Fundamental behaviour of Virtual-Build-to-Order systems

Philip G Brabazon*, Bart MacCarthy

Mass Customization Research Centre, Operations Management Division, Nottingham University Business School, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB

Abstract

Virtual-Build-to-Order (VBTO) is an emerging order fulfilment system within the automotive sector that is intended to improve fulfilment performance by taking advantage of integrated information systems. The primary innovation in VBTO systems is the ability to make available all unsold products that are in the production pipeline to all customers. In a conventional system the pipeline is inaccessible and a customer can be fulfilled by a product from stock or having a product Built-to-Order (BTO), whereas in a VBTO system a customer can be fulfilled by a product from stock, by being allocated a product in the pipeline, or by a build-to-order product. Simulation is used to investigate and profile the fundamental behaviour of the basic VBTO system and to compare it to a Conventional system. A predictive relationship is identified, between the proportions of customers fulfilled through each mechanism and the ratio of product variety / pipeline length. The simulations reveal that a VBTO system exhibits inherent behaviour that alters the stock mix and levels, leading to stock levels being higher than in an equivalent conventional system at certain variety / pipeline ratios. The results have implications for the design and management of order fulfilment systems in sectors such as automotive where VBTO is a viable operational model.

Keywords: Mass Customization, order fulfillment, simulation, VBTO

1. Introduction

The trends of increasing levels of product variety (Cox & Alm 1998; Bils & Kennow 2001; Barbiroli & Focacci 2003), of offering customers choices over product options (Kotler 1989), and of greater responsiveness required by the market place create challenges for order fulfilment processes (McCutheon *et al* 1994; Salvador & Forza, 2004). Interest has grown in how to engineer the product and process to enable product differentiation to be postponed until after the customer order is received (van Hoek 2001, Yang *et al* 2004). In the basic postponement order fulfilment model there is a decoupling point at which buffer stocks of components or semi-finished products are held in readiness for customer orders, as illustrated in Partanen & Haapasalo (2004), and when an order is received the product is produced and delivered to the requested specification. If the lead time demanded by customers is too short or if production capacity is a constraint there may be ways of satisfying demand by having several decoupling points or mixing an assemble-to-order with a make-to-stock (MTS) strategy (Denton *et al* 2003, Graman & Magazine 2002) but considerable effort may be needed to understand the product/process interaction as demonstrated in the selection of semi-finished vanilla boxes for computer servers by Swaminathan & Tayur (1998, 1999).

The characteristics of some products and supply chains hamper the adoption of the postponement model. For products such as passenger cars, some supply chain activities need to be triggered weeks or months before the date of manufacture. High levels of in-process inventories are unattractive in such systems. It is difficult to identify or create a dominant single decoupling point in the production system where a product can go from a 'generic' state upstream of the decoupling point to being fully specified downstream. How can the producer respond quickly to specific customer choice in such circumstances? Although a pure Build-to-Order (BTO) strategy can be adopted, it would result in customers having to tolerate long waiting times and it fails to satisfy those customers who may be prepared to accept some compromise from their preferred specification but are strongly averse to waiting (Elias 2002).

* Corresponding author. Tel: +44-(0)-115-951-4730/4025; fax +44-(0)-115-951-4000.
E-mail address: philip.brabazon@nottingham.ac.uk

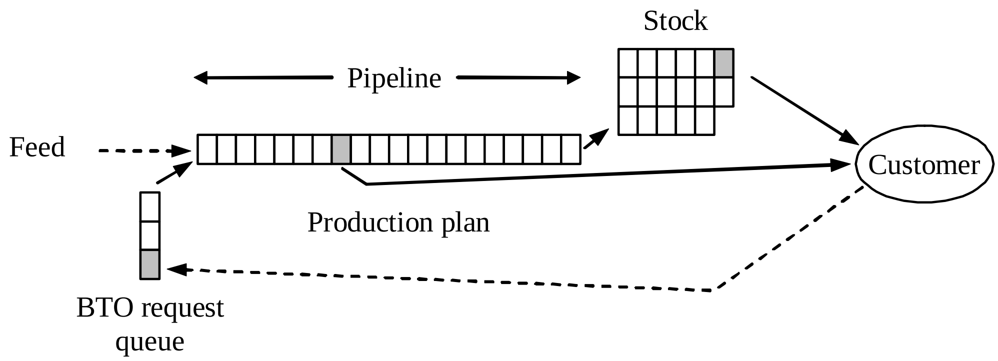
To accommodate these differences between customers, the producer could operate in a mixed BTO/MTS mode. However, information technology has opened the way to a new strategy termed Virtual-Build-to-Order (VBTO) that has evolved in the automotive sector but has applications more widely. It is described in the automotive context as connecting customers *'either via the internet or in dealer's showrooms, to the vast, albeit far-flung, array of vehicles already in existence, including vehicles on dealer's lots, in transit, on assembly line, and scheduled for production'*, with the expectation that *'customers are likely to find a vehicle with the colour and options they most want'* (Agrawal et al. 2001). The upstream part of the pipeline that is planned is the virtual segment, typically driven by demand forecasts but tempered by other considerations such as component availability, capacity bottlenecks and workflow and balancing issues.

The VBTO approach is being used by major automotive firms (see Holweg & Pil (2001) where it is referred to as 'amend-to-order'). The basic VBTO system has two segments – a 'finished stock' segment and a 'pipeline' segment that represents the sequence of products to be, or are being manufactured. Customers can be fulfilled in one of three ways: (1) by a product from stock; (2) by allocating the customer a product in the pipeline; (3) or by triggering a product to be built-to-order (in which case it enters the start of the pipeline). In the VBTO system there is no distinct, static, decoupling point in the process from which customers are linked to products. The ability to allocate products from anywhere from stock to the head of the pipeline has prompted the term 'floating decoupling point' to be coined (Brabazon & MacCarthy 2004). On the surface the operation of a VBTO system appears straightforward and to offer benefits over a Conventional fulfilment system that can fulfil customers from stock or by BTO only. For a mainstream passenger vehicle the pipeline can contain tens of thousands of products and it is intuitive to expect that a VBTO system can achieve the same level of customer fulfilment with less finished stock than a Conventional system. However, there is little analysis of such systems in the literature (Brabazon & MacCarthy 2004, Turner & Williams 2005). The aim of the research reported in this paper is to examine and characterise the fundamental behaviour of VBTO systems. Here we report results from a modelling and simulation study of the VBTO order fulfilment system that set out to investigate its behaviour with respect to the likelihood of fulfilling customer orders, the expected customer waiting time and the inventory characteristics of such systems. The structure of the model and the technical aspects of the simulation approach are first introduced. Results are presented for pipeline only systems with different search mechanisms and for conventional and VBTO systems. The observed behaviour is compared, analysed and discussed and the implications for the design and management of VBTO systems are noted.

2. Research approach

In this study the VBTO system is modelled as a system with three segments: a stock of unsold products, a production plan which is the sequence of products to be manufactured (referred to as the pipeline) and a BTO queue (Figure 1).

Figure 1: Schematic of the VBTO order fulfilment system showing three fulfilment mechanisms



In the model the production rate and customer arrival rate are constant and equal. Products enter the pipeline and customers arrive at a rate of one per period. Once products have entered the pipeline they cannot be re-sequenced. The sequence of products fed into the pipeline is random and uncorrelated. Each customer knows the specification of the product they are seeking and a search is made for a product that has exactly that specification. The customer can be fulfilled from any of the segments. Priority is given to selecting a product from stock, then searching for a product in the pipeline and lastly, if a suitable product has not been found, to initiating a BTO order. When a product is allocated to a customer it is made unavailable to subsequent customers. BTO requests are fed into the start of the pipeline at the earliest opportunity. Therefore, if the BTO request queue is empty an unallocated product is fed into the pipeline.

The performance metric of primary interest is the proportion of customers fulfilled by each mechanism. A second metric is average customer waiting time, for which it should be noted that customers fulfilled from stock are modelled as being fulfilled instantly.

The effects on the metrics of two system parameters and one operating policy are studied: product range, pipeline length, and direction of searching the pipeline. The product range is represented as a set of discrete variants, with each variant assigned a unique number. For example, if the product range has four variants, they are identified as 1, 2, 3 & 4. The effect of product range is studied by stepping from the lowest variety of 21 to a maximum level of 213 (8192). Several pipeline lengths are studied, from 64 to 2048, all of which are powers of 2. The purpose of having both product variety and pipeline length as powers of a number will be seen when results are plotted using the ratio of variety/ product range on logarithmic scales. The pipeline can be searched using either first-in-first-out or last-in-first-out logic. In the former case the product at the downstream end of the pipeline is the first to be inspected, and then each product in sequence back to the start of the pipeline is inspected, unless a match is found in which case the search is terminated. This form of search is referred to as a *backwards* search. The alternative is to inspect first the most recent product to enter the pipeline and then each product in sequence to the downstream exit of the pipeline, unless a match is found. This is referred to as a *forwards* search.

The simulation has been analysed as a non-terminating system. At the start of each simulation run the pipeline is primed with products before the first customer is fulfilled. An appropriate warm-up period is included but system performance is discarded. The replication method of batch-means has been used to capture system performance following the warm up period (Law & Kelton 2000) with the statistics from 9 batches (replications) combined to calculate the performance metrics for each experimental condition. A single length of warm-up period and batch size are not used because of different levels of variance in the performance metrics across the set of experimental conditions studied. In general the warm-up period is between 3 and 10 times the pipeline length and the batch size is between 200 and 5000 customers. The variance reduction method of Common Random Numbers has been employed (Law & Kelton 2000). Two random sequences are used within the model - the specification of products entering the production plan and the customer's desired specification – and each has its own seed. All runs use the same seeds.

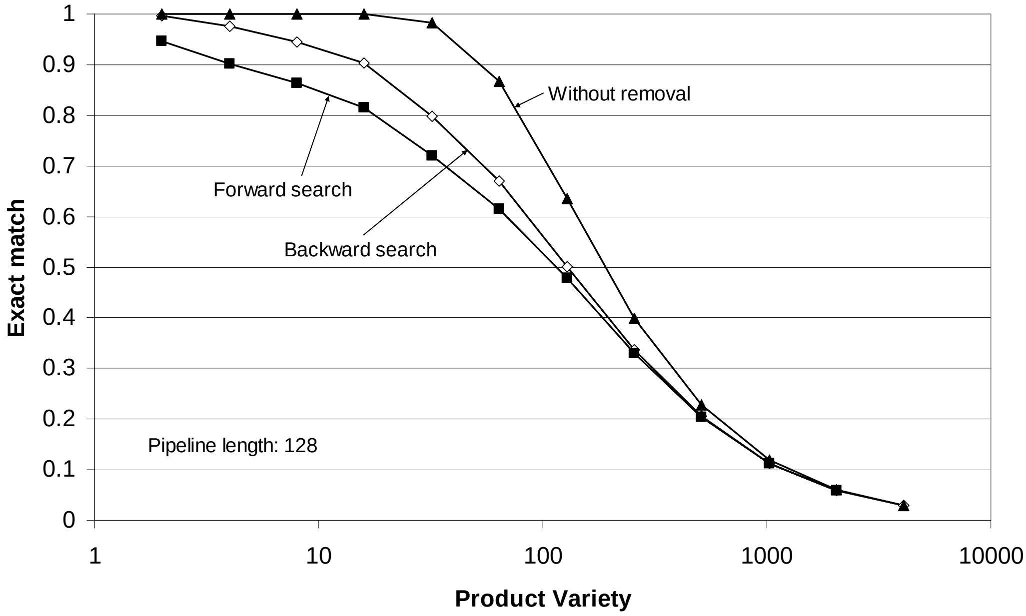
The GPSS simulation environment (Gordon 1975) was selected due its flexibility and to it having built-in functions (or 'Blocks' as they are called in GPSS) that greatly assist the construction of a VBTO model.

3. 'Pipeline only' system

The first stage of the analysis studies the pipeline in isolation. Products enter the pipeline and if they are unsold by the time they reach the end they are removed from the system and cannot be accessed by customers. If there is no matching product in the pipeline the customer is lost (i.e. no BTO request is made). The distribution of variety demanded by customers is equal to the distribution of produced variety and both are modelled as independent uniform random distributions, hence all variants have equal likelihood of being fed into the pipeline and of being ordered by a customer. The aim is to reveal the effect of search direction on the likelihood of finding a match in the pipeline for a customer.

The results of a pipeline of length 128 for a range of variety levels are shown in Figure 2 with logarithmic variety scale to illustrate the fulfilment behaviour of the system. For medium to low variety levels it shows that a lower proportion of customers are fulfilled when a forwards search is used compared to a backwards search. The difference disappears at high variety levels. The pattern of fulfilment is the same for both search methods. At low variety there is a high likelihood of a match in the pipeline, and as variety rises the likelihood of a match drops. A change in fulfilment likelihood from ~90% to ~10% is observed when the variety level increases by 2 orders of magnitude, approximately.

Figure 2: Probability of an exact match for each search method (Pipeline 128)

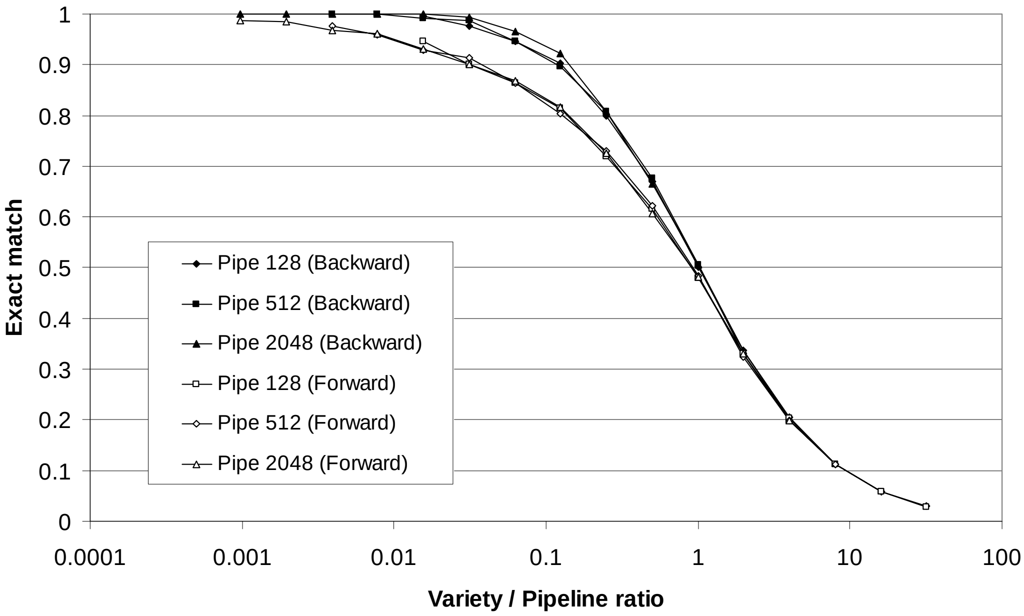


The difference between the two search methods can be explained by considering from where in the pipeline products are being allocated to customers. If two products in the pipeline match the customer's order, a backwards search policy will allocate the product that is furthest downstream, as opposed to the product that is furthest upstream that is allocated by the forwards search policy. The allocated product remains in the pipeline until it reaches the end and, until it leaves, it depletes the number of products available to subsequent customers. The residence time of products allocated through a backwards search is less than in a forwards search. Hence, compared to the forwards search, the backwards search leads to a lower reduction in the available number of products. At a low or medium variety level there is a significant likelihood of two or more suitable products being in the pipeline, but as shown in Figure 2, the difference between the forwards and backwards search policies diminishes as variety increases. This is as expected since the likelihood of there being two or more matching products in the pipeline also declines.

Figure 2 includes a third curve labelled 'without removal' that was generated by searching for a matching product in the pipeline and then leaving it available to subsequent customers, which is equivalent to a 'with replacement' selection process and means that every customer has a full pipeline available to them. Given the nature of the variety distributions for the feed into the pipeline and for the sequence of customer orders (with both being random sequences from independent uniform distributions) a binomial model should give the same results as the 'without removal' case, and this is found to be true. This result validates the basic simulation logic used.

An important and fundamental finding is that a consistent pattern is found when other pipeline lengths are simulated and the fulfilment likelihood plotted against the ratio of *variety / pipeline length*. In Figure 3 pipelines of 3 lengths have been simulated at 12 variety levels, from 2 to 4096, using both forward and backward searches.

Figure 3: Combined plot of backward and forward fulfilment vs. Variety / Pipeline ratio



An inference from this finding is that the disparity between the binomial model and the two search policies will be reproducible, and allow the performance of the pipeline-only system to be gauged roughly by reference to a simple Binomial calculation.

4. Comparison of the VBTO and Conventional systems

Here the VBTO system with three fulfilment mechanisms is compared to the Conventional two mechanism system. In a Conventional system a customer can be fulfilled only by a product from stock or by requesting a build-to-order (BTO) product since the unallocated products in the pipeline are hidden from the customer. The points of interest are how the addition of pipeline fulfilment alters the fulfilment patterns and how stock holding and customer waiting times are impacted.

In the simulations stock is allowed to form and no restriction is placed on the number of products in stock. However, because customer arrival rate and production rate are equal and constant, there is a constant number of unallocated products available in the system (which is equal to the pipeline length since in this simulation study the pipeline is always filled before the first customer arrives). To be clear, on each occasion that a customer is fulfilled from stock or from the pipeline they deplete the available products by one, but an unallocated product then enters the pipe. When a customer is fulfilled by a BTO product, the entry of the BTO product means that an unallocated product does not enter the pipe.

4.1. Observations on fulfilment mechanisms

For both systems, Conventional and VBTO, the ratio of variety to pipeline length has been found to be the fundamental determinant of behaviour when the feed into the pipeline and customer demand are modelled as sequences from independent uniform random distributions. Figures 4 and 5 are cumulative plots, e.g. in Figure 4 for the Conventional system, the area below the line is stock fulfilment, and above the line it is BTO fulfilment.

Figure 4: Fulfilment by mechanism for the Conventional system, *backwards* search. (Cumulative plot showing stock at bottom and BTO at the top)

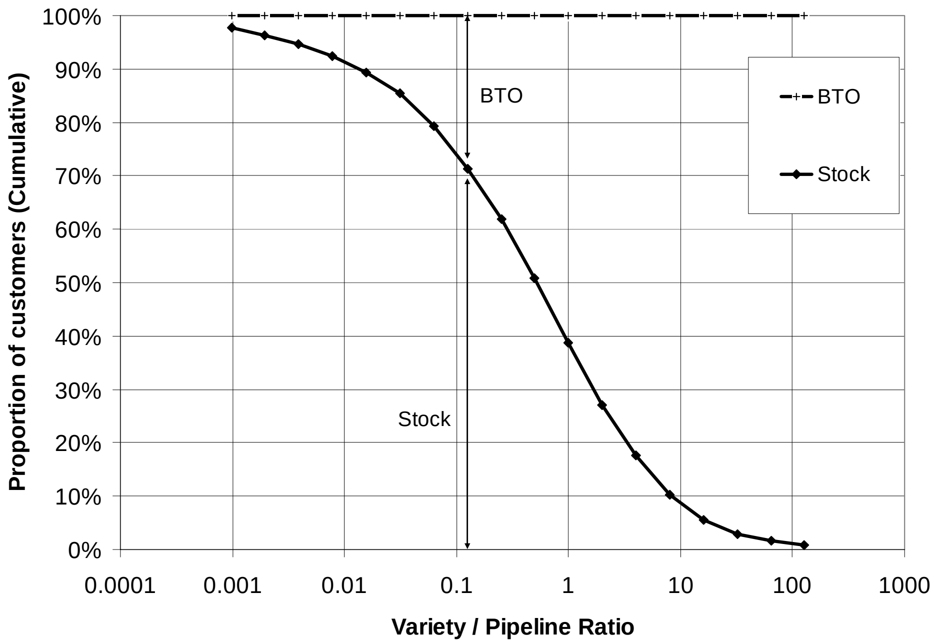
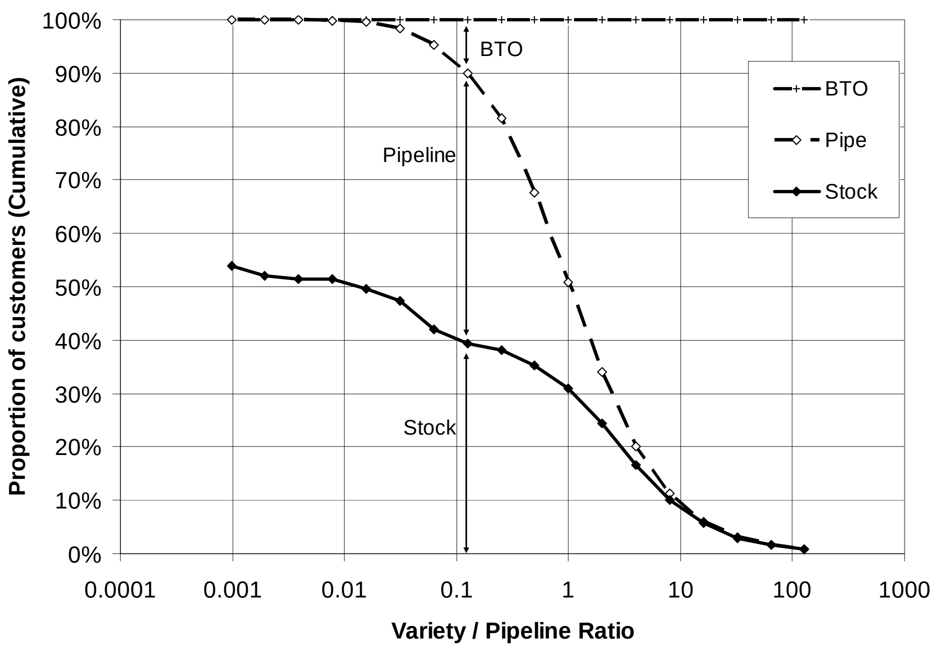


Figure 5: Fulfilment by mechanism for the VBTO system, *backwards* search. (Cumulative plot showing stock at bottom, pipeline in the middle, and BTO at the top)

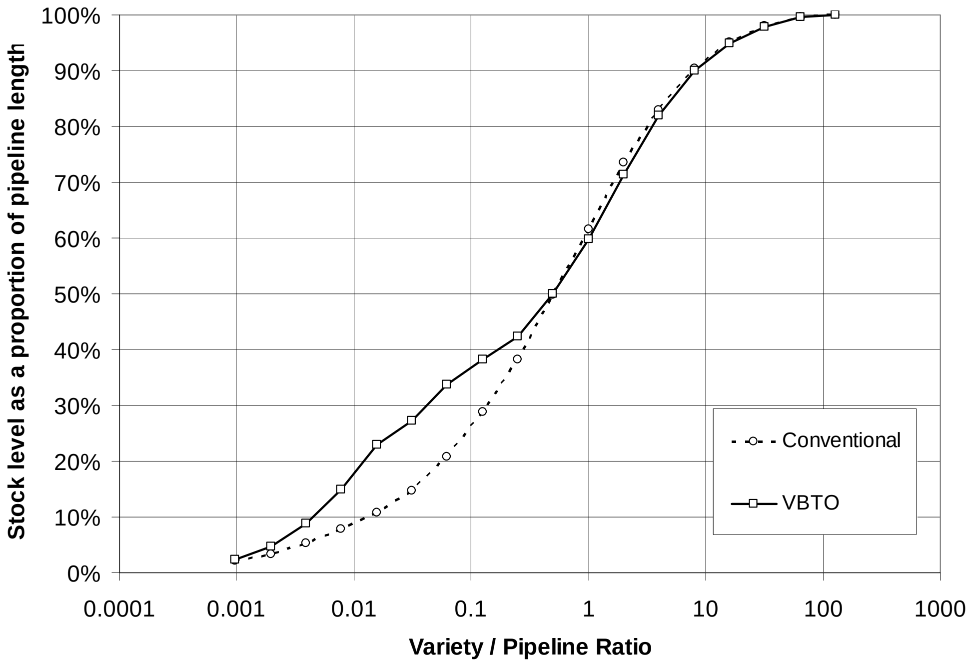


The results in Figure 4 show that at a variety/pipeline ratio of 1, just under 40% of customers are fulfilled from stock and 60% by BTO. In Figure 5 for the VBTO system, the bottom area is stock fulfilment, the middle zone is pipe fulfilment and the upper area is BTO fulfilment. In this graph, at a variety/pipeline ratio of 1 the fulfilment is 30% from stock, 20% from the pipeline and 50% by BTO. Compared to the Conventional system, in the VBTO system the switchover to BTO fulfilment is delayed to higher variety/pipeline ratios. When the variety/pipeline ratio is in the region of 0.01 to 0.1, the BTO proportion in a VBTO system is at the level it would be in a Conventional system which had an order of magnitude less variety. At a variety/pipeline ratio of ~10 and above, the VBTO and Conventional systems are equivalent.

4.2. Comparison of stock levels

Comparison of Figures 4 and 5 indicates that at low variety/pipeline ratios the dominant fulfilment mechanism is from stock in the Conventional system. Approximately double the proportion of customers is fulfilled from stock in the Conventional system compared to the VBTO system. From this observation it would be expected that there is double the amount of stock in the Conventional system. However, the reverse is the case as shown in Figure 6.

Figure 6: Comparison of average stock levels in the VBTO and Conventional systems



At variety/pipeline ratios below 1 the VBTO system carries more stock than the Conventional system, approaching twice as much at ratios close to 0.01.

Examination of the products in stock in the VBTO system reveals the mix to be highly skewed compared to the Conventional system. A random snapshot of the stock in a VBTO system (using a backward search, pipeline length of 512) with variety of 4 illustrates the condition: of the 57 products in stock there were none of type '1', 8 of type '2', 5 of type '3' and 44 of type '4'. The oldest 33 products in stock were all type '4'. The effect of this is that the stock in the VBTO system has less coverage than the stock in the Conventional system, where coverage is the proportion of product variants for which there is at least one available. Coverage for the two systems is presented in Figures 7 and 8 and comparison of these plots with Figures 4 and 5 reveals that fulfilment is closely linked to coverage.

Figure 7: Coverage in the Conventional system

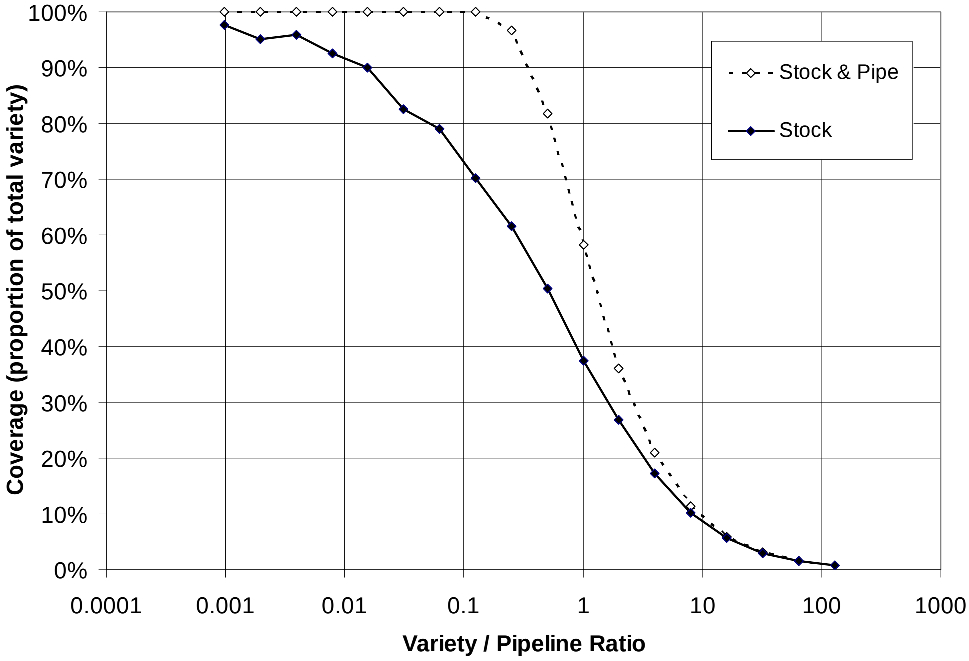
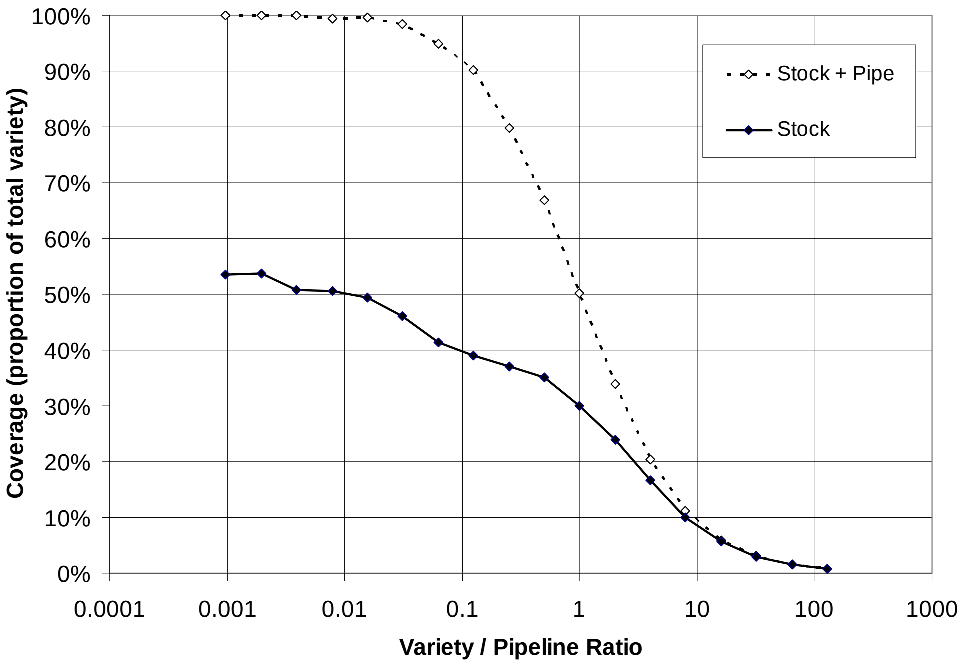


Figure 8: Coverage in the VBTO system

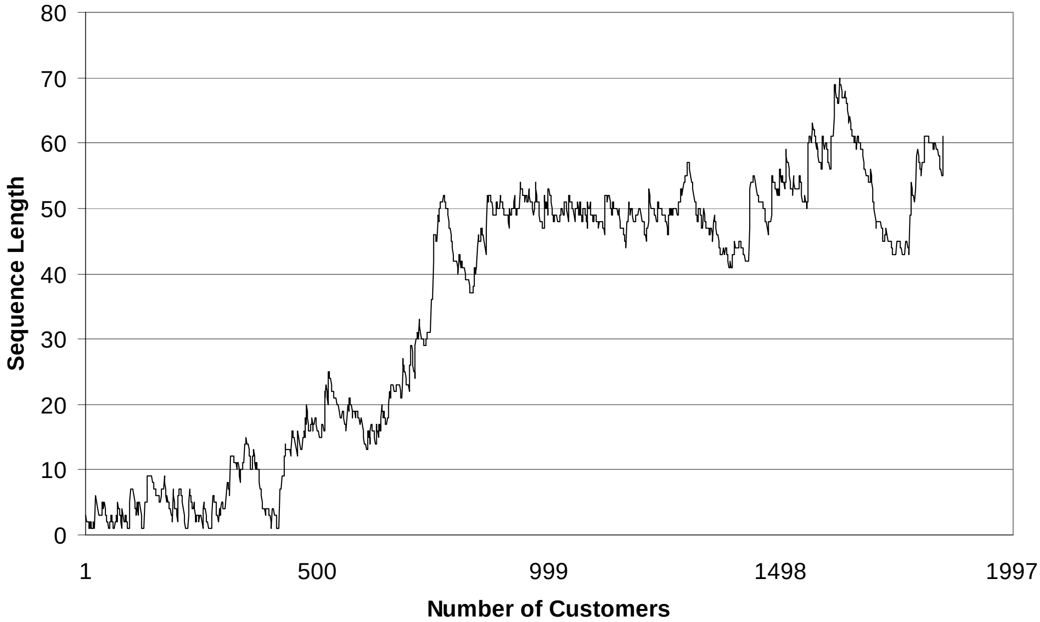


The cause of the stock mix in the VBTO system being different to a Conventional system has been found to lie in the interaction between the sequence of products in the pipeline and the sequence of customers. The feed of products into the pipeline is a random sequence and the customers also arrive in a random sequence, but the interaction between the two leads to the pipeline being stripped of its randomness, resulting in a non- random distribution of variants replenishing stock. In the Conventional system the stock is also replenished by products from the pipeline, but which in this case has been hidden from customers and is still in the random sequence that created it.

To illustrate this 'stripping' phenomenon, a long pipeline has been generated of a random sequence of 4 products, which has then been searched (using a backwards search), by a random sequence of customers. Figure

9 shows that after approximately 500 or so customers, near enough the first 20 available products are of the same type. After 1000 customers, the first 50 or so available products in stock are identical, and at one point the first 70 products are identical.

Figure 9: Length of sequence of identical products at the front of the pipeline



This outcome illustrates the stripping behaviour and shows that the sequence of available products in the pipeline of the VBTO system is no longer random. Further investigation has indicated that a statistically significant change in the mix occurs quickly. Using a chi-square test for randomness, when an infinite pipeline holds a random sequence of 32 variants, the downstream section of the pipeline (from where customers have been fulfilled) is no longer random after approximately 100 customers, and when the pipeline holds a variety of 256 variants randomness is lost after approximately 150 customers.

4.3. Comparison of customer waiting times

Although the VBTO system has a lower proportion of customers fulfilled by BTO products than in a conventional system until higher variety / pipeline ratios, at lower ratios the average customer waiting time is greater than in the Conventional system, as shown in Figure 10.

This comparison of average waiting time does not reflect that in the Conventional system there are two waiting times only since customers are either serviced immediately from stock or wait for a BTO product to come through the entire pipeline, whereas in the VBTO system there is a spread of fulfilment times. This is illustrated in Figure 11 which presents the distribution of waiting times in a VBTO system when the variety/pipeline ratio is 0.004.

Although the average customer waiting time in the VBTO system is 189 compared to 114 in the Conventional system, half of all VBTO customers are fulfilled in 117 time periods or less. In the Conventional system 22% of customers wait the length of the pipeline (which is 512 time periods) whereas only 7% of customers wait for longer than 500 time periods in the VBTO system.

Figure 10: Comparison of average customer waiting times in VBTO systems compared to Conventional systems

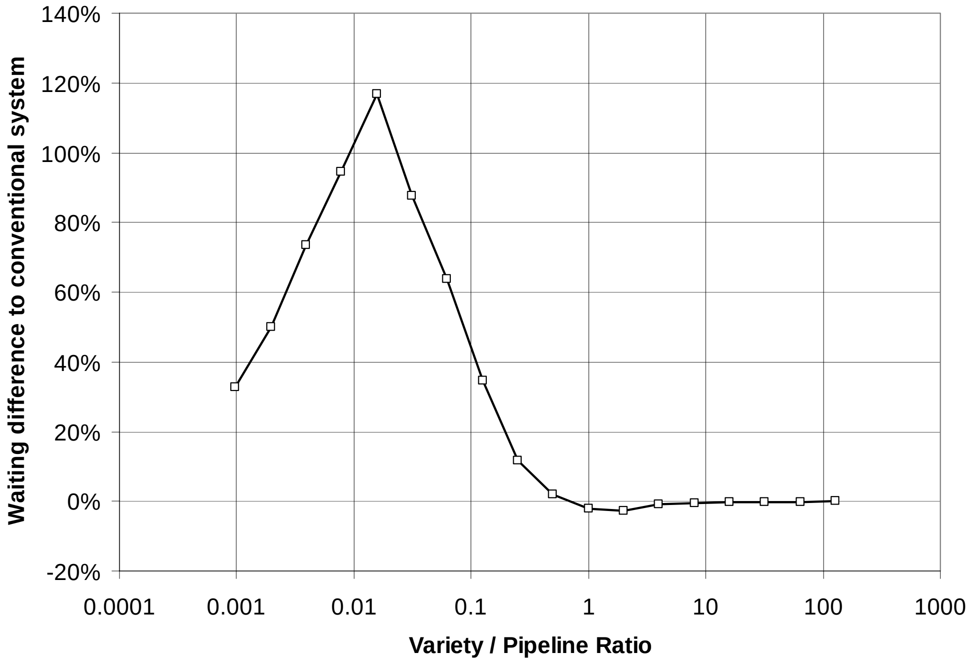
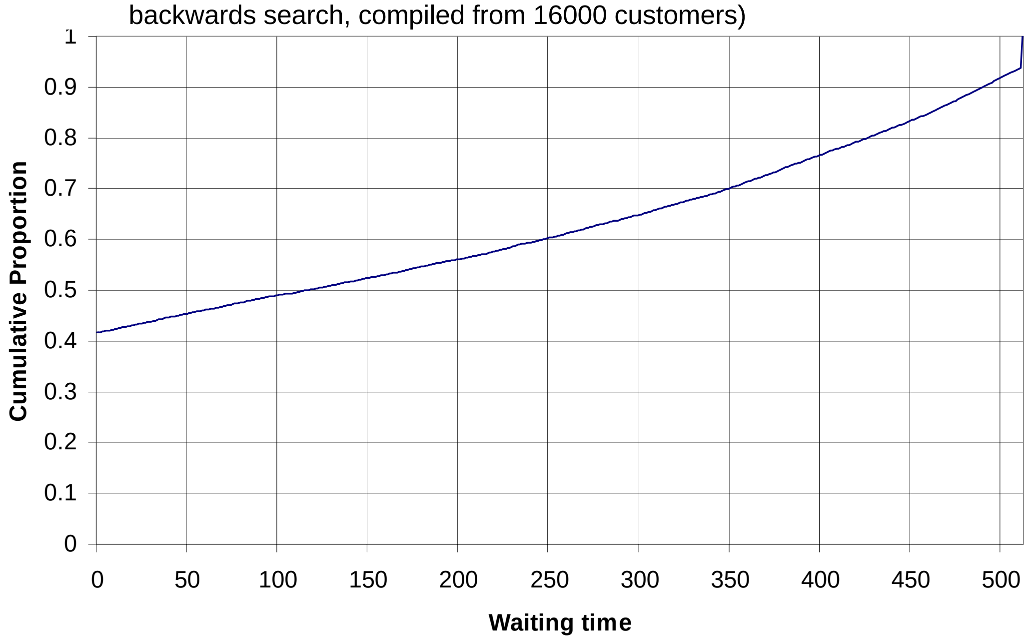


Figure 11: Cumulative distribution of customer waiting times for the VBTO system (variety 32, pipeline 512, backwards search, compiled from 16000 customers)



5. Discussion

This is the first study in the literature to examine the intrinsic behaviour of the Virtual-Build-to-Order system and to establish the parameters which determine its performance. The research approach has been to study the system in a basic form, with simple operating rules and without complex control logic, and this approach has been vindicated by the insights obtained. Although the simulation model has been kept simple, it structurally and logically captures key elements of real pipeline systems, enabling macro-behaviour to be investigated. An understanding of the response of the basic system to changes in the primary design parameters is necessary to provide a foundation for developing control policies and for providing direction for future research.

In the basic operating conditions studied the variety/pipeline length ratio has been found to be the principal determinant of the performance of both the VBTO system, which has an open pipeline, and the Conventional system, which has a closed pipeline. Using the charts above a system designer can gauge the balance between the fulfilment mechanisms in either system.

In comparison to the Conventional system, making the pipeline visible to customers changes the behaviour and fulfilment performance significantly. It is at variety/pipeline length ratios of between 0.001 and 1 that clear differences between the systems are observed. In this ratio range stock levels and average customer waiting time are greater when the pipeline is accessible. The differences are due to the stock mix being poorer in the VBTO system than in the Conventional system, and the cause for the poorer stock mix has been found to be the process of searching the visible pipeline. A major finding from this study is that the open pipeline, which is the key characteristic of the VBTO system, is also the characteristic that impairs its performance on two key metrics.

Although the average customer waiting time is longer in the VBTO system than in the Conventional system, there is a positive way of interpreting waiting time. As illustrated in section 4.3, fewer customers need to wait the length of time for a BTO product. In this study customers are modelled as being willing to wait, but in reality some customers will refuse to wait beyond a certain length of time and will be lost. Hence, even though the average waiting time is greater, it is possible that fewer customers will be lost to a producer operating the VBTO system. This is an issue for further study.

Further investigation is required to determine whether operational policies and practices can mitigate and manage the inherent behaviour. Areas to explore include introducing feedback between the sequence fed into the pipeline and the current stock mix, and to the use of rules for allocating pipeline products to customers. In this study there has been no feedback and the search rule has been to allocate the first available product. It is conceivable that a producer should restrict access to the pipeline, perhaps by making a fraction of products invisible. A simple method of doing so would be to make, say, one in ten of the products fed into the pipeline invisible, but more sophisticated approaches could be used, such as keeping a fraction invisible and a real-time monitoring process switching products from being visible to invisible and vice-versa depending on their specification and on the mix of available products in the pipeline and stock.

6. Conclusion

The paper has reported findings from a simulation study on an emerging order fulfilment model – VBTO- of particular interest in the automotive sector but which is also of interest to other sectors with similar characteristics i.e. with high levels of variety and necessary extensive planned product pipelines. The paper has highlighted the behaviour and performance of the VBTO model in comparison to the Conventional model without pipeline fulfilment. Insights have been presented on the mode of fulfilment, the behaviour characteristics of stock, the variety coverage provided by the system and the expected waiting time. The central role played by the variety to pipeline ratio is highlighted and the relevance of the findings for the design, management and control of open pipeline systems have been discussed. The work offers many avenues for development including pipeline control policies and methods.

References

Agrawal, M., T. Kumaresh and G. Mercer, 2001, The false promise of mass customization, The McKinsey Quarterly 3, 62-71.

Barbiroli, G. and A. Focacci, 2003, A techno-economic analysis of the results of product diversification in household appliance durables--evaluating concreteness, International Journal of Production Economics 83, 257-278.

Bils, M., and Kennow, P. J., 2001, The Acceleration in Product Variety, The American Economic Review, 91(2), 274-280.

Brabazon, P.G. and B.L. MacCarthy, 2004, Virtual-build-to-order as a Mass Customization order fulfilment model, Concurrent Engineering: Research and Applications 12, 155-165.

Cox, W.M. and R. Alm, 1998, The right stuff: America's move to mass customization (Federal Reserve Bank of Dallas).

Denton, B., D. Gupta and K. Jawahir, 2003, Managing Increasing Product Variety at Integrated Steel Mills, Interfaces 33, 41-53.

Elias, S., 2002, 3 Day Car Programme Research Survey Report: New car buyer behaviour (Cardiff University, Cardiff, UK).

Gordon, G., 1975, The application of GPSS V to discrete system simulation (Prentice Hall, Englewood, NJ, USA).

Graman, G.A. and M.J. Magazine, 2002, A numerical analysis of capacitated postponement, Production and Operations Management 11, 340-357.

Holweg, M. and F.K. Pil, 2004, The second century: reconnecting customer and value chain through Build-to-Order (MIT Press, Cambridge, Mass).
Law, A.M. and W.D. Kelton, 2000, Simulation modeling and analysis (McGraw-Hill, Singapore).
McCutcheon, D.M., A.S. Raturi and J.R. Meredith, 1994, The customization-responsiveness squeeze, Sloan Management Review 35, 89-99.
Partanen, J. and H. Haapasalo, 2004, Fast production for order fulfillment: Implementing mass customization in electronics industry, International Journal of Production Economics 90, 213-222.
Salvador, F. and C. Forza, 2004, Configuring products to address the customization-responsiveness squeeze: A survey of management issues and opportunities, International Journal of Production Economics 91, 273-291.
Swaminathan, J.M. and S.R. Tayur, 1999, Managing design of assembly sequences for product lines that delay product differentiation, IIE Transactions 31, 1015-1026.
Swaminathan, J.M. and S.R. Tayur, 1998, Managing Broader Product Lines through Delayed Differentiation Using Vanilla Boxes, Management Science 44, S161-S172.
Turner, K. and G. Williams, 2005, Modelling complexity in the automotive industry supply chain, Journal of Manufacturing Technology Management 16, 447-458.
Yang, B., N.D. Burns and C.J. Backhouse, 2004, Management of uncertainty through postponement, International Journal of Production Research 42, 1049-1064.
van Hoek, R.I., 2001, The rediscovery of postponement: a literature review and directions for research, Journal of operations management 19, 161-184.