Inventory optimization for Industrial-Scale assemble-to-order Systems

Willem van Jaarsveld
University Rotterdam - Erasmus School of Economics
&
Alan Scheller-Wolf
Carnegie Mellon University - Tepper School of Business

YEQT 2014, Eindhoven

Outline

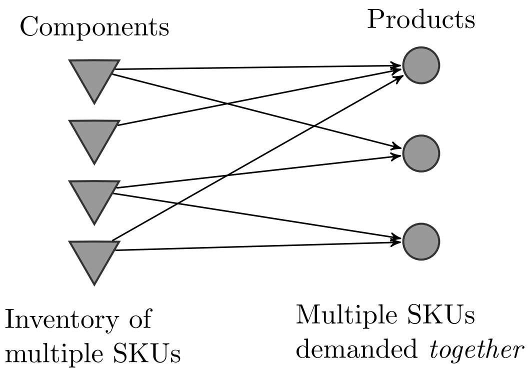
- Introduction & Motivation
- Model & methods
- Results
- Conclusions

Assemble-to-order (ATO) systems

- Large product variety managed by assembling the *products*, on demand, from multiple *components*
  - Broad assortment of products, yet cost efficient.
  - Good service: assembly is fast.
  - E.g. Dell (Kapuscinski et al. , 2004), IBM (Swaminathan and Tayur 1998; Cheng et al. 2002), . . .
- Additional motivation: *any situation* where SKUs of different types are needed together to attain performance:
  - Multiple spare parts and/or tools in a maintenance task, e.g. Fokker (v J, Dollevoet, Dekker, 2011), Philips Healthcare (Kampstra, 2012), ASML (Vliegen; 2009). . .
  - Multiple products in single customer demand, e.g. online retailers such as Amazon (Xu et al., 2009), . . .

ATO systems

ATO production:



M

Controlling ATO systems

- Proper control is important to attain the cost efficiency promised by the ATO production system.
- Two related control problems:
  1. Controlling replenishment/production of each component.
  2. Controlling allocation of components to product demands.
- Companies need to control production and allocation of hundreds/thousands of components and products.

Literature on controlling ATO systems

Incomplete review

1. Optimization of component **base-stock levels** (/(s, S) policies) under **FCFS allocation** (with committing):
   - Using approximate fillrate/waiting time: Song (1998), Zhang(1997), Song and Yao (2002), Cheng et al.(2002), Kapuscinski et al. (2004); Lu and Song (2005), v J et al. (2011)
     - Approximate optimization.
   - SP formulation: Swaminathan and Tayur (1998), Akçay and Xu (2004), Huang and de Kok (2011)
     - For positive leadtimes, SPs give rise to *non-convex* optimization problem.
     - Solved using "big-M" method. Thus limited to small problems.

Literature on controlling ATO systems

Incomplete review

2 Optimality results on non-FCFS and coordinated replenishment, for **special cases**:

- No-holdback allocation for base-stock inventory control (special BOM, e.g. generalisations of W-model): Song and Zhao (2009), Lu et al. (2010)
- Markovian models (special BOM): Benjafaar and ElHafsi (2006), ElHafsi et al. (2008), Nadar et al. (2012)
- SP based analysis (W/N-model): Doğru et al. (2010), Lu et al. (2012), Reiman and Wang (2012);
- High volume systems: Plambeck and Ward, (2007,2008), Reiman and Wang (2013)

Controlling ATO systems: companies

Companies use pragmatic approaches:

- Independent base-stock/(s, S) policies. Heuristics to determine parameters:
  - Ignore simultaneous stock-outs of multiple components (ISS):

$$\text{Product } i \text{ wait} = \max_{j \in \mathcal{J}^i}(\text{Wait for component } j) \approx \sum_{j \in \mathcal{J}^i}(\text{Wait for component } j)$$

    e.g. IBM (Cheng et al.;2002), Dell (Kapuscinski et al. ;2004), Fokker (vJ Dollevoet Dekker; 2011)

- Simple allocation rules: FCFS (commit inventory), FRFS (no-holdback)
- Because ISS and FCFS/FRFS are easy to implement.
- **But what is the performance of these approaches?**

Research questions

1. How to find provably (near-) optimal base-stock levels for large-scale ATO systems under FCFS allocation?
   - What can be achieved with FCFS and independent base-stock (no coordination)?
2. What are the costs of ignoring simultaneous stock-outs (ISS) while optimizing inventory?
3. What are the costs of using FCFS instead of optimal allocation in industrial-scale ATO systems?
   - How much further improvement if moving away from FCFS?

The model and notation

Ease of exposition

- **Continuous review**, multiple products ($\mathcal{I}$) and components ($\mathcal{J}$).
- Independent base-stock policies s_j for components $j \in \mathcal{J}$. Vector of base-stock policies $\vec{s}$.
- Costs for on-hand component inventory and product demand back-orders :

$$C = \sum_{j \in \mathcal{J}} h_j \mathbf{E} H_j(\vec{s}) + \sum_{i \in \mathcal{I}} b^i \mathbf{E} B^i(\vec{s})$$

- Poisson demand for products (λ^i)
- Deterministic replenishment leadtimes for components (l_j).
- Deterministic $0-1$ BOM:
  - $\mathcal{J}^i$ components used in product i.
  - $\mathcal{I}_j$ products in which component j is used.

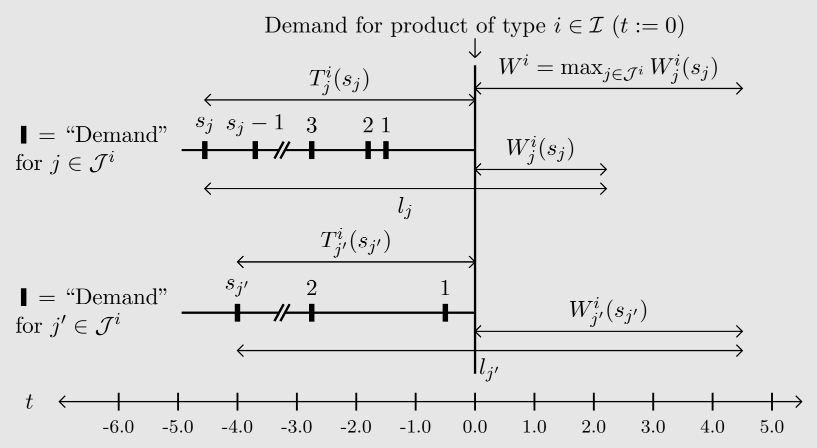
Near optimal base-stock levels under FCFS

- Lu and Song (2005) show:

$$C = \sum_{j \in \mathcal{J}} h_j s_j + \sum_{i \in \mathcal{I}} \tilde{b}^i \mathbf{E} B^i(\vec{s}) - D$$

  where $\tilde{b}^i = b^i + \sum_{j \in \mathcal{J}^i} h_j$, and D is constant.
- Use exact computation of $\mathbf{E}B^i(\vec{s})$ to optimize?
  - No: computation only tractable for $|J^i| \leq 3$.
  - No: non-linear, non-separable optimization; dimension of search space.
- Instead, we develop an SP formulation, and sampling procedures.
- $\mathbf{E}B^i(\vec{s}) = \lambda^i \mathbf{E}W^i(\vec{s})$, where W^i is the random waiting time incurred by an arbitrary demand of product i.
- Use *tagged job*, consider arbitrary arrival of demand for product i, and determine W^i.

Tagged job approach



SP formulation under FCFS

- x_{js} *indicates* that base-stock level s is used for part j.
- Two-stage stochastic program:

$$\min \sum_{j\in\mathcal{J}} \sum_{s=0}^{s_j^u} h_j s x_{js} - D + \sum_{i\in\mathcal{I}} \tilde{b}^i \lambda^i w^i$$

$$w^i = \mathbf{E} \max_{j\in\mathcal{J}^i} \left(\sum_{s=0}^{s_j^u} x_{js} W_j^i(s) \right) \qquad i \in \mathcal{I}$$

$$\sum_{s\in C_j} x_{js} = 1 \qquad j \in \mathcal{J}$$

$$x_{js} \in \{0,1\} \qquad j \in \mathcal{J}, s \in \{0,\ldots,s_j^u\}$$

- This SP can be *efficiently* solved using an SAA approach.
- To find feasible solutions and asymptotic lower bounds.

SAA approach

- Sample consists of N^i independent realizations of arrivals of each product i: $\xi^i_n = \{W^i_j(s)(\xi^i_n) | j \in \mathcal{J}^i, s \in \{0, \ldots, s^u_j\}\}$.
- SAA of SP for sample is given by:

$$\min \sum_{j \in \mathcal{J}} \sum_{s=0}^{s^u_j} h_j s x_{js} - D + \sum_{i \in \mathcal{I}} \frac{1}{|N^i|} \tilde{b}^i \lambda^i \sum_{n \in N^i} v^i_n,$$

$$v^i_n \geq \sum_{s=0}^{s^u_j} x_{js} W^i_j(s)(\xi^i_n), \qquad i \in \mathcal{I}, j \in \mathcal{J}^i, n \in N^i,$$

$$\sum_{s=0}^{s^u_j} x_{js} = 1, \qquad j \in \mathcal{J},$$

$$v^i_n \in \mathscr{R}, \qquad i \in \mathcal{I}, n \in N^i,$$

$$x_{js} \in \{0, 1\}, \qquad j \in \mathcal{J}, s \in \{0, \ldots, s^u_j\}.$$

Scalability

Size $\lvert\mathcal{J}\rvert = \lvert\mathcal{I}\rvert$ of test case	Computation time (Hrs.)		
	Tagged-Job approach	Akçay and Xu (2004) [1]	Huang and De Kok (2011)
5	0.01	0.11	1.2
10	0.03	0.11	4.5
15	0.04	1.5	40.1
20	0.07	88.8	-
50	0.17	-	-
100	0.4	-	-
200	0.88	-	-
300	1.45	-	-

Our tagged-job approach is the first scalable algorithm for determining base-stock levels in FCFS ATO systems.

[1] Ignores committed inventories.

Lower bound under optimal allocation

For better insights into non-FCFS policies: lower bound on the costs of the best base-stock policy under optimal allocation.

- Idea 1: under *any* allocation rule, on hand inventory at time $t := 0$ must be bigger than 0. Minimize back-orders at 0, subject to only this constraint:
  - Doğru et al, 2010: two-stage SP for equal leadtime case, LB on optimal policy.
  - Reiman and Wang, 2012: n-stage SP for unequal leadtime case, LB on optimal policy.
  - Employed to find the optimal replenishment + allocation policy for W-model under cost symmetry.

Lower bound under optimal allocation

- We: under base-stock policy, on-hand inventory at 0 equals

  $s_j + (\text{back-orders at } -l_j) - (\text{satisfied demands during } (-l_j, 0])$

- We use this relation to develop an SP lower bound on the costs of the optimal base-stock policy under optimal allocation.
- Two-stage SP, also for the non-equal leadtime case.
- We propose SAA algorithm to solve for general ATO systems.

Case: Assembly of products of multiple families

- Many OEMs divide their products into product families (e.g. medical equipment or wafer steppers, see De Kok (2003)).
- Developed cases with this property:
  - 3 product families, 20 common components, and 3×8 (more expensive) family specific components, .
  - Products in each family are assembled from 5 common components and $n_s \in \{1, \ldots, 8\}$ family specific components.
- Choose demand rates, Holding costs and leadtimes randomly. Leadtimes uniformly on $[1 - \delta_l, 1 + \delta_l]$.
- n_s and δ_l are varied.

Case: Assembly of products of multiple families (2)

- Two additional parameters - news-vendor fractile f and assymmetry δ_b - for setting back-order penalties:

$$b^i = \frac{f}{1-f} h^i (1 + \delta_b x^i)$$

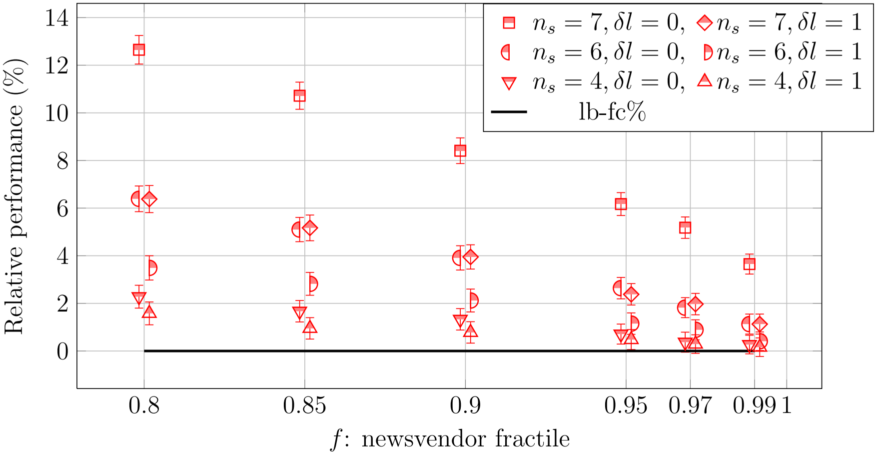
- Here, $h^i = \sum_{j \in \mathcal{J}^i} h_j$.
- And $x^i \in \{-0.5, 0.0, 0.5\}$ governs whether the product is a low, medium or high criticality.
- While setting x^i, we use two cases, representing *within* product family and *between* product family asymmetry.

Tested control policies & presentation

- Three policies for setting base-stock levels: ignoring simultaneous stockouts "iss", our tagged-job approach: "spfc"; the base-stock levels from the lower bound on optimal allocation "drw".
- Three allocation rules:
  - FCFS: commits inventory, denoted "-fc"
  - FRFS: No holdback/committing, clears back-orders FCFS "-fr"
  - No holdback, clear back-orders with highest modified penalty costs $\tilde{b}^i$ first: "-pr".
- Two lower bound estimates: a lower bound under FCFS allocation "lb-fc", and a lower bound for base-stock policies under optimal allocation "lb-all".
- We report relative differences with the lower bound under FCFS, for example:

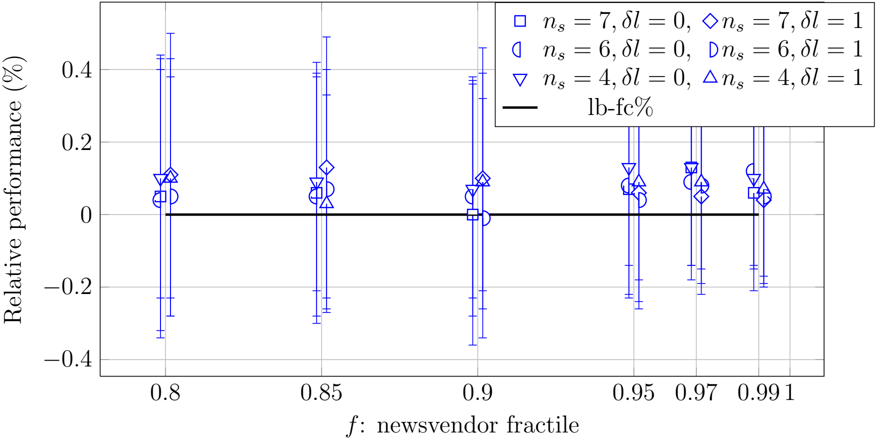
$$\text{iss-fc}\% := \frac{\text{iss-fc} - \text{lb-fc}}{\text{lb-fc}} \times 100\%.$$

Multiple family assembly : ISS-FC policies



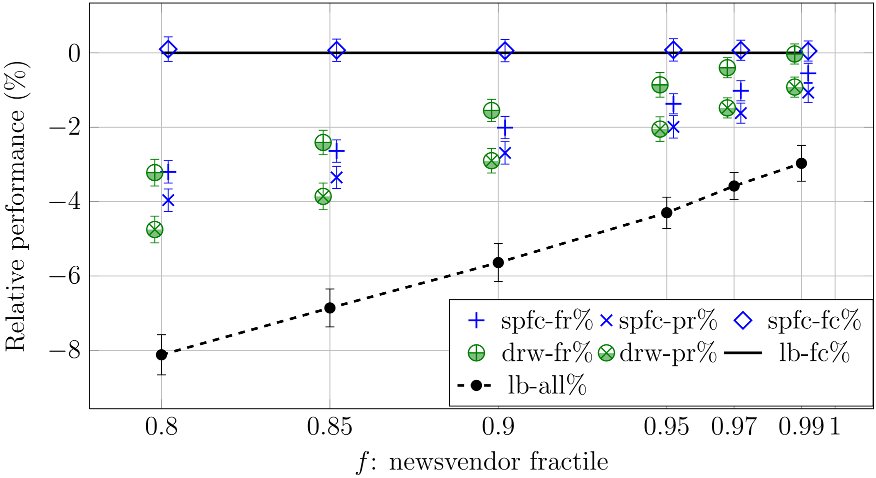
Bars represent 99.7% confidence intervals

Multiple family assembly: Tagged-job approach



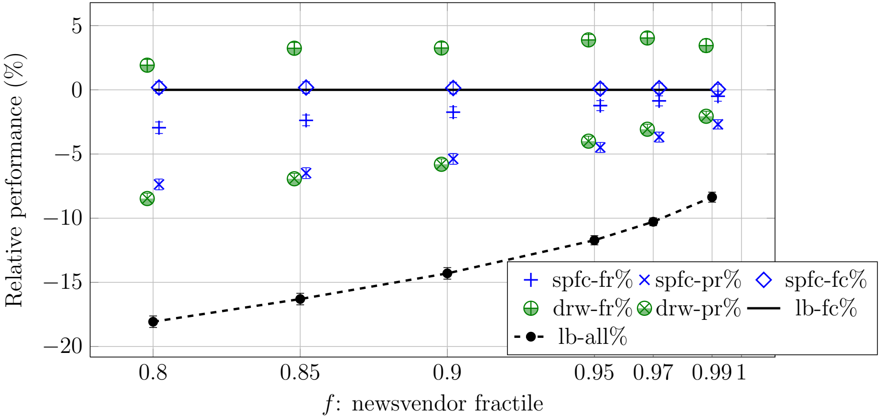
Multiple family assembly: non-FCFS

Between product family assymmetry ($n_s = 6, \delta_l = 1, \delta_b = 1$):



Multiple family assembly: non-FCFS

Within product family assymmetry ($n_s = 6, \delta_l = 1, \delta_b = 1$):



Repair shop case

- Based on data from aircraft component repair shop (Fokker Aerospace).
- 3 repair types ($\mathcal{I} = \{a, b, c\}$):
  - arrival rates $0.13, 0.10$ and 0.35.
- 110 spare parts:
  - holding cost (h) and leadtime (l) given.
  - Each part used in each repair type with given *probability* (p_a, p_b, p_c).
- Define holding costs:
- Define the backorder costs for repairs of type $i \in \{a, b, c\}$ as

$$b^i = h^i(f/(1-f))(1+x^i)$$

with

  - $x^a = 0.5\delta b$, $x^b = 0$, $x^c = -0.5\delta b$
  - $h^i := \sum_{j=1}^{110} p_j^i h_j$.

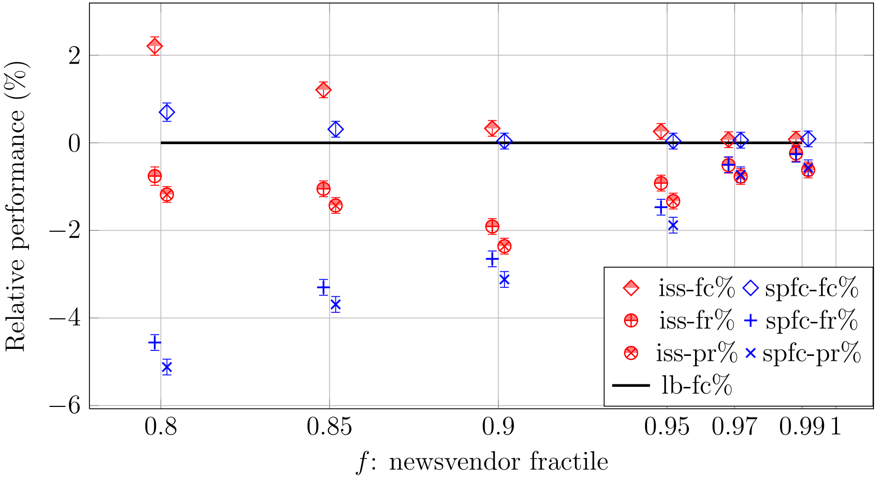
Repair shop case (continued)

Repair shop data (excerpt):

h	l	p_a	p_b	p_c	h	l	p_a	p_b	p_c	h
341	55	2.3%	-	-	56	21	4.7%	5.7%	1.8%	12
270	55	2.3%	2.9%	-	52	7	7%	8.6%	6.4%	11
270	55	-	-	8.3%	50	19	-	-	9.2%	11
249	41	4.7%	2.9%	4.6%	50	21	2.3%	2.9%	0.9%	10
240	41	2.3%	5.7%	2.8%	50	6	-	-	33%	10
213	55	16.3%	5.7%	-	50	13	-	-	3.7%	9
175	34	2.3%	2.9%	0.9%	50	17	2.3%	-	-	8
162	55	-	5.7%	-	50	19	18.6%	31.4%	20.2%	8
156	45	7%	-	4.6%	46	21	-	2.9%	-	7
156	31	-	-	0.9%	46	17	4.7%	-	4.6%	6
128	28	-	-	0.9%	44	12	-	-	0.9%	6
123	16	2.3%	2.9%	1.8%	43	12	4.7%	-	0.9%	6
123	13	4.7%	-	2.8%	41	50	2.3%	8.6%	-	5
105	37	-	-	1.8%	41	23	2.3%	8.6%	-	5
105	37	-	-	14.7%	41	3	4.7%	2.9%	5.5%	4
105	17	-	-	0.9%	39	13	-	-	0.9%	4

Repair shop: various policies

$\delta b = 1, \lambda = 2.0$:



Insights

Policy	Benchmark	Params		Cost increase per NV fractile 0.8	0.95	0.99
ISS	*lb-fc*	Dem corr.	L:	1-2%	0-0.1%	0%
			M:	5-8%	2-4%	1-2%
			H:	12-30%	6-19%	4-12%
SPFC	*lb-fc*	-		Always $< 0.5\%$		
FCFS (*lb-fc*)	*lb-opt*	P. asym.	L:	4-8%	2-4%	1-3%
			H:	8-18%	4-12%	3-8%
FCFS (*lb-fc*)	NHB-PR (*spfc*/*drw*)	P. Asym.	L:	3-4%	1-2%	0-1%
			H:	5-8%	2-5%	1-2%

Opportunity

Conclusion

- We developed the first algorithm that can find close to optimal solutions for large ATO systems, without using the ISS assumption.
- We generated insights for the performance of control policies for realistic cases.
- We found that ISS performs well as long as demand correlation during leadtime is low, or if NV fractiles are high.
- But can also perform poorly if dem. corr. during leadtime is high and NV fractiles are low. In those cases, consider SPFC.
- FCFS performs surprisingly well, even if there is significant penalty asymmetry.
- Simple no-holdback policies slightly outperform the *best* FCFS policy.