Production Planning with Resources Subject to Congestion

Reha Uzsoy

Edward P. Fitts Department of Industrial and Systems Engineering

North Carolina State University

NC STATE UNIVERSITY

Planning

- Planning is the allocation of the firm's productive capacity to different products or customers.
- Purpose is to match supply to demand
  - Need to deal with cycle times or flow times
  - Time from releasing work to its emerging as finished product
- Primary control variable is timing and quantity of releases into the plant

Manufacturing Capacity

- All planning algorithms have embedded some model of manufacturing capacity
- People talk all the time about capacity as if it is a single number
- In fact, situation is much more nebulous
  - Elmaghraby (1991)
- Capacity is generally manifested in cycle times
  - If we had infinite capacity, cycle time would be no greater than raw processing time
  - Estimated in planning models as lead time

Fundamental Circularity

- Planning models try to match supply to demand
- Hence they need estimates of cycle times
  - Lead times
- But cycle times depend on utilization
- Utilization is an output of the planning process
  - Determined by the quantities of work released into the resources over time

Implications

- Cycle times begin to increase well before utilization reaches 1
- This increase in cycle times also drives work in process inventories (WIP)
- Queueing tells us both increase nonlinearly with utilization
  - Once the queues are in steady state, linked by Little's Law
  - If not in steady state, considerably less clear

Planning Circularity

- Lies at the heart of production planning since its inception
- Until recently, very few direct attempts to address this
- Lies at the boundary of mathematical programming and queuing
- Generally speaking, queues are likely to be in transient regime
  - Not clear how good an approximation obtained by steady-state models

Planning Circularity vs. Random Lead Times

- Not the same things!
- In practical systems, uncertain lead times are pervasive
  - Subject to some probability distribution
- Planning circularity implies that the probability distribution changes with planning decisions

Optimization models

- Planning horizon divided into discrete periods
- Decision variables for each period
- Flow conservation constraints across periods
  - Enforced at boundary between periods
  - Activity rates constant within periods
- Aggregate capacity constraints for key resources in each period
- Costs assessed on period variables
  - Separable by product and period
- Basic paradigm is unchanged since the 1950s
  - Modigliani and Hoh(1955); Holt et al.(1960)

Simple LP model

$$\text{MIN} \quad \sum_{t=1}^{T}\sum_{i=1}^{n} c_i X_{it} + h_i I_{it}$$

Subject to :

$$I_{it-1} + X_{it} = D_{it} + I_{it} \qquad i = 1..n,\ t = 1..T$$

$$\sum_{i=1}^{n} a_i X_{it} \le C_t \qquad t = 1..T$$

$$X_{it},\ I_{it} \ge 0 \qquad i = 1..n,\ t = 1..T$$

Treatment of lead times

- Now we need to distinguish between
  - Production (X_{it})
  - Releases (R_{it})
- Questions of timing emerge
  - When does the released material emerge as finished product?
  - When does the released material consume capacity?

Basic lead times

- Relationship between releases and output:
  - $X_t = R_{t-\Lambda}$
- Inventory balance
  - $I_t = I_{t-1} + X_t - D_t$
- Capacity
  - $X_t \leq C_t$
- Where is the WIP?

$$W_t = \sum_{\tau=1}^{t} R_\tau - \sum_{\tau=1}^{t} D_\tau$$

Observations

- WIP is implied in these formulations, but very seldom explicitly treated or costed
- If $R_{t-L} < C_t$, only R_{t-L} units of WIP are available to be worked on in period t
  - Replanning will fix this…
- Assumes WIP will not accumulate in the system, but will move through as it was released, as if on a conveyor (Graves 1988)
- Fractional lead times can be modelled under assumptions of uniform activity level over periods

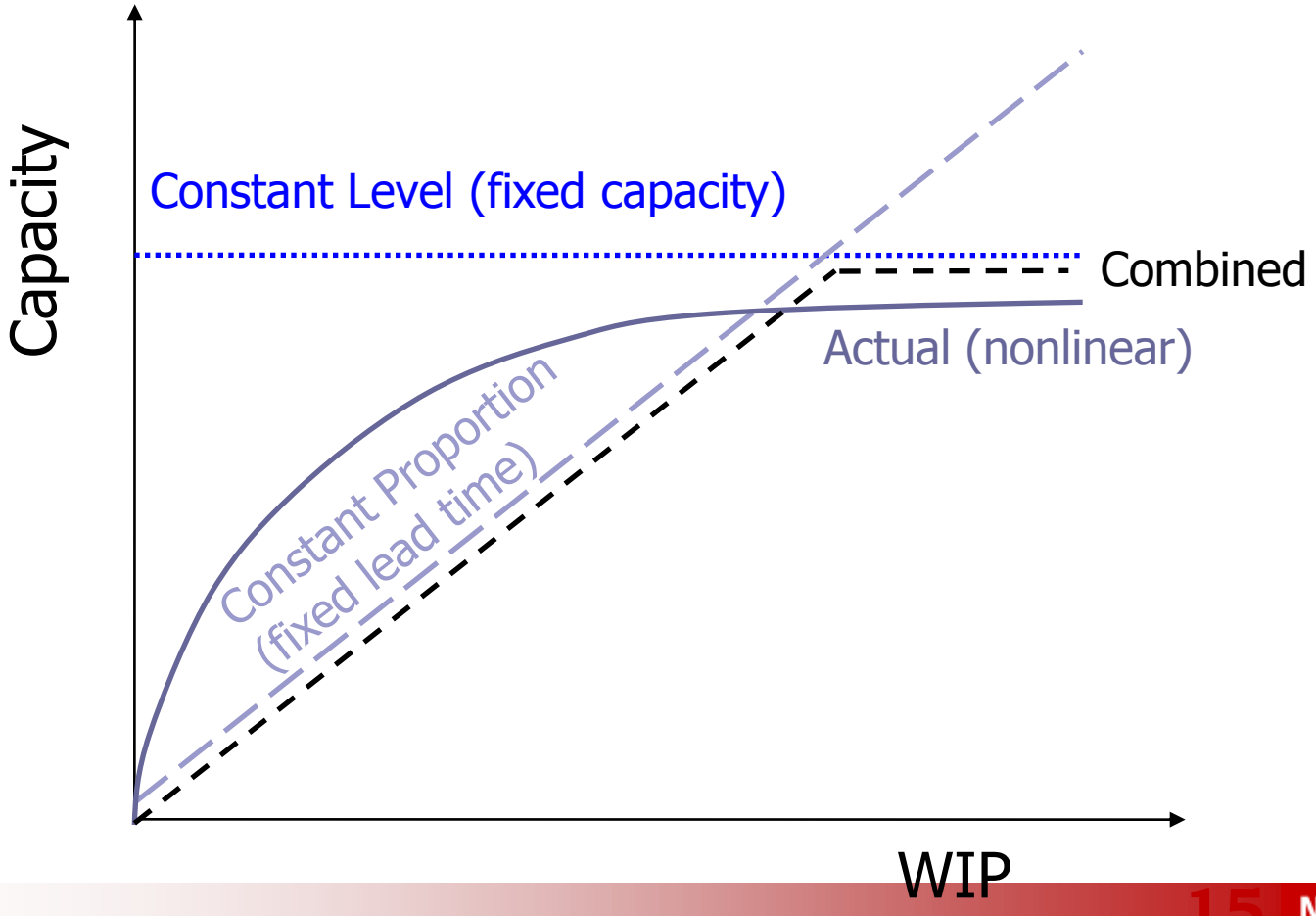
Interesting anomalies

- WIP cannot accumulate
  - Quantity available to be worked on in period t is always R_{t-L}
- Will never hold finished goods inventory unless utilization = 1 in some period
  - Clearly not the case in practice
- Nonzero dual prices only when utilization = 1 in a period
  - Should not be the case when WIP is considered in the objective function

Clearing function models

- Proposed initially by Graves(1986), Karmarkar(1989) and Srinivasan et al.(1988)
- Expresses expected output in a planning period as a function of some measure of planned workload in the period
  - Actual workload is a random variable…

Clearing Functions



Deriving clearing functions

- Lots of open questions remain
  - Historical data, simulation models
  - Variety of functional forms have been proposed
  - Assume single-variable form…for now

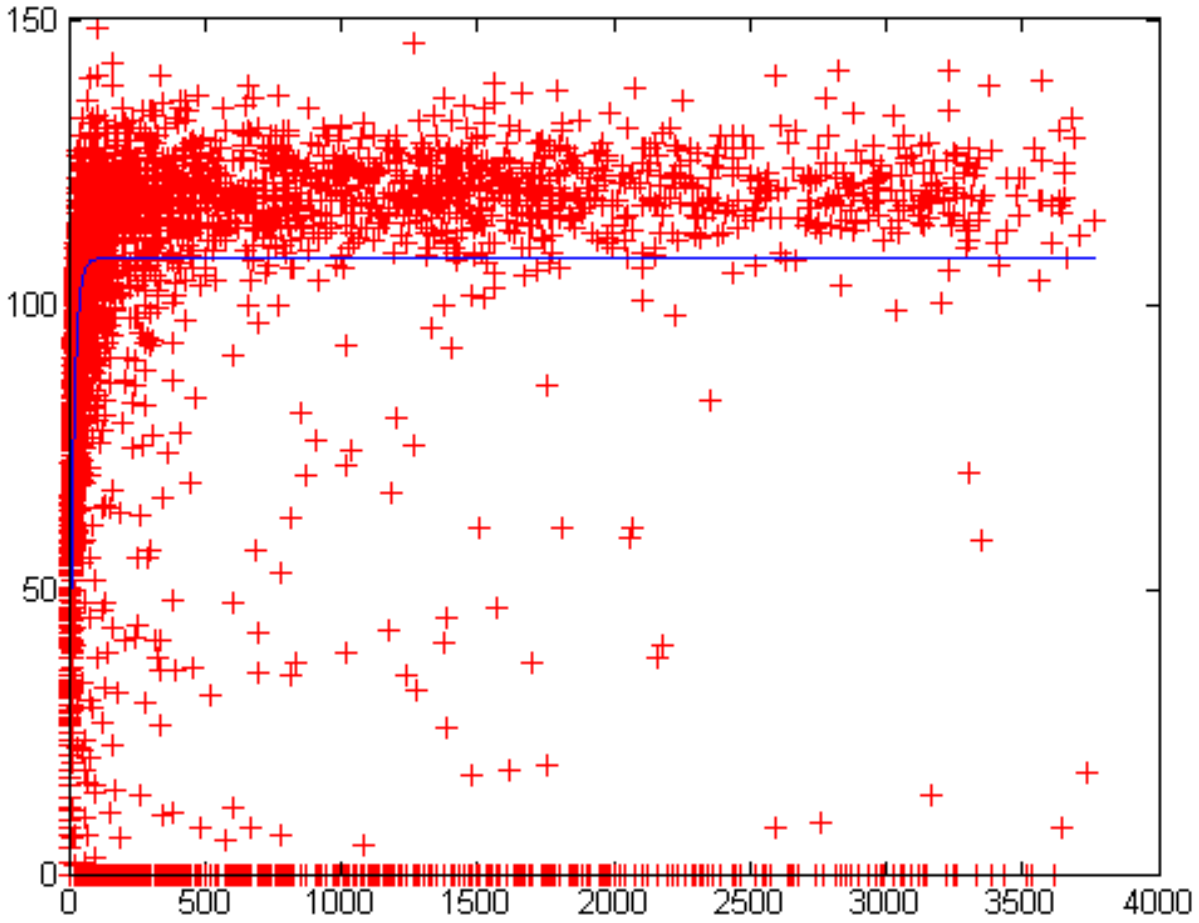
$$f(W) = K_1\left(1 - e^{-K_2 W}\right)$$

(Srinivasan et al. 1988)

$$f(W) = \frac{K_1 W}{K_2 + W}$$

(Agnew 1976, Karmarkar 1989)

The reality…



Clearing function model

(Karmarkar 1989)

$$MIN \quad \sum_{t=1}^{T} (h_t^w W_t + h_t^I I_t + m_t R_t + c_{tw} X_t)$$

Subject to :

$$W_{t-1} + R_t = W_t + X_t \qquad t = 1..T$$

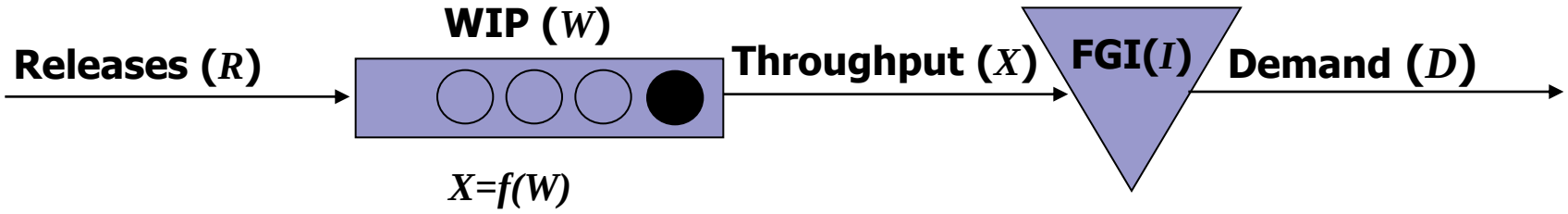
$$I_{t-1} + X_t = D_t + I_t \qquad t = 1..T$$

$$X_t \le f(W_{t-1}, R_t, C) \qquad t = 1..T$$

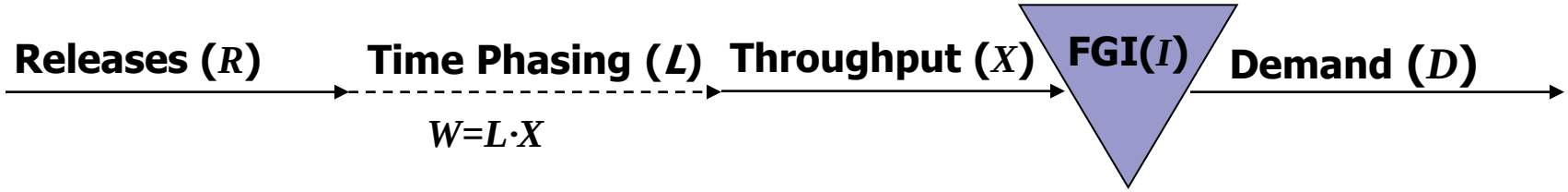
$$W_t,\ I_t,\ R_t,\ X_t \ge 0 \qquad t = 1..T$$

Different View of the World!

- Clearing function model



- Fixed lead time



Multiple Products

- Difficulties in handling the effects of product mix
- Can create capacity for one product by holding high WIP of another
  - Long lead times for some products, very short for others
- Need to make sure production is consistent with WIP mix
  - “No passing” idea

Idea

- Want to get a constraint that
  - Involves the clearing function
  - Which relates to the total WIP in the system
  - But uses only the WIP of the individual product
- Assume all products see the same average time in system, which is equal to the system average
  - FIFO processing
  - Processing time dominated by queue time
  - Implies system is at relatively high utilization

Allocated CF Formulation

$$\min \sum_t (\phi_{it} X_{it} + \omega_{it} W_{it} + \pi_{it} I_{it} + \rho_{it} R_{it})$$

subject to

$$W_{it} = W_{i,t-1} - X_{it} + R_{it}, \text{ for all i, t}$$

$$I_{it} = I_{i,t-1} + X_{it} - D_{it}, \text{ for all i, t}$$

$$X_{it} \le Z_{it} f_t \left(\sum_i \frac{\xi_{it} W_{it}}{Z_{it}} \right), \text{ for all i and t}$$

$$\sum_i Z_{it} = 1 \text{ for all t}$$

$$X_{it}, W_{it}, I_{it}, R_{it}, Z_{it} \ge 0 \text{ for all i, t}$$

Outer linearization of CF

$$\min \sum_t \sum_i (\phi_{it} X_{it} + \omega_{it} W_{it} + h_{it} I_{it} + \rho_{it} R_{it})$$

subject to

$$W_{it} = W_{i,t-1} - X_{it} + R_{it} \text{ for all i and t}$$

$$I_{it} = I_{i,t-1} + X_{it} - D_{it} \text{ for all i and t}$$

$$\xi_{it} X_{it} \leq \alpha^c \xi_{it} W_{it} + Z_{it} \beta^c \text{ for all i, t and c}$$

$$\sum_i Z_{it} = 1 \text{ for all t}$$

$$Z_{it}, X_{it}, W_{it}, I_{it} \geq 0 \text{ for all i and t}$$

Observations

- Classical LP formulation is a relaxation if parameters set consistently
- Operates by creating an aggregate clearing function for the system, and then allocating this out among products
- How far does this get us beyond LP?

Lead time iteration

- Proposed by Hung and Leachman(1996)
- Uses an LP model with fixed, non-integer lead time estimates to develop a release plan
- Simulation model to estimate realized cycle times under that release plan
- Update lead time estimates and resolve LP
- Iterate until convergence (hopefully…)
- Ought to do better than fixed lead time estimate….

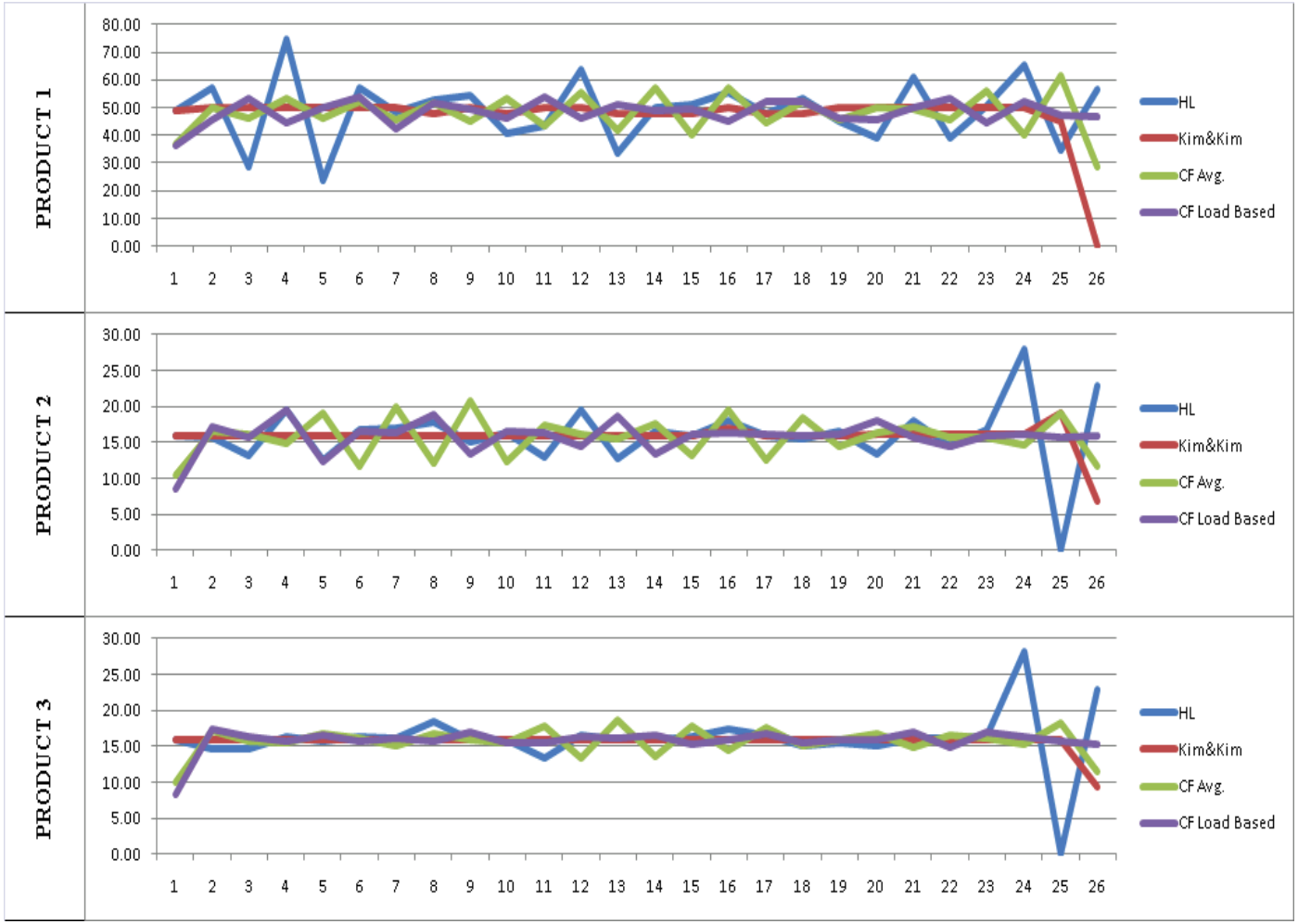
Other iterative methods

- Kim and Kim(2001) – use weights observed from simulation, rather than calculating them using flow times
  - Shows excellent convergence behavior.
- Byrne and Bakir(1999) – special case of Kim and Kim(2001) when flow times are 1 period
  - Poor convergence when this is not valid!
- Riano(2003): sophisticated weight calculation and iteration derived from transient queueing analysis

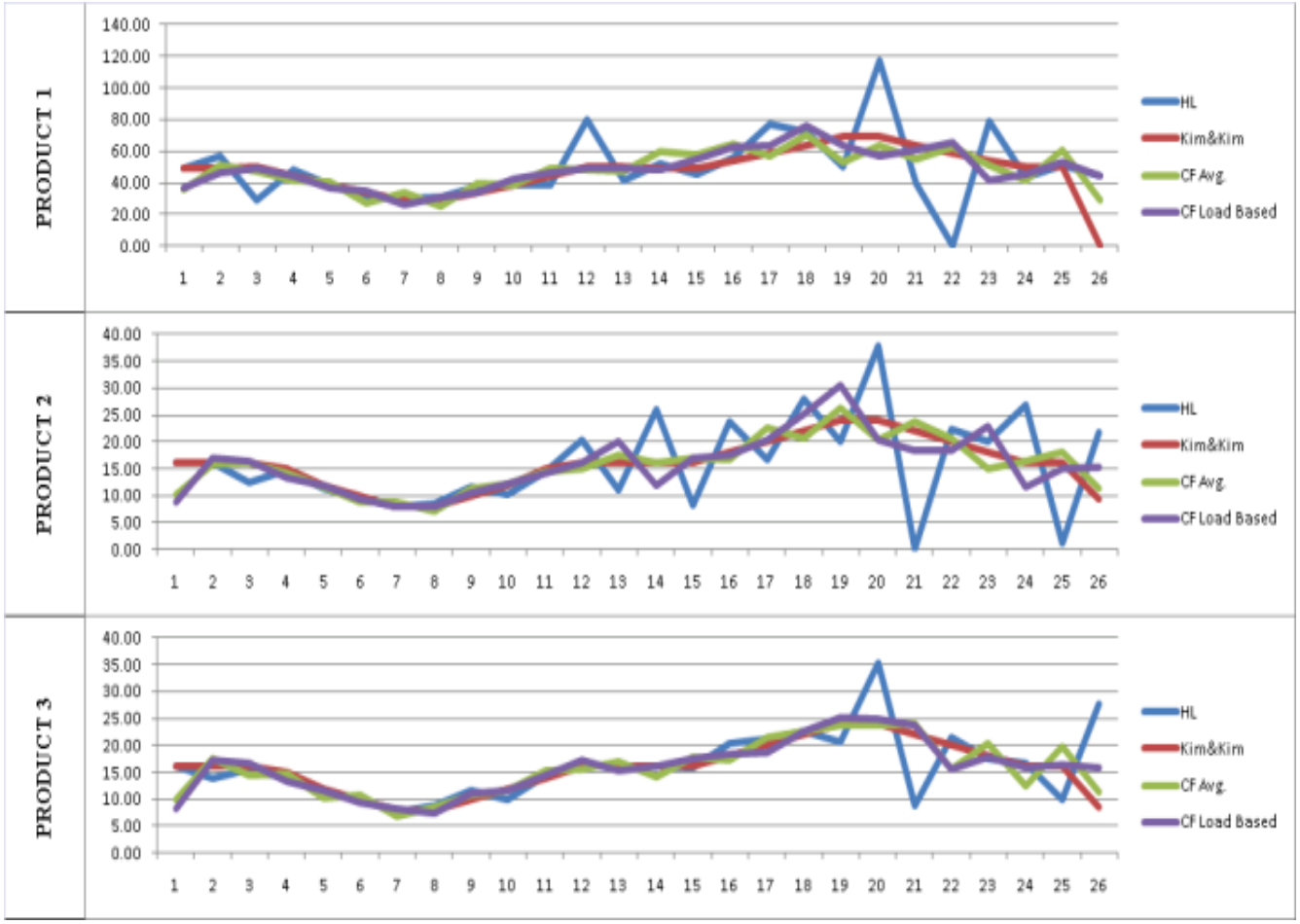
Observations

- Kim and Kim(2001) converges very clearly where Hung and Leachman appears to fluctuate
- Subtle differences in LP models allowed for a much as possible
- Weights from simulation differ significantly from those estimated using the flow times
- Exhibits some cycling behavior – also observed by Riano

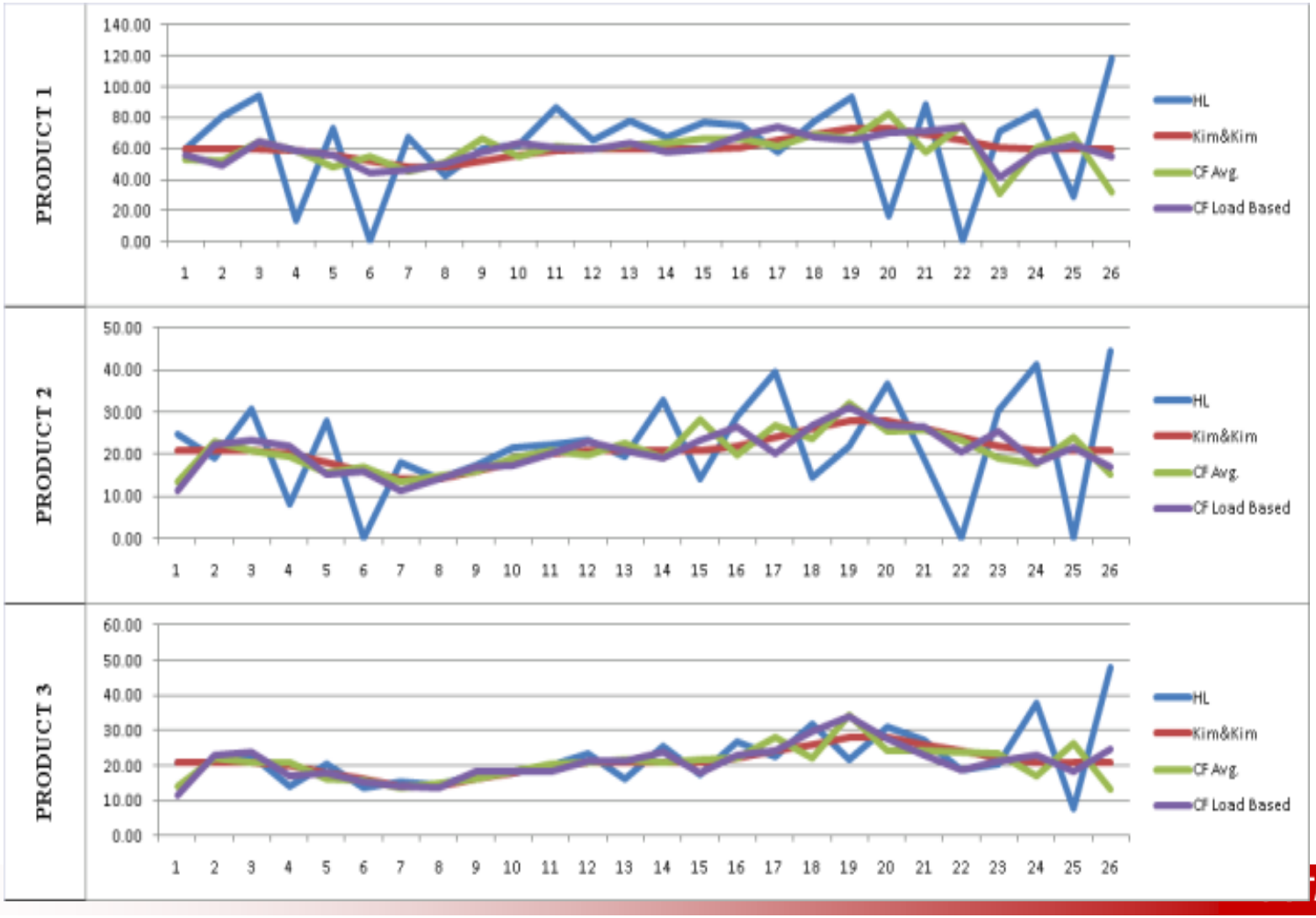
70% - Short - Constant



70%-Short-Varying



90% - Short - Varying



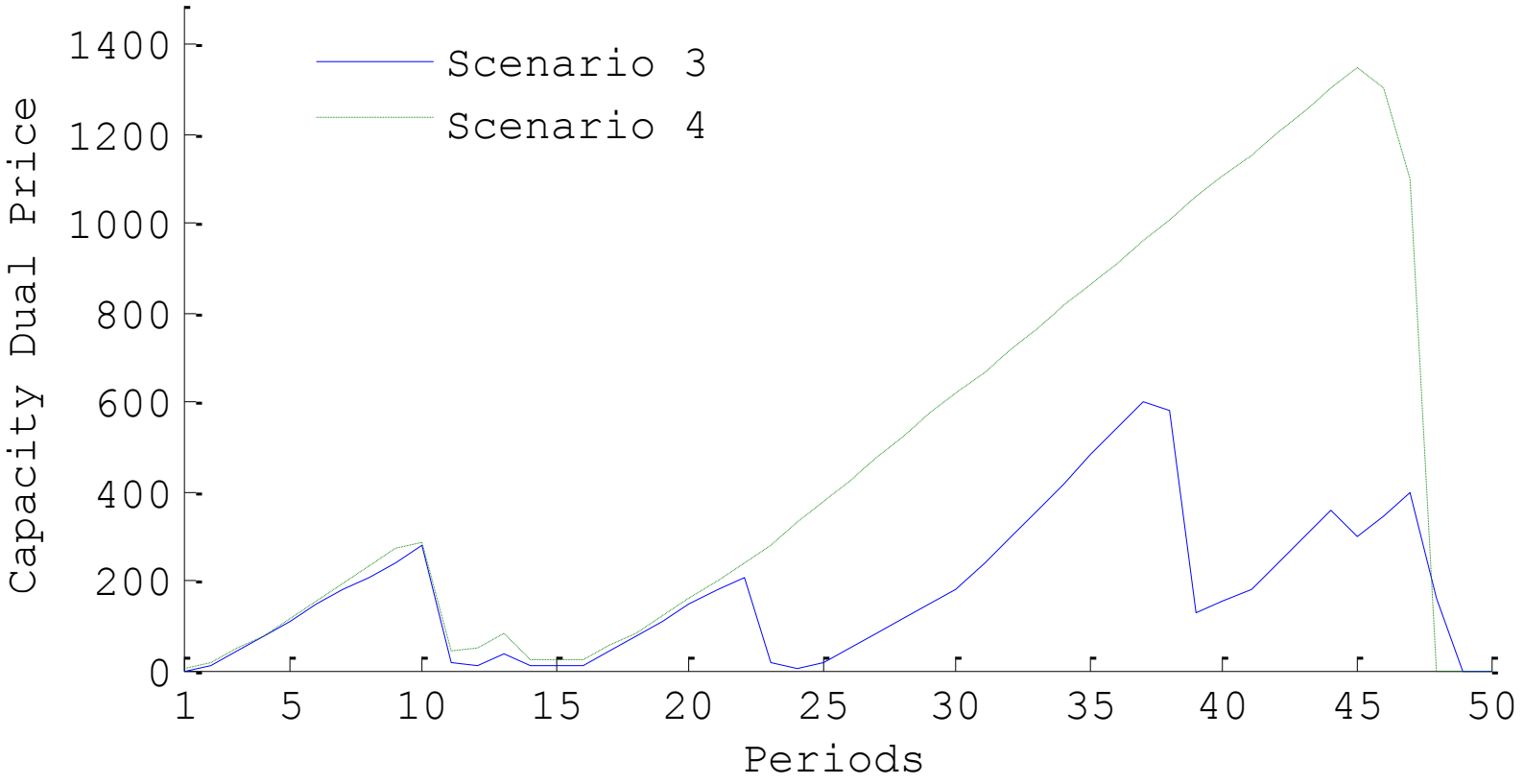

Total Profit
$120,000
$100,000
$80,000
$60,000
$40,000
$20,000
$0
S-70-C
S-70-V
S-90-C
S-90-V
L-70-C
L-70-V
L-90-C
L-90-V
HL
Kim&Kim
CF Avg.
CF Load Based

Observations

- Clearing function models consistently yield statistically better profit
- Also somewhat different solutions
- Clearly, iterative methods converge to something different…
  - Local optimality issues…
- The basic clearing function form used here is quite primitive, but still leads to better plans

Dual variables

- Studied by Kefeli et al.(forthcoming) for simple single product single stage CF model
- CF is piecewise linearized
- Similar results hold for single stage multiproduct systems
- Multistage multiproduct systems yield very interesting insights into shifting bottlenecks as product mix changes


Scenario 3
Scenario 4
Capacity Dual Price
1400
1200
1000
800
600
400
200
0
1
5
10
15
20
25
30
35
40
45
50
Periods

Observations

- Clearing function model yields the intuitively correct dual price behavior
  - Nonzero prices when utilization below 1
  - Exploding prices when utilization is high
  - Cost of queues in the system
- Quality of dual price estimates at high utilization depends on accuracy of outer linear approximation

Where does this leave us?

- When properly parameterized, clearing function models have repeatedly been found to give better solutions than LP models
- Intuitively correct dual price behavior
- Yield smooth, “practically acceptable” release schedules
- Allocated CF model addresses multiproduct issues, yields tractable LP model

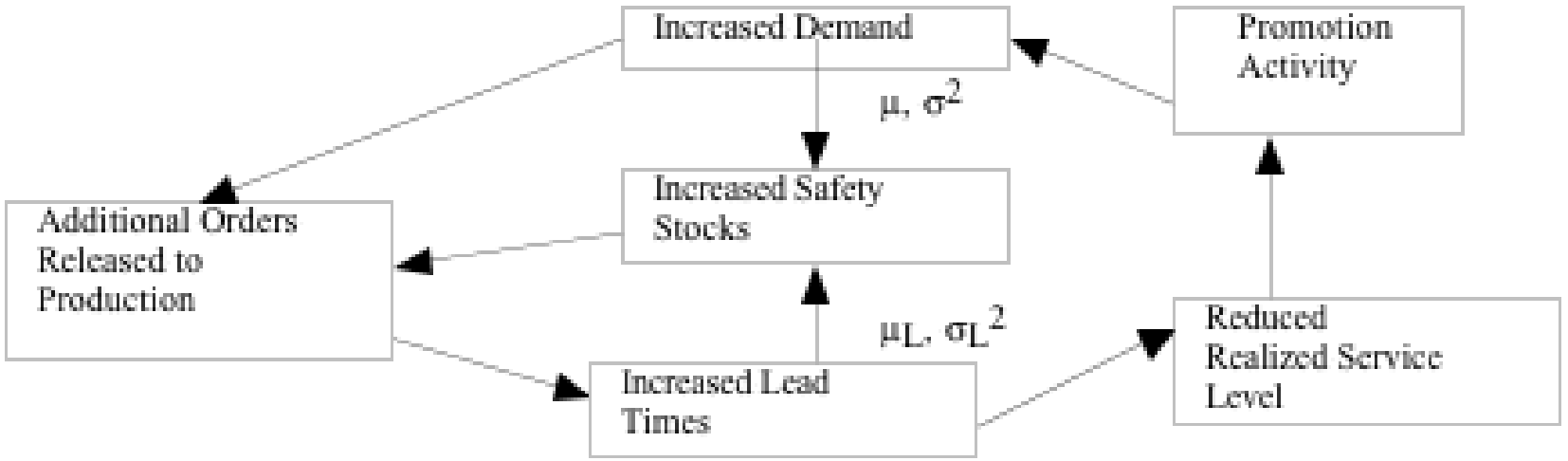
However…

- The single-variable clearing functions used here are a VERY crude deterministic equivalent for a stochastic production system governed by queuing behavior
  - Likely to be transient – Selcuk et al.(2007), Missbauer(2010), Riano(2003)
  - Other factors, such as operator interference
- Even estimation from simulation data is not a straightforward process
- Represents an expectation over all states the simulation entered

Incorporating Demand Uncertainty

- In most industrial applications, production planning and safety stock determination are addressed in tandem
- However, safety stock levels driven by demand over lead time
- As utilization goes up, so does lead time, so does safety stock requirements, driving lead times even higher

Feedback loops…



The challenge…

- A planning algorithm that
  - Understands demand uncertainty at least better than the deterministic models in wide use today
  - Comprehends the relationship between utilization, lead times and safety stocks
  - Is transparent to users, and solvable with off the shelf software
  - Seems to rule out an optimal solution, but…

Basic approach

- Capture load-dependent lead times using the clearing function construct
- Model demand uncertainty using chance constraints
- Start with a single-stage, single product model to begin with
  - If we can't do this, we might as well give up…

On chance constraints…

- Chance constraints have been largely superseded by multistage stochastic programming with recourse
- Valid modeling issues regarding how CC models recourse actions
  - We can violate a constraint with some probability, but the cost of that violation is not captured accurately

However…

- In our context, chance constraints have a natural interpretation as service level
  - Probability of inventory going negative
- Under the right conditions, can be solved with standard LP solvers
- Can accommodate rudimentary recourse actions through decision rules
- Far more user-friendly than stochastic programming

Basic formulation

- Clearing function models explicitly represent WIP inventory
- Write chance constraint on probability of inventory position exceeding demand over lead time
- Assume independent normal demand
- Model plans inventory position; decision rules allow dynamic adaptation as demand is realized

DYNIP Formulation

$$\textbf{\textit{Minimize}}\ E\left(\sum_{t=1}^{T} h_t \mathbf{I_t} + h_t W_t\right) \qquad \text{subject to}$$

$$E(\mathbf{I_t}) = E(\mathbf{I_{t-1}}) + X_t - E(\mathbf{D_t}) \qquad \forall t = 1,\ldots,T \qquad \text{(FGI BALANCE)}$$

$$E(\mathbf{W_t}) = W_0 + \sum_{i=1}^{t} (Y_i + E(\mathbf{D_i}) - X_i) \qquad \forall t = 1,\ldots,T \qquad \text{(WIP BALANCE)}$$

$$W_0 + I_o + \sum_{i=1}^{t} Y_i \geq G^{-1}_{t+1,t+L_t}(\alpha) \qquad \forall t = 1,\ldots,T \qquad \text{(SERVICE LEVEL)}$$

$$X_t \leq a_k E(\mathbf{W_{t-1}}) + b_k \qquad \forall t = 1,\ldots,T\ ; \forall k = 1,\ldots,n \qquad \text{(CAPACITY)}$$

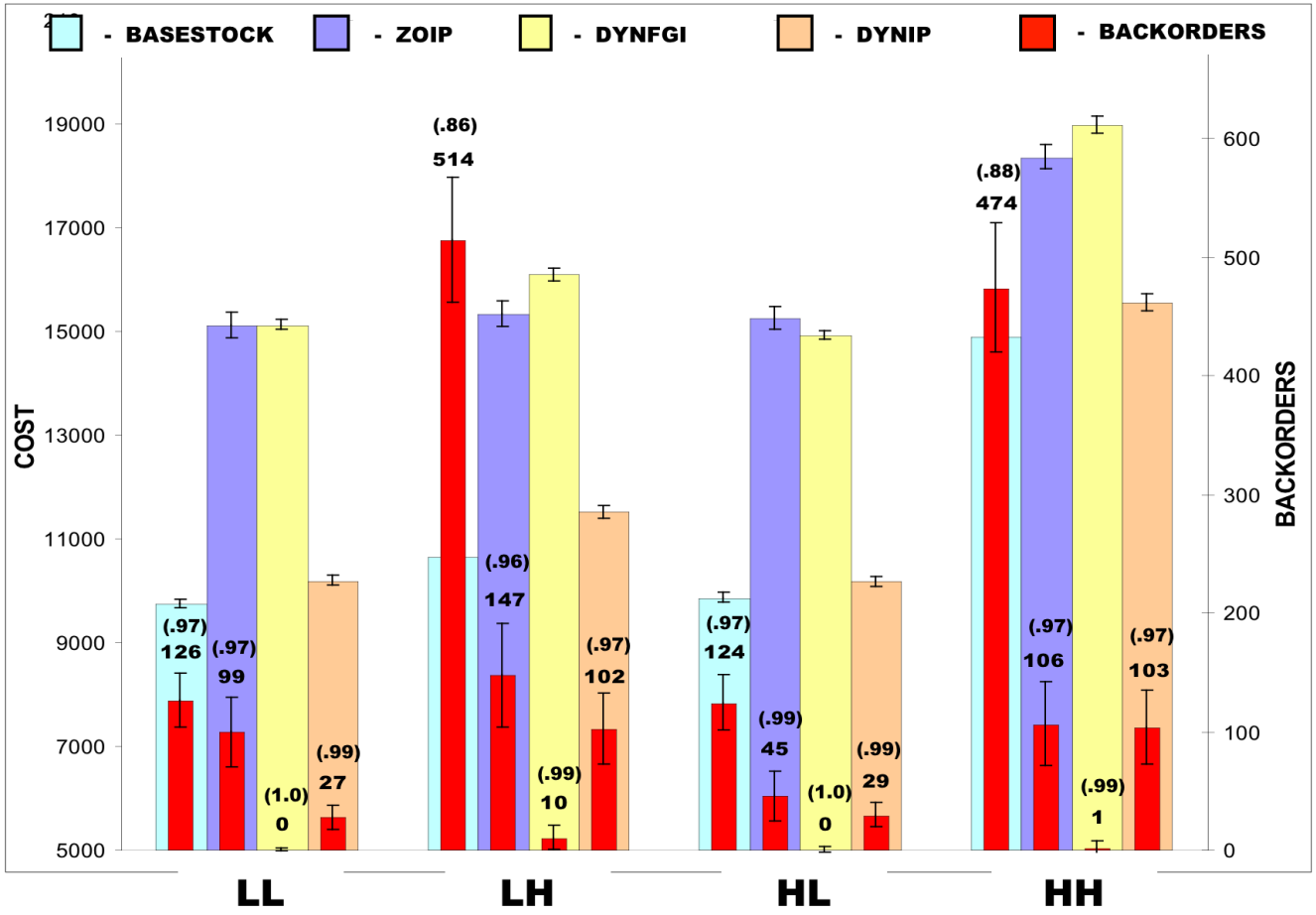
$$Y_t + (\mathbf{D_t})_{\min} \geq 0 \qquad \forall t = 1,\ldots,T \qquad \text{(REL. NON-NEG.)}$$

$$X_t,\ E(\mathbf{I_t}),\ E(\mathbf{W_t}) \geq 0 \qquad \forall t = 1,\ldots,T$$

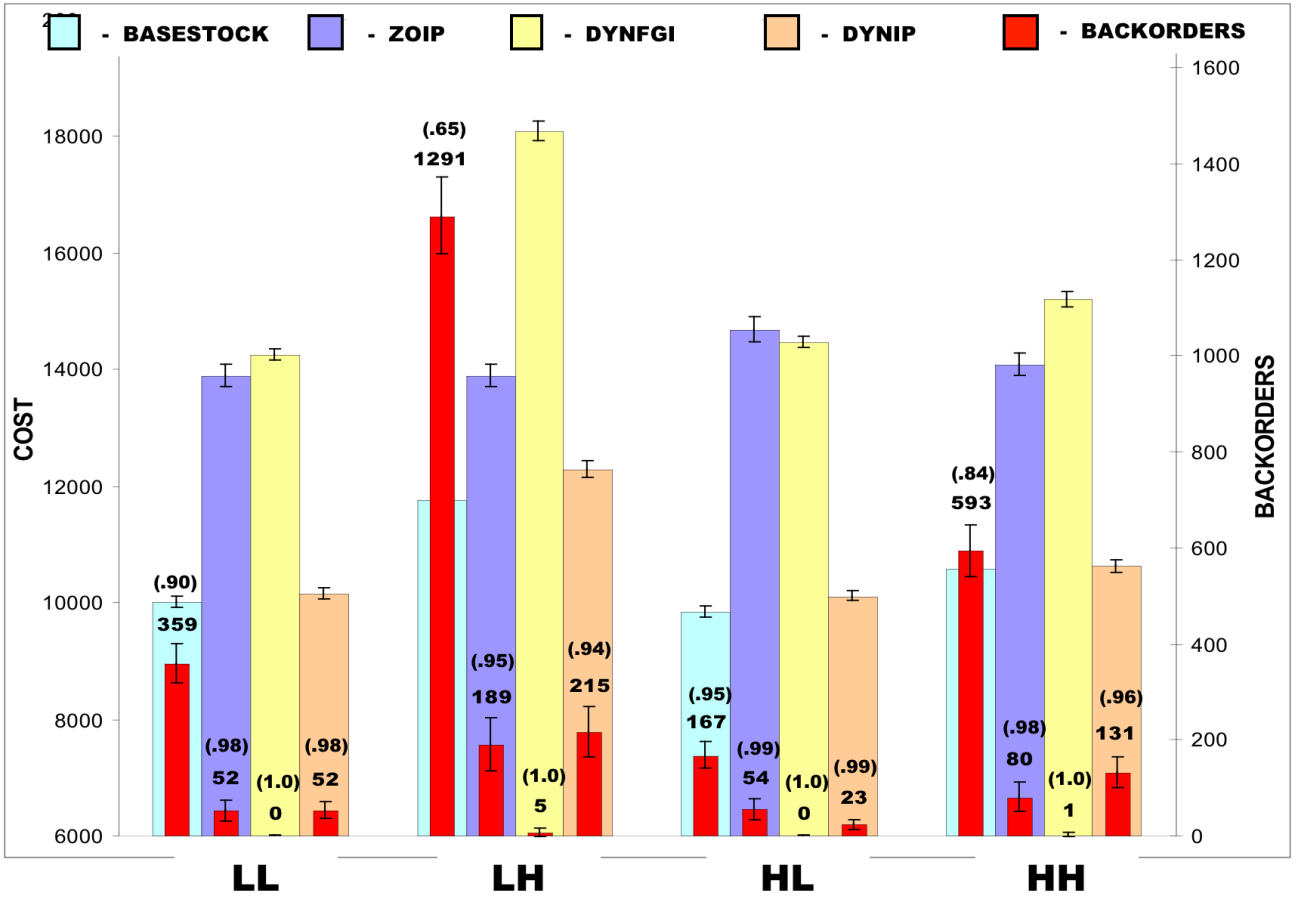
Computational Experiments

- Generate a set of demand means, assuming fixed CV
- Solve the planning model
- Simulate with multiple realizations of demand from the same distributions
- Collect statistics, relaxing several modeling assumptions

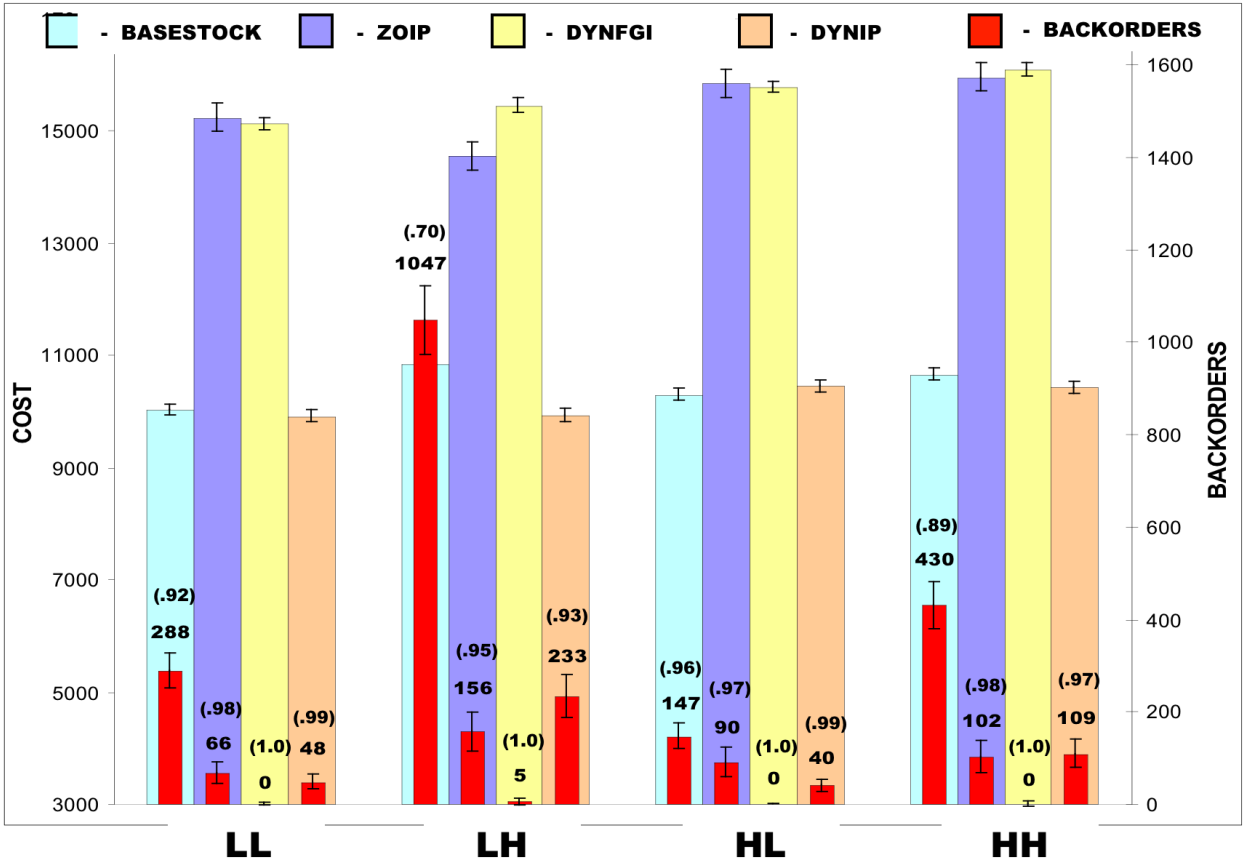
Random demand means



Increasing demand means



Increasing/decreasing demand means



Observations

- The ZOIP and DYNFGI models behave quite similarly
  - Inventory accumulates over time in both
- Base stock model and DYNIP have similar inventory levels, but DYNIP does better on backorders
- When LT underestimated, DYNIP does a lot better on service level
- Remember both have the same service level wired in!

Observations

- Chance constraint models stock out by very little, even when stockouts are quite frequent
- ZOIP, DYNFGI have service levels far above the designated level
- Treat “demand” LT as an exogenous parameter!
- Can apply lead time iteration to this model
  - Approximate $L_t = W_t/X_t$
  - Consistent convergence, good performance – Orcun et al.(2009)

In summary…

- We have been acting like production planning is a solved problem
- In reality, it is far from it
- Lots of interesting, challenging issues remain
- Many of these we have not addressed here today
  - Incorporation of supply and demand uncertainty
- Lots of new work in the last ten years

Clearing functions

- Quite primitive versions give quite good results under a surprising variety of experimental conditions
- This is not an application where we need six decimal places of precision for the solution to be useful
- Theory is well supported by queuing for single-stage steady state environments
  - VERY long planning periods…

Future directions

- Transient systems, multistage with dependent arrivals pose serious conceptual problems
  - Decisions at one stage change shape of CF downstream
- Relationship to iterative LP models needs to be elucidated
- Conjecture: Given the solution to the CF model, a set of weights can be constructed that would allow an LP yielding the same solution to be formulated

Future directions

- Extensions to dynamic lot sizing problems
- Yield models with complex multivariate CFs
  - Functions of both WIP and lot sizes of ALL products
  - Highly nonconvex optimization models
  - No setup costs
- Incorporating uncertainties is also an important direction
  - Find useful, usable approximations as opposed to narrowly optimal approaches