MEC-E7001 Examination 20.2.2024 – Some answers

1. A factory manufactures 200 products a year starting from raw materials. Average throughput time from start to completion is 5 weeks and manufacturing cost is about 100 000 euros/product.

a) What is the average value of work in process? (solve this after b)

Because manufacturing starts from raw
materials, WIP in euros is less than 19.25 x 100 000 = 1925000 euros, somewhere between 0,5 to 1 times of 1900000 euros => Range is [962500, 1925000]. Probably about 1000000 euros.

b) What is the average number of products found at the factory (incomplete or complete) at any time?

$$L = \lambda W$$

L – Average number of items within the system

λ – Average arrival rate of items into and out of the system

W – Average time an item spends in the system

==========

Step 1: Understanding the Terms

Throughput time: This is the average time it takes for a product to be completed from start to finish, which is 5 weeks.
Manufacturing cost: The cost to manufacture one product is €100,000.

Step 2: Annual Output

The factory manufactures 200 products a year.

Step 3: Calculating WIP in units

To calculate WIP, we first need to determine how many products, on average, are in process at any given time. This can be estimated using the formula derived from Little's Law:

WIP (units) = Throughput rate (units/time) x Throughput time (time)

Here, the throughput rate is 200 products per year. Since the year has 52 weeks, the weekly throughput rate is:

Throughput Rate = WIP (over a year) / Throughput time (year) = 200 products / 52 weeks = 3.85 products/week

Multiplying this rate by the throughput time gives:

WIP (5 weeks) = Throughput Rate x 5 weeks = 3.85 products/week × 5 weeks = 19.25 products

This indicates there are about 19 to 20 products in the WIP inventory at any time.

2. Compare linear regression modeling and neural network modeling. What is the purpose of such modeling?

• Generally input-output models fitted to data. Used for cost, work content estimation, process modelling, stock exchange price prediction etc. Easy to use.

What are the main characteristics of each, how do they differ and give examples of typical applications for both of them.

• Linear regression: linear, goodness of fit can be evaluated with defined statistics, model provides information about the process (system behavior).
• Neural networks: can be used for complicated, nonlinear systems, “black box” character, can be unpredictable. Need a lot of data for fitting.

MEC-E7001 Examination 23.2.2021

Write full sentences using clear and readable fonts

1. Describe how and why the following factors affect products’ throughput time in production:
    a) System utilization rate
    b) Processing time variation
    c) Batch size

In a stochastic process throughput time depends on
• Utilization rate
• Variation
• Processing time (and in case of batch production also of the batch size and set-up time)
• G/G/1 approximation1
Cycle Time = VUT + T,
where

V = (CV^2 processing time + CV^2 interarrival time)/2, CV is Coefficient of Variation.
U = u / (1 - u), where u = utilization rate
T = processing time.
• Because high utilization rate increases U strongly, it is beneficial
to focus improvement efforts on bottle necks

Queuing theory

For queuing systems:

$L = \lambda W$ (WIP)

$L_q = \lambda W_q$ (Queue length)

$L_p = \lambda * 1/\mu = \lambda/\mu$ (In process, machine, being served)

$W = W_q + 1/\mu$ (Throughput time)

$L = L_q + L_p = L_q + \lambda/\mu$ (WIP),

where

λ = Arrival rate (note: $1/\lambda$ is interarrival time)

μ = Processing rate (products per time unit, $1/\mu$ is processing time)

L = Products in system (= WIP, Work in Progress)

W = Throughput time

L_q = Queue length

W_q = Waiting time in queue

a) Utilization rate:

Queuing theory

If queue length or waiting time is known, all other interesting values can be calculated. For given utilization rate, queue length L_q depends mainly on interarrival time and processing time distributions.

For M/M/1

$L_q = \lambda^2 / \mu(\mu - \lambda) = \rho^2/(1 - \rho)$, $L = \rho^2/(1 - \rho) + \rho = \rho/(1 - \rho)$

where

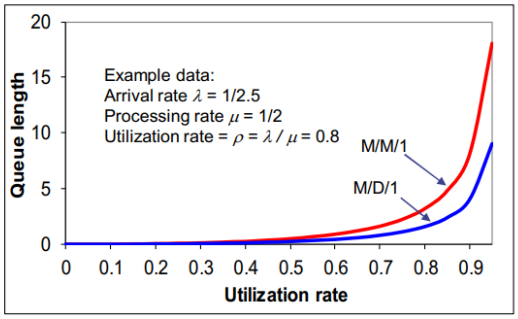
$\rho = \lambda / \mu = L_p$, utilization rate

For example $1/\lambda = 2.5$ ja $1/\mu = 2$:

$L_q = 0.8^2/(1-0.8) = 3.2$

$L = L_q + \lambda/\mu = 3.2 + 0.4/0.5 = 4$

$W = L / \lambda = 4 / (1 / 2.5) = 10$



The utilization rate is proportion of time that a production system is actively used to process products compared to the total available time. High utilization rates can lead to decreased throughput times because resources are maximally employed, thus reducing idle time. However, if the utilization rate is too high, it can also lead to bottlenecks, as there is little to no buffer for handling variability in production, which can ultimately increase throughput times.

b) Processing Time Variation:

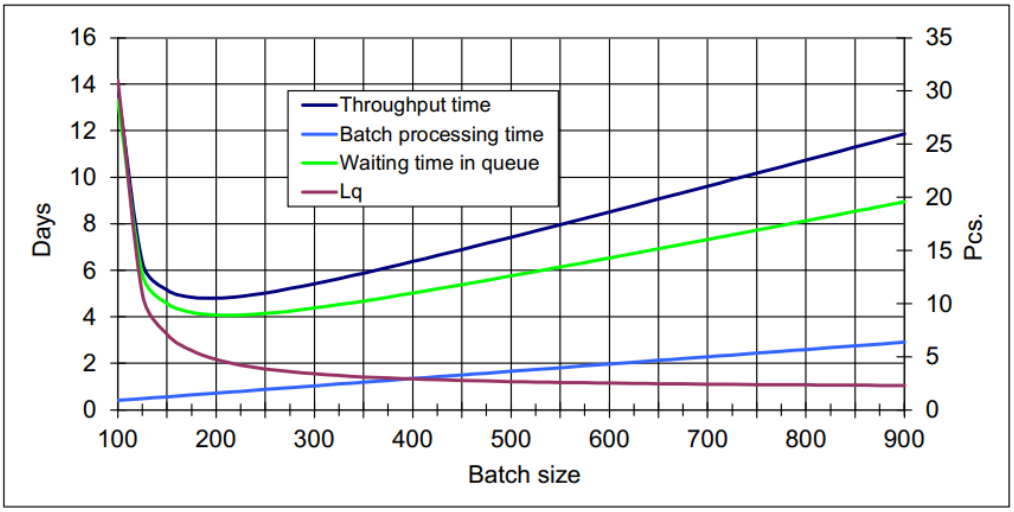
Processing time variation refers to the fluctuation in the amount of time it takes to complete different production tasks or process different items. When processing times are consistent and predictable, it is easier to schedule work and manage the flow of products through the production system, potentially reducing throughput times. In contrast, high variation can cause disruptions and delays, as subsequent processes may need to wait for the previous processes to complete, thereby increasing the overall throughput time.

V = (CV^2 processing time + CV^2 interarrival time)/2, CV is Coefficient of Variation.

c) Batch Size:

Waiting and batch sizes – throughput times

When batch size is increased, utilization rate decreases, because the number of set-ups reduces. This shortens queues when measured in number of batches, but the larger batches increase waiting time.



Batch size is the quantity of items that are processed or produced together as a group. Smaller batch sizes often lead to shorter throughput times because each batch moves more quickly through the production system. This can be advantageous in reducing WIP and can be more responsive to changes in demand. However, smaller batches may require more frequent setups or changes, which can increase throughput time. Conversely, larger batch sizes reduce throughput time but it can also result in longer queues at each production station, increasing the throughput time.

2. Describe the following optimization problem types, especially from the difficulty of optimal solving point of view. Which solving methods can be used for each?
    a) Linear problem
    b) Linear integer problem
    c) Convex nonlinear problem
    d) Nonlinear non-convex problem with integer constraints

a) Linear Problem:

A linear problem is an optimization problem where both the objective function and the constraints are linear. Linear problems are "easy" to solve efficiently (in polynomial time) using well-established methods such as the simplex Method Aapopular algorithm that iteratively moves towards the optimal solution by traversing the edges of the feasible region defined by the constraints.

b) Linear Integer Problem:

A linear integer problem is a particular case of a linear problem where some or all the decision variables are constrained to be integers. This makes the problem harder to solve than the continuous case because it is often necessary to search through a discrete, large set of possibilities.

Solving Methods:

Branch and Bound: This method involves dividing the problem into smaller subproblems (branching) and then calculating lower and upper bounds on the optimal solution, cutting off (bounding) subproblems that cannot contain the optimal solution.

c) Convex Nonlinear Problem:

A convex nonlinear problem is an optimization problem with a convex objective function and convex constraints. Convexity ensures that any local optimum is also a global optimum, which simplifies the problem significantly compared to non-convex problems.

Solving Methods:

Gradient Descent: A first-order iterative optimization algorithm for finding the minimum of a function, requiring that we have access to the gradient of the problem.

d) Nonlinear Non-convex Problem with Integer Constraints:

This is the most challenging class of optimization problems. The objective function and/or constraints are non-convex, and some or all of the variables are constrained to be integers. Such problems are generally NP-hard, meaning that no polynomial-time algorithms are known for their solution.

Solving Methods:
Genetic/Evolutionary Algorithms: These are search heuristics that mimic the process of natural selection to generate high-quality solutions to optimization and search problems.

3. Describe production control of a job shop using priority rules. Describe typical rules. Why are they used? What are the advantages and disadvantages of such a system compared to real optimization when making production scheduling decisions?

Production control in a job shop using priority rules involves managing the job flow and resource allocation based on a set of predefined criteria. These priority rules help determine the order in which jobs should be processed. The use of priority rules simplifies the complex decision-making processes involved in job shop scheduling.

Typical Priority Rules:

First Come, First Served (FCFS): Jobs are processed in the order they arrive.
Shortest Processing Time (SPT): Jobs with the shortest processing time are done first. This is for total/average flow time (or throughput) minimization
Earliest Due Date (EDD): Jobs with the earliest due date are given priority. This is for Maximum tardiness minimization. Tardiness is the amount of time the job need to be completed pass its due date

For total tardiness minimization no simple algorithm exists. Many heuristics are based on that
• If the schedule is loose and there are few late jobs, EDD rule probably works well
• If the schedule is tight and most jobs are late, SPT rule probably works well

Why Priority Rules are Used:

Priority rules are used in production control to efficiently manage job flow and resource allocation by establishing a clear and structured method for determining the sequence in which jobs should be processed.

Advantages of Priority Rule Systems:

Robustness: Priority rules provide a dependable framework that consistently handles varied production demands and uncertainties.
Reactive: They enable quick adjustments to scheduling in response to changes in job status or workshop conditions.
Ease of Use: Priority rules simplify complex decision-making, making them accessible and straightforward for managers to implement.

Disadvantages of Priority Rule Systems:

Sub-Optimality: They may not provide the most efficient or cost-effective schedule.

Real optimization involves creating schedules based on mathematical models that aim to find the best possible solution given a set of constraints and objectives.

Advantages of Real Optimization: It seeks to find the best possible solution for the entire system, resulting in the global optimum

Disadvantages of Real Optimization:

Complexity: It requires specialized knowledge and tools to implement.
Computationally Intensive: It can be time-consuming and require significant computational resources.
Rigidity: Once an optimal schedule is set, it may be difficult to adapt to changes quickly.

4. How can
   a) average throughput time,
   b) maximum tardiness,
   c) number of tardy products,
   d) total tardiness
   be minimized for a single machine? You do not need to specify models in detail, just describe principle. Describe the assumptions you make concerning your system.
   e) makespan

a) Average Throughput Time
Shortest Processing Time (SPT): Prioritize jobs with the shortest processing times to minimize the overall time jobs spend in the system.

b) Maximum Tardiness:
Earliest Due Date (EDD): Prioritize jobs with the closest due dates to minimize the risk of missing deadlines.

c) Number of Tardy Products:
Minimizing number of tardy jobs is necessary when on time delivery is maximized. Efficient algorithms exist for this. For example:

Jobs are first sorted according to due dates, then

1. Find first tardy job k
2. Find longest job 1..k and remove it and move last in schedule
3. Return to step 2 until there are no more tardy jobs (in the original schedule)

Example:

Job i	Duration P_i	Due date D_i
1	1	2
2	5	7
3	3	8
4	9	13
5	7	11

1. {1,2,3,5,4} k = 3, longest in 1..3 is job 2
2. {1,3,5,4} {2} k = 4, which itself is longest
3. {1,3,5} {2,4}
4. No more late jobs, final schedule: {1,3,5,2,4}

d) Total Tardiness:

For total tardiness minimization no simple algorithm exists. Many heuristics are based on that
• If the schedule is loose and there are few late jobs, EDD rule probably works well
• If the schedule is tight and most jobs are late, SPT rule probably works well

or Weighted Tardiness Scheduling: Assign weights to jobs based on importance or penalty costs and schedule to minimize the weighted sum of tardiness.

e) Makespan:

Makespan is the total processing time for all jobs, unlike total throughput time which also includes waiting time and also setup time. Makespan is not affected by the order of jobs, so any processing order is okay. Makespan optimization is a relevant criterion when set-ups are taken into account

Assumptions:
The system cannot have extra constraints like earliest starting time for a) and b) to hold.

Single Machine: There is only one machine available for processing all jobs.

Deterministic Processing Times: The processing times for all jobs are known and constant.

No Preemption: Once a job starts processing, it continues uninterrupted until completion.

MEC-E7001 Example examination

1. Create single machine schedules for the jobs below that minimize a) work in process and b) maximum tardiness. What are the reults and how did you end up with them?

Job	Duration	Due date
1	7	5
2	4	12
3	6	17
4	3	24
5	4	9
6	3	21

To create schedules that minimize a) work in process and b) maximum tardiness, we will need to apply different scheduling strategies:

a) To minimize work in process (WIP), we can use the Shortest Processing Time (SPT) rule, which schedules jobs in order of increasing duration. This tends to reduce the time jobs spend in the system (i.e., WIP).

Applying SPT to the given jobs, we would order them by duration as follows:

Job 4 - Duration 3
Job 6 - Duration 3
Job 2 - Duration 4
Job 5 - Duration 4
Job 3 - Duration 6
Job 1 - Duration 7

b) To minimize maximum tardiness, we could use the Earliest Due Date (EDD) rule, which schedules jobs in the order of increasing due dates.

Applying EDD to the given jobs, we would order them by due dates as follows:

Job 1 - Due Date 5
Job 5 - Due Date 9
Job 2 - Due Date 12
Job 3 - Due Date 17
Job 6 - Due Date 21
Job 4 - Due Date 24

Results and Explanation:

For a) Minimizing WIP with SPT:

The jobs are processed in the order of 4, 6, 2, 5, 3, 1.
The total processing time is the sum of durations, which is 3+3+4+4+6+7 = 27.
This strategy may lead to some jobs being late, but the average time a job spends in the system is minimized.

For b) Minimizing Maximum Tardiness with EDD:

The jobs are processed in the order of 1, 5, 2, 3, 6, 4.

To calculate tardiness, we sum the job durations as we go along and compare the cumulative time to the due dates.

Job 1 starts at time 0 and ends at time 7 (tardiness = max(7-5, 0) = 2).
Job 5 starts at time 7 and ends at time 11 (tardiness = max(11-9, 0) = 2).
Job 2 starts at time 11 and ends at time 15 (no tardiness since 15 < 12).
Job 3 starts at time 15 and ends at time 21 (no tardiness since 21 < 17).
Job 6 starts at time 21 and ends at time 24 (no tardiness since 24 < 21).
Job 4 starts at time 24 and ends at time 27 (no tardiness since 27 < 24).

Maximum tardiness in this case is 2, which occurs for both Job 1 and Job 5.

These results show the trade-off between different scheduling objectives. SPT is good for minimizing WIP, but may not minimize tardiness. EDD focuses on due dates, which minimizes tardiness but can increase WIP.

2. Describe an FM system from the production planning and control point of view.

• FM systems (FMS, Flexible Manufacturing System) are automatic centrally controlled machining systems that consist typically of NC machine tools and supporting machines like washing machines etc., central workpiece changing and storing system (AS/RS, Automatic Storage / Retrieval System) and loading and unloading stations

• The purpose is to increase utilization rate and weekly running time of machines with automation
• Throughput time is secondary objective in the sense that long daily running time guarantees a high turnover rate of orders anyway
• Use of flexible technology makes manufacturing a large variety of products possible
• Set-ups are mostly external and therefore small batch sizes or one-off manufacturing is possible

FM systems' properties

From the production control point of view FM systems are job shops with the following features

• Alternative parallel machines, but
• Similarity of machines is affected by tool set-up
• AS/RS is a scheduled or at least loaded resource in the system
• Number of pallets is limited
• Set-ups are external and they mainly consist of change of fixtures on the pallets
• Change of fixtures takes considerable time
• Workpiece change and fixture change are done manually during manned shifts

FM system control constraints
• Tool set-up must allow execution of manufacturing program
• Fixture set-up must allow execution of manufacturing program
• There should be enough work for the unmanned shifts
• Fixture and workpiece change must be executed during the manned time
• AS/RS system must not become too limiting a bottleneck
• In addition, normal manufacturing targets must be met: required parts must be manufactured within the given time

FM system planning and control

• The products to be manufactured are selected based on their relative fit and economy
• In the medium term the products are determined in the master production schedule
• In the short term the job selection can be formulated as an optimization problem with the following constraints
• Tool set-up is such that the jobs can be executed
• There is sufficient work (loaded pallets) for the unmanned shifts
• Workpiece loading and fixture changes can be done during the manned shifts
• System control is based on prioritizing, where the AS/RS system serves machines based on priorities set to jobs and machines
• Operator selects jobs to be loaded on pallets from the job queue and calls for pallet and material to the loading station
• Unloading has its own job queue

4. How does a linear regression model work? What is the meaning of coefficients in the model and how can their importance and contribution to the model output be evaluated?

A linear regression model is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. Here's how it works and how the coefficients play a role in the model:

How Linear Regression Works:

Modeling Relationship: The model assumes that the dependent variable (often denoted as Y) can be explained or predicted by a linear combination of independent variables (denoted as X1, X2, X3, ..., Xn).

Fitting the Model: The linear equation typically has the form Y = b0 + b1X1 + b2X2 + ... + bn*Xn + e, where b0 is the intercept term, b1 to bn are the coefficients of the independent variables, and e is the error term.

Least Squares: The method of least squares is commonly used to find the values of the coefficients that minimize the sum of the squared differences between the observed values and the values predicted by the model.

Meaning of Coefficients:

Intercept (b0): This is the predicted value of Y when all the X variables are 0. It's the point where the regression line crosses the Y axis.

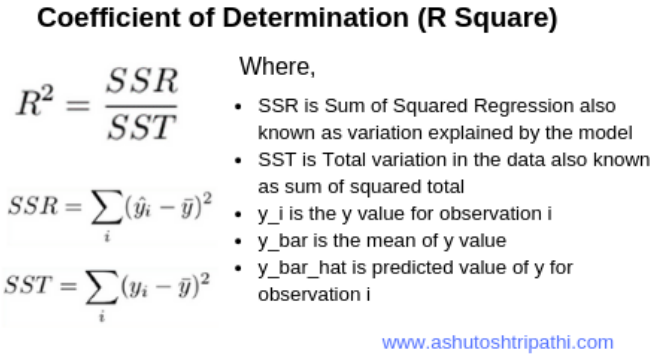
Slope Coefficients (b1, b2, ..., bn): These represent the change in the dependent variable for a one-unit change in the independent variable, assuming all other variables are held constant.

Evaluating Model Output and Coefficients Importance:

Coefficient Significance: Statistical tests, like the t-test, can be used to determine whether the coefficients are significantly different from zero

R-squared (R^2): This statistic measures the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher R-squared value indicates a better fit of the model to the data.

Adjusted R-squared: It adjusts the R-squared value based on the number of predictors in the model and the number of observations, preventing overestimation of the model



2. Explain how linear regression modeling works. What is the function of the constants in the model and how can their effects and significances be evaluated? Why is such evaluation useful? How are the values of the constants determined?

Evaluating the Effects and Significance:

The effects and significance of the coefficients are typically evaluated using hypothesis testing with t-tests. The null hypothesis states that the coefficient is equal to zero (no effect). A p-value is calculated for each coefficient to determine whether to reject the null hypothesis. If the p-value is less than the significance level (usually 0.05), the coefficient is statistically significant.

Determining the Values of the Constants:

The values of the coefficients are determined using a method called Ordinary Least Squares (OLS). OLS estimates the coefficients by minimizing the sum of the squares of the residuals (the differences between the observed values and the values predicted by the model). In essence, OLS finds the best-fitting straight line through the data points by adjusting the coefficients to minimize the discrepancy between the predicted and actual data points.

By doing so, the regression model allows for the making of predictions or inferences about the relationship between the independent and dependent variables. It is widely used in various fields for trend analysis, forecasting, and inferential statistics.

Write full sentences using clear and readable fonts

1. A factory manufactures 1000 products a year starting from raw materials. Average throughput time from start to completion is 5 weeks and manufacturing cost is about 100 000 euros/product.
    a) What is the value of average work in process?
    b) What is the average number of products found at the factory (incomplete or complete) at any time?

a) The value of average work in process for the factory is approximately 9,615,385 euros.

b) The average number of products (incomplete or complete) found at the factory at any time is approximately 96.15, which we can round to 96 products considering that you can't have a fraction of a product.

These calculations are based on Little's Law, which states that the long-term average number of items in a queuing system L (work-in-process) is equal to the long-term average effective arrival rate, λ (annual output), times the average time that an item spends in the system, W (throughput time). Here, the arrival rate is the annual output of the factory, and the average time in the system is converted from weeks to a year for the calculation

Python code

```
# Given data
annual_output = 1000  # products per year
throughput_time_weeks = 5  # weeks from start to completion
cost_per_product = 100000  # euros per product

# Calculations

# Little's Law: L = λ * W
# L is the average number of items in the system (work-in-process)
# λ is the arrival rate of items to the system (annual output in this case)
# W is the average time an item spends in the system (throughput time)

# Since we're dealing with a year and the output is annual, we convert throughput time to a yearly basis
```

```
weeks_per_year = 52
throughput_time_yearly = throughput_time_weeks / weeks_per_year

# Calculate the average number of items in the system using Little's Law
average_items_in_system = annual_output * throughput_time_yearly

# To calculate the value of average work in process, we multiply the average items by the cost per product
value_of_average_work_in_process = average_items_in_system * cost_per_product

average_items_in_system, value_of_average_work_in_process
```

Write full sentences using clear and readable fonts

1. A factory manufactures 200 products a year starting from raw materials. Average throughput time from start to completion is 4 weeks and manufacturing cost is about 100 000 euros/product.
    a) What is the value of average work in process?
    b) What is the average number of products found at the factory (incomplete or complete) at any time?

(Duplicate from above assignment, just different numbers).

```
# Re-defining the given data for the new problem after the reset
annual_output = 200  # products per year
throughput_time_weeks = 4  # weeks from start to completion
cost_per_product = 100000  # euros per product

# Calculations using Little's Law as before

# Convert throughput time to a yearly basis
weeks_per_year = 52
throughput_time_yearly = throughput_time_weeks / weeks_per_year

# Calculate the average number of items in the system using Little's Law
average_items_in_system = annual_output * throughput_time_yearly

# Calculate the value of average work in process
value_of_average_work_in_process = average_items_in_system * cost_per_product

average_items_in_system, value_of_average_work_in_process
```

a) The value of average work in process for the factory is approximately 1,538,462 euros.

b) The average number of products (incomplete or complete) found at the factory at any time is approximately 15.38, which we can round to 15 products considering that you can't have a fraction of a product.

These figures were determined using Little's Law, which relates the number of items in the system (L), the arrival rate (λ), and the average time an item spends in the system (W). The arrival rate is the annual output of the factory, and the throughput time is converted from weeks to a year for the calculation

Write full sentences using clear and readable fonts – you may answer in Finnish or Swedish, too.

1. How can capacity of a manufacturing resource in short or medium term be adjusted in practice? How can capacity planning be formulated as an optimization model? What could be the objective and which constraints would be needed? You do not have to write the equations, textual description is sufficient.

Capacity planning
• In planning of production any further than immediate near future one has to use (inaccurate) estimates
• Unit of workload in planning is usually working hour or average product
• Timing resolution is rough, usually day, week or month
• The objective is to match predicted work load with capacity at minimum cost
• In modeling and reality decision variables concerning capacity are, depending on local circumstances
• Work force adjustment
• Hiring and firing
• Overtime
• Personnel leasing
• Subcontracting
• Manufacturing to stock
• All measures involve cost aspects and limitations on quantity

• In its basic form Capacity planning is easy to formulate as an optimization model
• Time proceeds in discrete steps determined by detail level of planning
• Here we manufacture average products and examine the whole factory as a single resource
• Notation, parameters:
t = time period index, t = 1,..,T
D_t = demand during period t
B = production (products) per worker on a time period
C = cost of one worker for a time period
O = overtime cost of one worker for a full time period
P = Hiring cost of a worker
E = Firing cost of a worker
H = inventory holding cost of a product for a time period

Capacity planning

- Decision variables:
  - q_t = production amount on time period t
  - o_t = overtime done in worker input on time period t
  - p_t = workers hired in the beginning of time period t
  - e_t = workers fired in the beginning of period t
  - W_t = Number of workers on period t; intermediate result, not a real decision variable
  - I_t = stock level on period t; intermediate result
- Optimization model:

$$\text{Min } c = \sum_{\forall t} \left(CW_t + Oo_t + Pp_t + Ee_t + HI_t \right)$$

$s.t.$

$$W_t = W_{t-1} + p_t - e_t, \quad \forall t$$

$$I_t = I_{t-1} + q_t - D_t, \quad \forall t$$

$$q_t \le (W_t + o_t)B, \quad \forall t$$

$$q_t, o_t, p_t, e_t, I_t \ge 0, \quad \forall t$$

Make-to-stock production - MTS

- Make-to-stock (MTS) production is used to manufacture parts or products in batches and store them in stock to wait for need
- An example could be a screw manufacturer, that makes screws to stock and delivers them to customers when orders arrive
- The previous dynamic lot sizing formulations are appended with capacity constraints
- We add more products and later more departments to the system
- One-time costs are usually related to set-ups in manufacturing
- In MTS production problems related to timing are usually easier than in E/MTO production
- This is because there often is no immediate need for the product and scheduling is more flexible

Aggregate planning in MTO-production

- In MTO-production batch formation is not usually a relevant problem, because product variation is large, which in fact is the reason for MTO-production in the firs place
- Product structures and production processes are typically complicated
- Orders appear randomly and resource load is fluctuating
- Pressure for short delivery times is severe
- Leveling of resource load is important for economical reasons
- Levelling affects resource utilization rate and need of capacity and reduces marginal cost due to overtime work, and emergency subcontracting
- Aggregate planning systems appear as MRP (Manufacturing Resource Planning) functions in ERP (Enterprise Resource Planning systems

Aggregate planning in MTO-production

- Purpose of aggregate planning is to schedule orders based on experience, present status and forecasts so that
  - The ability to take new orders is good
  - Resource loading matches capacity
  - Delays are minimized
  - Work in process is kept at a low level
  - Need for rescheduling is kept minimal
- MTO aggregate planning system's basic inputs:
  - Order release dates
  - Order due dates
  - Material delivery times
  - Resource capacities
  - Orders' resource loading (work contents)
  - Process precedence requirements (process flow)

Aggregate planning in MTO-production

- System user sees the status of the system when allocating new orders and knows basic specifications of the orders
- In addition an order forecast may be available
- Two extreme cases are *forward planning*, in which orders are scheduled as early as possible, and *backward planning*, in which orders are started as late as possible taking capacity constraints and release and due dates into consideration
- Slack and possibilities to affect timing are the bigger the bigger the difference between shortest possible delivery time and due date is and the more capacity is available
- In practice room for optimization exist, because only some orders are taken with minimum delivery time

MTO aggregate planning strategies

- **Forward planning** provides the best ability to take new orders, but work in process is maximized and need for rescheduling may be considerable
- On the other hand, reserving capacity for rush orders is important, because these orders are taken at best price and make your customers happy
- In **backward planning** orders may be lost for lack of capacity
- But, in MTO production customers often make changes to orders afterwards. These changes are the easier to make the later manufacturing takes place

- In practice one has to use sales forecasts and adjust timing tactics accordingly
- Rescheduling causes confusion, errors and generates indirect administrative costs

MTO aggregate planning optimization

Optimization criteria

- In the following models the main optimization criteria are resource levelling or peak load minimization
- The idea is that in practice production needs to deliver products as promised to customers and therefore due dates are constraints that can not be violated and capacity has to be flexible
- Tardiness minimization is also a common optimization criterion with capacity taken as a constraint
- Both am. criteria can be combined in multi-objective optimization, where penalties are determined to resource load limit and due date violations

Optimization model

- Because different kinds of orders load resources at different amounts and at different times in relation to finishing time, the problem is difficult to solve manually
- As a linear optimization model it works well