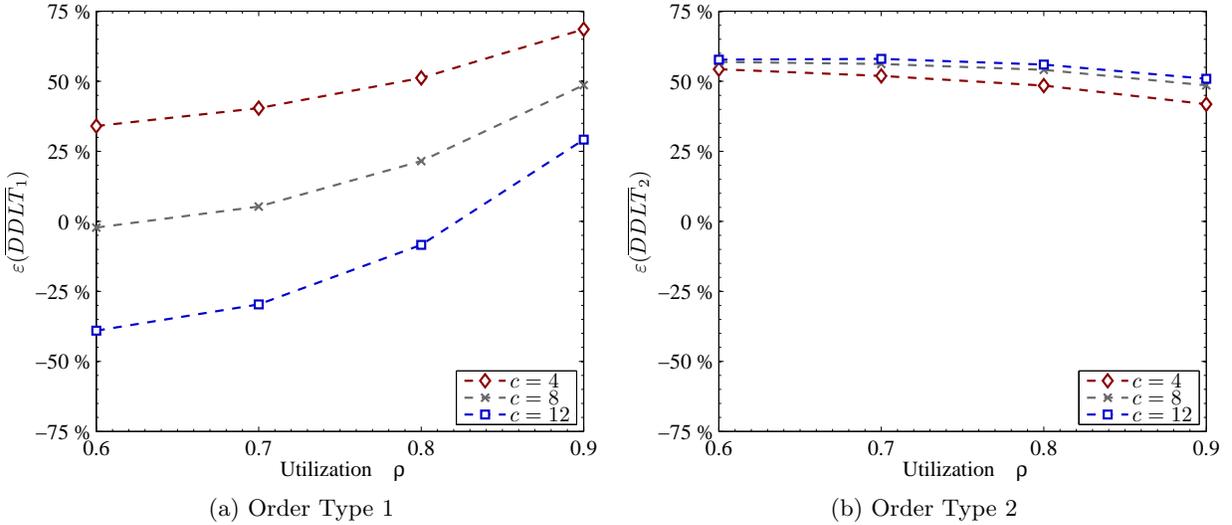


Figure 3: Pooling gains (in %) for the DDLT of each order type under the balanced case, the DYN-SEPT policy combination and different values of c and ρ

For the order type with long expected processing time, we observe the opposite trend although much weaker when $\rho_1 \geq \rho_2$ i.e. $\delta \geq 1$ (see Table 1 and Table 2). For $\rho_2 > \rho_1$ i.e. $\delta < 1$ and higher differences in traffic loads (see Table 2) the pooling gain for type 2 increases in ρ . This is because the facility will face more orders of that type while there will be considerably less orders of the type 1. Then, since the processing times are n times faster, the average DDLT of the type 2 will be considerably lower than in the dedicated configuration.

Moreover, the pooling gains for type 2 slightly increase in c . In the pooled configuration the incoming order of type 2 would wait less time for its processing than in the dedicated configuration when c increases because the incoming orders can find in the pooled queue orders of type 1 with much shorter processing times. The policy which gives higher priority to type 1, however, does not hurt considerably type 2. We later show in the analytical study on the pooling gains for each order type (see Section 5.1) that, for a multi-type system, order types with longer processing times benefit always from pooling.

5 Analytical study and the impact of the modeling assumptions

In this section we address several questions that arise when analyzing the pooling gain on the DDLT performance measures. The first question aims to provide additional insights about the impact of heterogeneity and the utilization rate on the pooling gains. In order to answer this question we will study analytically the pooling gains for the expected sojourn time when using the SEPT policy. There are several reasons for this choice. The DDLT that can be promised is influenced by an order's expected sojourn time. Also, the SEPT policy minimizes the expected sojourn time and its computation is analytically tractable. The next questions deal with two key assumptions we made in our preliminary analysis. In Section 5.2 we examine the impact of the number of pooled systems, and in Section 5.3 we question the modeling assumption of a single server for the pooled system.

5.1 Pooling gains for the expected sojourn time under the SEPT policy

Let $E[S_i^d]$ be the expected sojourn time of orders of type i in configuration d with n order types. We calculate $E[S_i^d]$ based on the formula for an M/M/1 queueing system:

$$E[S_i^d] = \frac{\tau_i}{(1 - \rho_i)}. \quad (5)$$

For configuration p with the SEPT policy, let $E[S_i^p]$ be the expected sojourn time of orders of type i in p . In order to calculate $E[S_i^p]$ we assume the order types $1, 2, \dots, n$ are sorted in ascending order of their expected processing times so that $\tau_1 \leq \tau_2 \leq \dots \leq \tau_n$. Thus, $E[S_i^p]$ can be computed based on the formula for the expected waiting time (Cox and Smith, 1961):

$$E[W_i] = \frac{\sum_{j=1}^n \rho_j \cdot E(R_j)}{(1 - \sum_{j=1}^i \rho_j) \cdot (1 - \sum_{j=1}^{i-1} \rho_j)}, \quad (6)$$

where $E(R_j)$ represents the expected residual processing time of order type j and is equal to τ_j for the exponentially distributed processing times.

Noting that in configuration p the traffic load and the mean processing time of type i are ρ_i/n and τ_i/n , respectively, $E[S_i^p]$ is calculated with the following expression:

$$E[S_i^p] = \frac{\sum_{j=1}^n \rho_j \tau_j}{(n - i \cdot \rho_i)(n - (i - 1)\rho_i)} + \frac{\tau_i}{n}. \quad (7)$$

The global (total) expected sojourn time for production configuration $s \in \{d, p\}$ with n order types, $E[S_g^s]$, is calculated as follows:

$$E[S_g^s] = \frac{\sum_{i=1}^n \lambda_i E[S_i^s]}{\sum_{i=1}^n \lambda_i}. \quad (8)$$

In the following two subsections we start by computing the pooling gains for the expected sojourn times under the SEPT policy when the workloads of the different order types are equal (balanced case) and then, we show how our results generalize to the case where the workloads are different (unbalanced case).

5.1.1 Balanced case

For the balanced production configuration with n order types, the traffic loads across order types are equal i.e. $\rho = \rho_i \forall i = 1, \dots, n$. Then, based on the equations (5)-(8), the pooling gain for the global expected sojourn time $\varepsilon(E[S_g])$ can be expressed in terms of ρ , n , and τ_i with the following expression:

$$\varepsilon(E[S_g]) = 1 - \frac{1 - \rho}{n} \left(1 + \frac{\rho \cdot \sum_{i=1}^n \tau_i}{\sum_{i=1}^n \tau_i (n - i\rho) \cdot (n - (i - 1)\rho)} \right). \quad (9)$$

Note that pooling outperforms the dedicated configuration when $0 < \varepsilon(E[S_g]) \leq 1$.

In order to better understand how heterogeneity in processing times/arrival rates influences the pooling gain represented by equation 9, we consider that the coefficient of heterogeneity, defined as the ratio of the expected processing times or arrival rates between successive order types, is constant. In other words, $c = \tau_i/\tau_{i-1}$ and $c = \lambda_{i-1}/\lambda_i$, where $c \geq 1$. As a consequence $\tau_i = c^{i-1} \cdot \tau_1$ and $\lambda_i = \lambda_1/c^{i-1}$. In this way, lower values of c (i.e., close to 1) reflect the case when order types

are homogeneous, while higher values of c reflect higher heterogeneity between processing times or arrival rates. This stylized model enables us to summarize a large variety of instances with only three parameters ρ , c and n . The pooling gain can now be expressed in terms of these three parameters:

$$\varepsilon(E[S_g]) = 1 - \frac{1-\rho}{n} \left(1 + \frac{\rho \cdot \sum_{i=1}^n c^{i-1}}{\sum_{i=1}^n c^{i-1} (n-i\rho) \cdot (n-(i-1)\rho)} \right). \quad (10)$$

Similarly, based on equations (5) and (7) and replacing τ_i by $c^{i-1} \cdot \tau_1$, we obtain the pooling gain on the expected sojourn time of order type i , $\varepsilon(E[S_i])$:

$$\varepsilon(E[S_i]) = 1 - \frac{1-\rho}{n} \left(1 + \frac{\sum_{j=1}^n c^{j-1}}{c^{i-1}} \cdot \frac{n\rho}{(n-i\rho)(n-(i-1)\rho)} \right). \quad (11)$$

Note that from equations (10) and (11) the pooling gains increase with the utilization rate ρ . For example, for the configuration with 2 order types, pooling is advantageous for both order types when $c < 8.33$ and $\rho = 0.6$, while when $\rho = 0.9$ it is advantageous for much higher values of c ($c < 22.22$).

5.1.2 Unbalanced case

For the unbalanced production configuration with n order types, the traffic loads across order types are unequal i.e. $\rho_1 \neq \rho_2 \neq \dots \neq \rho_n$. In order to obtain the insights into the pooling gains for different levels of heterogeneity in traffic loads of different order types, we assume that the traffic load of order type i can be represented in terms of the traffic load of its preceding type $i-1$ and coefficient $\delta > 0$ so that $\rho_i = \rho_{i-1}/\delta$. Then, ρ_i can be represented in terms of ρ_1 i.e. $\rho_i = \rho_1/\delta^{i-1}$. In other words, by setting $\delta > 1$ the order types will be ranked in decreasing order of traffic load. Furthermore, when $\delta > 1$ the highest traffic load corresponds to the type with the shortest processing time (recall also that $\tau_i = c^{i-1} \cdot \tau_1$).

Then, based on the expressions (5)-(8), and representing ρ_i and τ_i in terms of δ and c , respectively, the pooling gain for the global expected sojourn time under the unbalanced case is provided by the following expression:

$$\begin{aligned} \varepsilon(E[S_g]) &= 1 - \frac{\rho_1 \sum_{i=1}^n \left(\frac{c}{\delta}\right)^{i-1}}{\sum_{i=1}^n (\delta^{i-1} - \rho_1)^{-1}} \cdot \sum_{i=1}^n \frac{(c\delta)^{1-i}}{\left(n - \rho_1 \sum_{j=1}^i \delta^{1-j}\right) \left(n - \rho_1 \sum_{j=1}^{i-1} \delta^{1-j}\right)} \\ &\quad - \frac{1}{n} \cdot \frac{\sum_{i=1}^n \delta^{1-i}}{\sum_{i=1}^n (\delta^{i-1} - \rho_1)^{-1}}. \end{aligned} \quad (12)$$

Similarly, the pooling gain for the expected sojourn time of order type i is given by equation (13).

$$\varepsilon(E[S_i]) = 1 - \frac{\delta^{i-1} - \rho_1}{(c\delta)^{i-1}} \cdot \left(\frac{\rho_1 \cdot \sum_{j=1}^n \left(\frac{c}{\delta}\right)^{j-1}}{\left(n - \rho_1 \sum_{j=1}^i \delta^{1-j}\right) \cdot \left(n - \rho_1 \sum_{j=1}^{i-1} \delta^{1-j}\right)} + \frac{c^{i-1}}{n} \right). \quad (13)$$

These gains depend on c , δ , ρ_1 and n .

Finally, in order to analyze whether the sojourn time is a good indicator of the performance of the pooling strategy observed in the previous section, we present in Figure 4 results for the global pooling gains for the DDLT under the DYN-SEPT policy combination and for the expected sojourn time under the SEPT policy. The results are presented for the balanced and unbalanced case.

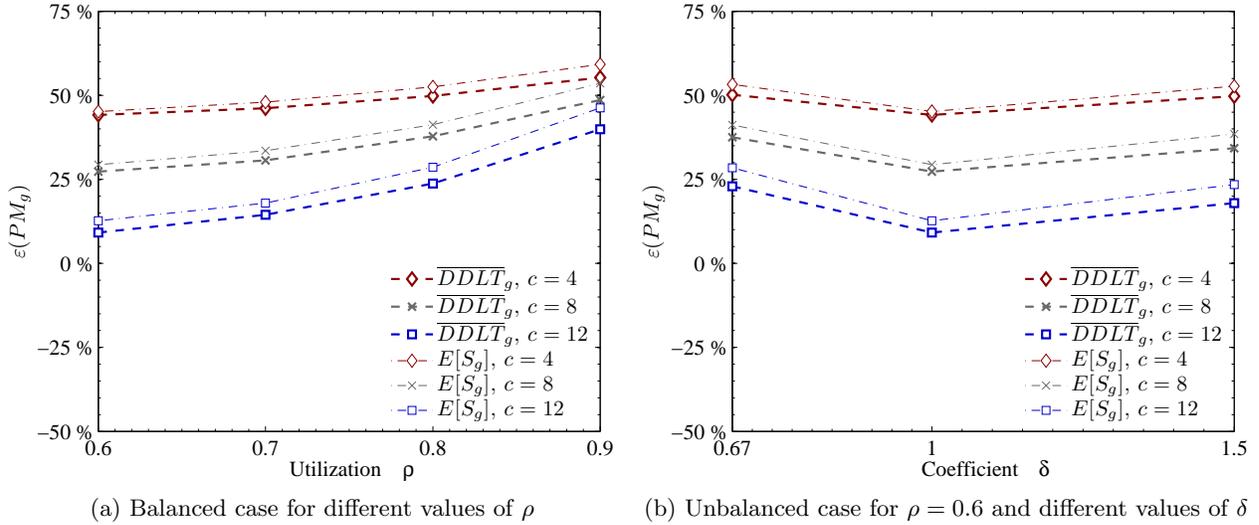


Figure 4: Pooling gain for the global expected sojourn time under the SEPT policy and the global average DDLT under the DYN-SEPT policy combination for different values of c

We observe a remarkable correspondence between both sets of curves. This is important because to obtain the pooling gains in terms of the due-date performance measures is very difficult task, while there exists a simple analytical treatment which can be used as a good indicator of these gains. We nevertheless observe a small increase in the gap between the two sets of curves when ρ or asymmetry in workloads increases.

5.2 Impact of increasing the product variety on the pooling gain

Given that we could observe the high degree of similarity between the pooling gains for the expected sojourn time and the DDLT, we will base our further analysis on the expected sojourn time which is analytically tractable for the pooled single-facility configuration. In order to study how increasing product variety influences the pooling gain for the expected sojourn time, we represent c in terms of the shortest and the longest expected processing times and the number of order types, $c = \sqrt[n]{k}$, where $k = \tau_n/\tau_1 > 1$. In this way, we can compare the pooling gains if we increase the number of order types between the slowest and the fastest order types.

In a similar way, let us define $\delta = \sqrt[n-1]{b}$ where $b = \rho_1/\rho_n$. Then, to study the pooling gains for different values of b and n under the same global utilization $\rho = \sum_{i=1}^n \rho_i/n$, ρ_1 can be replaced by $\sum_{i=1}^n \rho_i b / \sum_{i=1}^n b^{i-1}$ in equation (12). The additional constraint for the stability is $\sum_{i=1}^n \rho_i < \min \left\{ \sum_{i=1}^{n-1} b^{i-1}/b, \sum_{i=1}^{n-1} b^{i-1} \right\}$.

The global pooling gains are illustrated in Figure 5a when n and k vary and in Figure 5b when n and b vary.

The two-order type system gives consistently the smallest improvement on the expected sojourn time when pooling (a proof is presented in Appendix B). Another observation is that the sensibility towards the heterogeneity of the orders is also maximized in the two-order type case. As a result, the two-order type model was reasonable to consider for our analysis as it constitutes a lower bound on the benefit of pooling.

It is noteworthy that Figure 5b is almost symmetric in terms of the global pooling gain, because

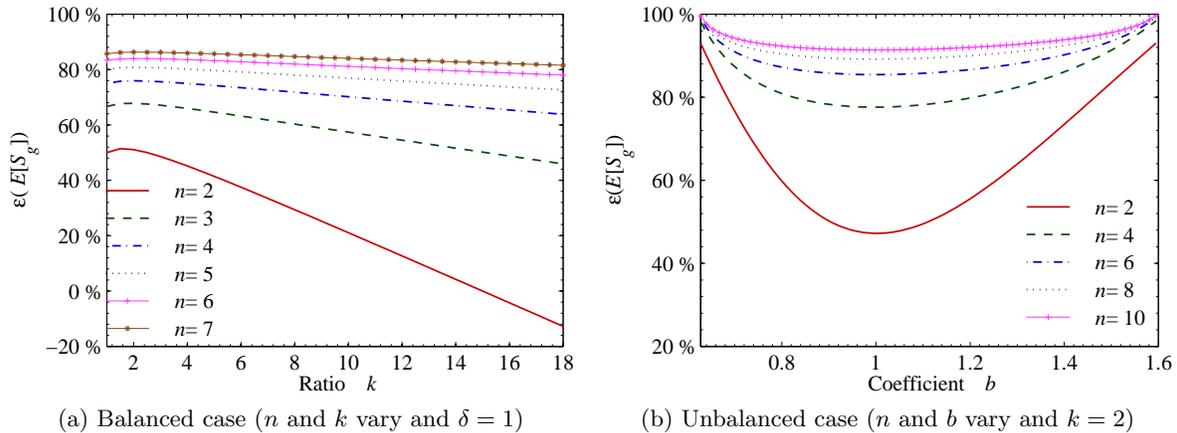


Figure 5: Pooling gain for the global expected sojourn time when $\rho = 0.6$

this gain is similar when $\rho_2 = \delta\rho_1$ and $\rho_1 = \delta\rho_2$. Similar tendency is observed when the average DDLT is considered (see Table 2) where, as c decreases these gains become more similar under all the considered policy combinations.

As a final observation let us also mention that $n = 2$ seems also to be the most relevant case from a practical point of view since the coordination cost for pooling will increase very quickly with n .

5.3 Pooled multi-server configurations

In the previous sections, we modeled the pooled facility as a single server with an increased capacity. We will now investigate the impact of this modeling assumption by comparing it with an alternative model for the pooled facility. Instead of having a high capacity $M/G/1$ queueing system we have an $M/G/n$ queueing system.

It can easily be shown that the expected sojourn time will be longer in the multiple-server model than in the single-server model with equivalent capacity. On the other hand, the expected queueing time will be shorter; this gives an indication that the sojourn time will not be too different. In order to evaluate the impact of our modeling assumption on the DDLT we performed a simulation study for the 2-order type case. We choose this setting because with two order types the contrast between orders is the largest and the pooling decision is the most difficult. Note that the mean processing times for the two order types are the same as in the dedicated configuration. We compute the pooling gains for the global average DDLT for the balanced case when the CON-SEPT policy combination is employed. We present the results for the two pooled configurations in Figure 6. We add superscript $M/G/2$ to the performance measure in the pooled two-facility configuration.

We observe that the pooling gains of both pooled configurations are similar especially under the higher utilization rate. This confirms that our single server modeling assumption does not significantly influence our results. Note that, the higher difference is only observed under the very high heterogeneity ($c = 12$) and low utilization rate. However, in both configurations the pooling gains are negative.

Putting those results in perspective with the literature on flexibility it seems that the same type of conclusion would be reached with a large variety of configurations for the pooled system. Just how different configurations would affect the pooling gains is beyond the scope of this article.

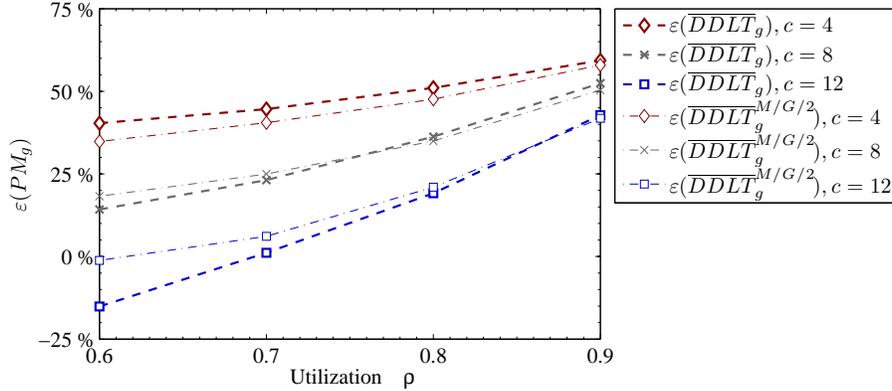


Figure 6: Pooling gains for the global average DDLT for the two pooled configurations under the CON-SEPT for the balanced case and different values of c and ρ

6 Conclusions

Reaping economies of scale is a tempting idea for companies in their continuous struggle to increase their competitiveness. Pooling operations with uncertain demand can be an important source of economies of scale. Indeed, above the direct scale effect that usually brings cost savings, the pooling of the uncertainty of demand has been widely shown to bring important economies of scale in terms of waiting time and inventory. Nevertheless the lean management philosophy warns us that any increase in variability should be repressed. For firms pondering over the possibility of pooling resources (internally or externally) of heterogeneous product lines, it is thus necessary to get a deeper understanding about the interplay between economies of scale and variability.

In this article we shed a new light on this dilemma by taking into account the fact that in practice most firms operate with due dates. For the global planning of a supply chain, reliable due dates are imperative. The goal is then to keep those due dates as short as possible in order to increase the attractiveness of the firm. The good news is that we could show that the additional flexibility provided by the due-date mechanism can be used to overcome the additional variability introduced by the heterogeneity of the pooled products. This is not as obvious as might seem given that, to the best of our knowledge, the literature so far only studied pooling based on FCFS queueing and consequently reached very different conclusions.

Our first conclusion is thus that the level of heterogeneity in terms of processing time does not have a significant negative impact on the due-date performance if the system is adequately managed. Even more interesting is the fact that this conclusion is even stronger when the utilization rate is high – the case where pooling should be particularly attractive – This means that the economies of scale will in general largely dominate the heterogeneity impact. This represents a complete reversal from the conclusions reached so far for FCFS queues where heterogeneity has in general a significant negative impact and this negative impact increases with the utilization rate. In such situation the conclusion is in general that in the presence of heterogeneity pooling should not be pursued.

Our second conclusion is that in terms of workload, heterogeneity brings benefits. Indeed it appears that pooling products with different workloads will in general bring benefits that increase with the workload imbalance between the products. As a conclusion, one might really say that when pooling manufacturing operations with due dates, opposites should attract each other.

This work also raises new questions. As mentioned earlier, it would be interesting to study further how to best exploit different flexible configurations in the framework of manufacturing

systems with heterogeneous orders and lead times. This type of environment is absent from the extant literature on flexibility. Another interesting research avenue would be to analyze the case where the customers are also heterogeneous in terms of their sensitivity to the due date lead time. Would it be possible to get additional benefits from pooling in such circumstances? This probably requires analyzing more sophisticated scheduling rules which is itself a topic that could help further increase our understanding of the potential benefits of pooling for manufacturing systems.

Appendix A

Table 1: The pooling gains for the DDLT under the balanced case

ρ	DDSP-SP	$c = 4$			$c = 8$			$c = 12$		
		$\varepsilon(\overline{DDLT}_1)$	$\varepsilon(\overline{DDLT}_2)$	$\varepsilon(\overline{DDLT}_g)$	$\varepsilon(\overline{DDLT}_1)$	$\varepsilon(\overline{DDLT}_2)$	$\varepsilon(\overline{DDLT}_g)$	$\varepsilon(\overline{DDLT}_1)$	$\varepsilon(\overline{DDLT}_2)$	$\varepsilon(\overline{DDLT}_g)$
0.6	CON-SEPT	30.58	50.12	40.36	-25.12	53.53	14.25	-82.63	53.60	-15.13
	CON-EDD	17.10	54.71	35.91	-49.07	56.43	3.70	-113.26	56.59	-29.13
	CON-SLACK	4.97	57.72	31.35	-70.86	60.06	-5.35	-149.24	60.54	-45.28
	DYN-SEPT	34.02	54.33	44.15	-2.22	56.92	27.32	-39.03	57.75	9.15
	DYN-EDD	31.77	56.90	44.31	-5.66	58.84	26.55	-43.40	59.69	7.92
	DYN-SLACK	29.99	57.48	43.71	-8.69	59.51	25.36	-47.02	60.29	6.44
0.7	CON-SEPT	42.15	47.15	44.62	-4.36	50.66	23.13	-50.72	52.67	1.13
	CON-EDD	26.45	52.96	39.65	-31.28	56.21	12.50	-90.92	57.48	-16.48
	CON-SLACK	9.47	57.73	33.57	-62.97	60.93	-0.95	-135.76	61.43	-36.83
	DYN-SEPT	40.45	51.93	46.17	5.27	56.21	30.63	-29.63	58.04	14.47
	DYN-EDD	36.73	56.58	46.61	-0.17	59.16	29.34	-36.10	60.99	12.71
	DYN-SLACK	33.60	57.86	45.67	-4.27	60.32	27.89	-43.72	61.42	9.12
0.8	CON-SEPT	58.20	43.85	51.04	24.66	48.48	36.51	-10.70	49.56	19.23
	CON-EDD	42.06	50.84	46.41	-3.62	54.53	25.36	-51.15	54.88	1.52
	CON-SLACK	18.33	56.70	37.44	-42.41	59.29	8.26	-99.83	60.18	-20.37
	DYN-SEPT	51.22	48.48	49.83	21.53	54.10	37.81	-8.39	55.97	23.75
	DYN-EDD	45.51	55.84	50.67	14.47	59.76	37.11	-18.41	60.14	20.77
	DYN-SLACK	41.18	58.22	49.69	6.56	60.89	33.69	-27.91	62.36	17.16
0.9	CON-SEPT	77.08	42.25	59.34	58.79	45.79	52.45	40.14	45.52	42.79
	CON-EDD	62.11	48.35	55.09	35.13	49.91	42.49	5.55	51.49	28.11
	CON-SLACK	28.42	56.64	42.79	-8.84	57.28	23.99	-35.10	56.13	9.71
	DYN-SEPT	68.61	41.83	55.29	48.63	48.58	48.62	29.18	50.89	39.91
	DYN-EDD	61.95	52.66	57.31	39.40	57.24	48.29	18.33	59.50	38.76
	DYN-SLACK	55.68	56.46	56.06	28.60	60.16	44.34	3.77	62.39	32.87

Table 2: The pooling gains for the DDLT under the unbalanced case

ρ	δ	DDSP-SP	$c = 4$			$c = 8$			$c = 12$		
			$\varepsilon(\overline{DDL T}_1)$	$\varepsilon(\overline{DDL T}_2)$	$\varepsilon(\overline{DDL T}_g)$	$\varepsilon(\overline{DDL T}_1)$	$\varepsilon(\overline{DDL T}_2)$	$\varepsilon(\overline{DDL T}_g)$	$\varepsilon(\overline{DDL T}_1)$	$\varepsilon(\overline{DDL T}_2)$	$\varepsilon(\overline{DDL T}_g)$
0.60	0.67	CON-SEPT	7.99	64.62	49.54	-69.10	66.34	30.56	-148.00	65.85	8.46
		CON-EDD	-12.50	67.41	46.14	-104.69	68.48	22.75	-196.70	67.84	-3.19
		CON-SLACK	-30.21	69.10	42.68	-142.19	70.45	14.29	-253.04	70.08	-16.69
		DYN-SEPT	16.29	64.13	50.16	-33.22	66.59	37.54	-84.93	67.09	22.95
		DYN-EDD	13.51	65.86	50.58	-38.33	67.52	36.69	-90.76	67.90	21.84
		DYN-SLACK	11.39	66.47	50.39	-42.42	67.86	35.74	-97.03	68.46	20.41
0.60	1.50	CON-SEPT	52.80	36.36	48.46	16.50	40.67	22.87	-20.75	42.09	-4.13
		CON-EDD	45.27	43.04	44.68	3.74	46.09	14.91	-39.25	47.17	-16.40
		CON-SLACK	38.50	47.97	41.01	-8.88	50.87	6.86	-57.01	52.10	-28.16
		DYN-SEPT	51.72	44.94	49.75	27.93	49.59	34.26	4.24	50.85	17.93
		DYN-EDD	49.80	49.15	49.59	26.05	52.50	33.80	1.65	53.72	16.96
		DYN-SLACK	48.69	50.48	49.20	24.20	53.41	32.74	-0.51	54.69	15.69
0.75	0.67	CON-SEPT	19.79	78.83	70.42	-47.24	79.37	60.87	-115.15	79.77	51.28
		CON-EDD	-9.28	80.73	67.90	-100.78	81.22	54.61	-187.32	81.60	42.30
		CON-SLACK	-42.25	82.33	64.56	-161.02	82.55	46.94	-287.55	83.22	29.06
		DYN-SEPT	19.92	76.43	66.18	-32.18	77.59	57.25	-85.14	77.65	47.15
		DYN-EDD	13.34	78.20	66.44	-42.85	79.09	56.46	-99.46	79.14	45.64
		DYN-SLACK	8.10	78.87	66.02	-52.82	79.62	55.06	-112.35	79.65	43.68
0.75	1.50	CON-SEPT	79.70	11.74	69.93	64.14	19.73	57.77	48.94	23.03	45.23
		CON-EDD	73.95	24.50	66.84	54.33	30.65	50.92	34.66	33.74	34.51
		CON-SLACK	67.45	35.57	62.87	43.41	40.47	42.98	17.86	42.33	21.37
		DYN-SEPT	74.25	29.05	66.06	59.97	37.92	55.94	46.63	40.29	45.49
		DYN-EDD	71.57	39.52	65.74	57.05	45.53	54.96	42.85	46.83	43.55
		DYN-SLACK	70.08	42.78	65.11	54.58	47.53	53.29	39.46	49.01	41.16

Appendix B

For the comparison of the production configurations with different number of order types, we add an additional superscript n to the appropriate notation. In order to show that $\varepsilon(E[S_g^n]) \leq \varepsilon(E[S_g^{n+1}])$ it is sufficient that $E[S_g^{d,n}] \leq E[S_g^{d,n+1}]$ and $E[S_g^{p,n}] \geq E[S_g^{p,n+1}]$.

The inequality $E[S_g^{d,n}] \leq E[S_g^{d,n+1}]$ can be written as follows:

$$\frac{n}{n+1} \leq \frac{\sum_{i=1}^n k^{\frac{i-1}{n-1}}}{\sum_{i=1}^{n+1} k^{\frac{i-1}{n}}}. \quad (14)$$

The right side of inequality 14 is increasing function of k and it achieves the minimum (i.e. $\frac{n}{n+1}$) when $k = 1$. Then, the inequality is satisfied for any value of $k \geq 1$.

Based on equations 7 and 8, the global expected sojourn time for system p with n order types can be expressed as:

$$E[S_g^{p,n}] = \rho\tau_1 k \sum_{i=1}^n \left[k^{\frac{i-1}{n-1}} (n-i\rho)(n-(i-1)\rho) \right]^{-1} + \tau_1 k \left[\sum_{i=1}^n k^{\frac{i-1}{n-1}} \right]^{-1}. \quad (15)$$

Equation 15 shows that increasing the number of order types between the order types with the minimum and maximum expected processing times will decrease the variance of the processing times of the pooled facility and will consequently lead to a decrease in the global expected sojourn time. Therefore, $E[S_g^{p,n}]$ is decreasing in n .

References

- Baker, K. R., Bertrand, J. W. M., 1982. A dynamic priority rule for scheduling against due-dates. *Journal of Operations Management* 3 (1), 37–42.
- Barman, S., 1998. The impact of priority rule combinations on lateness and tardiness. *IIE Transactions* 30, 495–504.
- Bassamboo, A., Randhawa, R., Mieghem, J. V., 2010a. A little flexibility is all you need: Asymptotic optimality of tailored chaining and pairing in queuing systems. Tech. rep., Working paper, Northwestern University, Evanston, IL.
- Bassamboo, A., Randhawa, R., Mieghem, J. V., 2010b. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Science* 56 (8), 12851303.
- Benjaafar, S., 1995. Performance bounds for the effectiveness of pooling in multi-processing systems. *European Journal of Operational Research* 87, 375–388.
- Benjaafar, S., Cooper, W. L., Kim, J.-S., 2005. On the benefits of pooling in production-inventory systems. *Management Science* 51 (4), 548 – 565.
- Bookbinder, J. H., Noor, A., 1985. Setting job-shop due-dates with service-level constraints. *Journal of the Operational Research Society* 36 (11), 1017–1026.
- Buzacott, J., 1996. Commonalities in reengineered business processes: Models and issues. *Management Science* 42, 768–782.

- Cheng, T., Gupta, M., 1989. Survey of scheduling research involving due date determination decisions. *European Journal of Operational Research* 38, 156–166.
- Cox, D., Smith, W., 1961. *Queues*. Methuen Co. Ltd, London.
- Enns, S. T., Grewal, C. S., 2010. Lot size optimisation under pooled and unpooled scenarios in batch production system. *International Journal of Production Research* 48 (20), 6085–6101.
- Graves, S., Tomlin, B., 2003. Process flexibility in supply chains. *Management Science* 49 (7), 907–919.
- Gupta, D., Gerchak, Y., 2002. Quantifying operational synergies in a merger/acquisition. *Management Science* 48 (4), 517–533.
- Gurumurthi, Benjaafar, S., 2004. Modeling and analysis of flexible queueing systems. *Naval Research Logistics* 51 (5), 755–782.
- Hella Australia Pty Ltd, 07 July 2007. Behr hella service - bringing the world's best thermal management products to the Australian automotive aftermarket. [press release]. Available at http://www.hella.com/produktion/HellaAU/WebSite/Channels/Company/Press/National_Press/Company_News/BehrHellaPR.jsp, [Accessed 15 November 2012].
- Hopp, W. J., Sturgis, M. L. R., 2000. Quoting manufacturing due dates subject to a service level constraint. *IIE Transactions* 32, 771–784.
- Iyer, A., Jain, A., 2004. Modeling the impact of merging capacity in production-inventory systems. *Management Science* 50 (8), 1082–109.
- Jain, A., 2007. Value of capacity pooling in supply chains with heterogeneous customers. *European Journal of Operational Research* 177, 239–260.
- Jordan, W., Graves, S., 1995. Principles on the benefits of manufacturing process flexibility. *Management Science* 41 (4), 577–594.
- Kelly, F., Massouli, L., Walto, N., 2009. Resource pooling in congested networks: proportional fairness and product form. *Queueing Systems: Theory and Applications* 63 (1-4), 165–194.
- Keskinocak, P., Tayur, S., 2004. Due-date management policies. In: Simchi-Levi, D., Wu, S., Shen, Z. J. (Eds.), *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*. International Series in Operations Research and Management Science, Kluwer Academic Publishers, Norwell, MA, pp. 485–553.
- Koulamas, C., Kyparisis, G., 2008. Single-machine scheduling with waiting-time-dependent due dates. *European Journal of Operational Research* 191, 577–581.
- Mandelbaum, A., Reiman, M. I., 1998. On pooling in queueing networks. *Management Science* 44 (7), 971–981.
- Rothkopf, M. H., Rech, P., 1987. Perspectives on queues: Combining queues is not always beneficial. *Operations Research* 35 (6), 906–909.
- Sheikhzadeh, M., Benjaafar, S., Gupta, D., 1998. Machine sharing in manufacturing systems: Total flexibility versus chaining. *The International Journal of Flexible Manufacturing Systems* 10, 351–378.

- Smith, D. R., Whitt, W., 1981. Resource sharing for efficiency in traffic systems. *The Bell System Technical Journal* 60 (1), 39–55.
- So, K., Song, J., 1998. Price, delivery time guarantees and capacity selection. *European Journal of Operational Research* 111, 28–49.
- Sridharan, V., Li, X., 2008. Improving delivery reliability by a new due-date setting rule. *European Journal of Operational Research* 186, 1201–1211.
- Stidham, S., 1970. On the optimality of single-server queueing systems. *Operations Research* 18, 708–732.
- Swinnen, J., Herck, K. V., 2011. Poland, bulgaria and romania: Social impact of discount and organized food retail formats on remote regions. Tech. Rep. 16, FAO Investment Centre/EBRD Cooperation Programme. Report Series, nr.16.
- Tekin, E., Hopp, W., van Oyen, M., 2009. Pooling strategies for call center agent cross-training. *IIE Transactions* 41 (6), 546–561.
- van Dijk, N. M., van der Sluis, E., 2008. To pool or not to pool in call centers. *Production and Operations Management* 17, 296 – 305.
- Veral, E., 2001. Computer simulation of due-date setting in multi-machine job shops. *Computers and Industrial Engineering* 41 (1), 77–94.
- Vinod, V., Sridharan, R., 2011. Simulation modeling and analysis of due-date assignment methods and scheduling decision rules in a dynamic job shop production system. *International Journal of Production Economics* 129 (1), 127–146.
- Wein, L., Chevalier, P. B., 1992. A broader view of the job-shop scheduling problem. *Management Science* 38 (7), 1018–1033.
- Wein, L. M., 1991. Due date setting and priority sequencing in a multiclass M/G/1 queue. *Management Science* 37 (7), 834–850.
- Whitt, W., 1992. Understanding the efficiency of multiserver service systems. *Management Science* 38 (5), 708–723.

Recent titles

CORE Discussion Papers

- 2012/62 Francis BLOCH and Axel GAUTIER. Strategic bypass deterrence.
- 2012/63 Olivier DURAND-LASSERVE, Axel PIERRU and Yves SMEERS. Sensitivity of policy simulation to benchmark scenarios in CGE models: illustration with carbon leakage.
- 2013/1 Pierre PESTIEAU and Maria RACIONERO. Harsh occupations, health status and social security.
- 2013/2 Thierry BRECHET and Henry TULKENS. Climate policies: a burden or a gain?
- 2013/3 Per J. AGRELL, Mehdi FARSI, Massimo FILIPPINI and Martin KOLLER. Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems.
- 2013/4 Adel HATAMI-MARBINI, Per J. AGRELL and Nazila AGHAYI. Imprecise data envelopment analysis for the two-stage process.
- 2013/5 Farhad HOSSEINZADEH LOTFI, Adel HATAMI-MARBINI, Per J. AGRELL, Kobra GHOLAMI and Zahra GHELEJ BEIGI. Centralized resource reduction and target setting under DEA control.
- 2013/6 Per J. AGRELL and Peter BOGETOFT. A three-stage supply chain investment model under asymmetric information.
- 2013/7 Per J. AGRELL and Pooria NIKNAZAR. Robustness, outliers and Mavericks in network regulation.
- 2013/8 Per J. AGRELL and Peter BOGETOFT. Benchmarking and regulation.
- 2013/9 Jacques H. DREZE. When Borch's Theorem does not apply: some key implications of market incompleteness, with policy relevance today.
- 2013/10 Jacques H. DREZE. Existence and multiplicity of temporary equilibria under nominal price rigidities.
- 2013/11 Jean HINDRIKS, Susana PERALTA and Shlomo WEBER. Local taxation of global corporation: a simple solution.
- 2013/12 Pierre DEHEZ and Sophie POUKENS. The Shapley value as a guide to FRAND licensing agreements.
- 2013/13 Jacques H. DREZE and Alain DURRE. Fiscal integration and growth stimulation in Europe.
- 2013/14 Luc BAUWENS and Edoardo OTRANTO. Modeling the dependence of conditional correlations on volatility.
- 2013/15 Jens L. HOUGAARD, Juan D. MORENO-TERNERO and Lars P. OSTERDAL. Assigning agents to a line.
- 2013/16 Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. First-order methods with inexact oracle: the strongly convex case.
- 2013/17 Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. Intermediate gradient methods for smooth convex problems with inexact oracle.
- 2013/18 Diane PIERRET. The systemic risk of energy markets.
- 2013/19 Pascal MOSSAY and Pierre M. PICARD. Spatial segregation and urban structure.
- 2013/20 Philippe DE DONDER and Marie-Louise LEROUX. Behavioral biases and long term care insurance: a political economy approach.
- 2013/21 Dominik DORSCH, Hubertus Th. JONGEN, Jan.-J. RÜCKMANN and Vladimir SHIKHMAN. On implicit functions in nonsmooth analysis.
- 2013/22 Christian M. HAFNER and Oliver LINTON. An almost closed form estimator for the EGARCH model.
- 2013/23 Johanna M. GOERTZ and François MANIQUET. Large elections with multiple alternatives: a Condorcet Jury Theorem and inefficient equilibria.
- 2013/24 Axel GAUTIER and Jean-Christophe POUDOU. Reforming the postal universal service.
- 2013/25 Fabian Y.R.P. BOCART and Christian M. HAFNER. Fair re-valuation of wine as an investment.
- 2013/26 Yu. NESTEROV. Universal gradient methods for convex optimization problems.
- 2013/27 Gérard CORNUEJOLS, Laurence WOLSEY and Sercan YILDIZ. Sufficiency of cut-generating functions.

Recent titles

CORE Discussion Papers - continued

- 2013/28 Manuel FORSTER, Michel GRABISCH and Agnieszka RUSINOWSKA. Anonymous social influence.
- 2013/29 Kent WANG, Shin-Huei WANG and Zheyao PAN. Can federal reserve policy deviation explain response patterns of financial markets over time?
- 2013/30 Nguyen Thang DAO and Julio DAVILA. Can geography lock a society in stagnation?
- 2013/31 Ana MAULEON, Jose SEMPERE-MONERRIS and Vincent VANNETELBOSCH. Contractually stable alliances.
- 2013/32 Jean-François CAULIER, Ana MAULEON and Vincent VANNETELBOSCH. Allocation rules for coalitional network games.
- 2013/33 Georg KIRCHSTEIGER, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Limited farsightedness in network formation.
- 2013/34 Ana MAULEON and Vincent VANNETELBOSCH. Relative concerns and delays in bargaining with private information.
- 2013/35 Kristof BOSMANS, Koen DECANCO and Erwin OOGHE. What do normative indices of multidimensional inequality really measure?
- 2013/36 Alain PHOLO BALA, Dominique PEETERS and Isabelle THOMAS. Spatial issues on a hedonic estimation of rents in Brussels.
- 2013/37 Lionel ARTIGE, Antoine DEDRY and Pierre PESTIEAU. Social security and economic integration.
- 2013/38 Nicolas BOUCKAERT and Erik SCHOKKAERT. Differing types of medical prevention appeal to different individuals.
- 2013/39 Pierre M. PICARD. Trade, economic geography and the choice of product quality.
- 2013/40 Tanja B. MLINAR and Philippe CHEVALIER. Pooling in manufacturing: do opposites attract?

Books

- G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.
- J. HINDRIKS and I. VAN DE CLOOT (2011), *Notre pension en héritage*. Itinera Institute.
- M. FLEURBAEY and F. MANIQUET (2011), *A theory of fairness and social welfare*. Cambridge University Press.
- V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
- I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
- W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
- L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
- J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
- M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
- J. HINDRIKS (2012), *Gestion publique*. De Boeck.

CORE Lecture Series

- R. AMIR (2002), Supermodularity and complementarity in economics.
- R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
- A. SHAPIRO (2010), Stochastic programming: modeling and theory.