How many?

Risk-based decision making for MRO inventory

Mark Horton, September 2014

Why take risks at all?

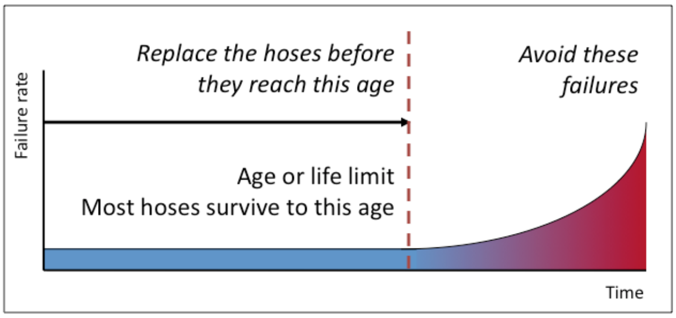
Because, whether you like it or not, you have no choice.

You may want to believe that every inventory demand should be predictable. If you read the last paper in this series, *No surprises*, perhaps you are optimistic that new maintenance technologies will help. I have bad news for you: they won't. They may help with some decisions, but they won't solve all of your problems.

It is true that some MRO inventory demands can be predicted. Take a simple example: if hydraulic hoses last for five years, then your maintenance system probably includes a task to replace them when they reach the end of their life. That task will demand replacement hoses from the store, on schedule, every five years. You can plan to order parts in advance of the demand, and the only risks you take are those associated with the supply chain: will you be able to get the right parts into stock before they are needed?

This demand can be predicted because the underlying failure—deterioration of the hoses—has a characteristic minimum life. If the hoses are left for longer, then eventually the material of which they are made will degrade and develop leaks. The obvious maintenance response is to replace the hoses before they start to deteriorate.

"This response was fundamentally flawed"



There are many other types of failure that have the same characteristics, specifically those associated with mechanisms such as corrosion, erosion, fatigue and chemical changes. At one time, the majority of industrial maintenance was based on fixed interval replacement of parts, and a great deal of time and effort was spent in trying to identify each part's "life". However, research from civil aviation in the 1960s and 1970s revealed that this response was fundamentally flawed for a number of reasons.

- The vast majority of failures show no relationship at all between age and failure rate, so fixed interval replacement does nothing to improve reliability
- Maintenance intervention can be positively harmful because it can introduce new failures
- Even where an age or life limit can be identified, some failures can occur before the limit is reached
- Fixed-interval discard tasks replace parts without taking account of their condition at the time, so any remaining useful life that they have is wasted

The aviation industry's response to this research was to introduce a rigorous process for selecting maintenance tasks to ensure that they dealt effectively with the consequences of each failure. Reliability-centred Maintenance (RCM) has been used since the 1970s to establish maintenance schedules for aircraft, and over the past three decades it has also become the technique of choice in all asset-intensive industries.

"Great news for the maintenance team can bring unwelcome surprises for MRO inventory"

RCM has revolutionised maintenance by ensuring that every equipment failure is appropriately managed. The results have been dramatic: maintenance costs have been reduced, but uptime has increased, safety improved, and environmental compliance has been easier to ensure. This is great news for the maintenance team; however, if you are responsible for supporting the MRO inventory it can bring unwelcome surprises.

New Maintenance Techniques

In the old world of fixed-interval part replacement, the inventory planner's job could sometimes be simple: just ensure that parts are available when the task is scheduled. In today's world, new maintenance techniques have completely changed the way in which we manage failures, and the planner's job has become far more complex.

RCM works because it recognises an awkward truth about failures: that most of them don't happen after some convenient and easily defined life. In fact, the original study by United Airlines showed that only 11% of the failure modes surveyed had failure rates that increased with age, implying that nearly nine out of ten failures cannot not be effectively managed by fixed interval replacement. It isn't just that scheduled discard is too expensive: it is technically and absolutely the wrong thing to do for these failures. The best that it can achieve is to leave their reliability unchanged; but in often it could actually make their reliability worse through maintenance-induced failure.

Modern maintenance schedules recognise this reality, and they manage failures in four different ways in addition to fixed-interval overhaul and replacement.

Condition-based maintenance	If there are detectable signs that a failure is developing before it happens, it may be possible to monitor the equipment's condition and only replace parts just before they fail.
Failure-finding	Protective devices such as alarms and trips can fail at any time without any signs that they would not work. Schedules include formal failure-finding tasks to test devices at regular intervals.
Do nothing (corrective)	"Do nothing" is a positive decision in RCM. It means that either there is no task that could detect or prevent the failure, or that the task is so expensive that leaving the failure to happen is a better option.
Redesign or one-time action	Recognises that maintenance cannot achieve the required performance, so the equipment or the way in which it is used needs to be changed.

The inventory planner's problem is obvious: from the stores viewpoint, the shift from fixed interval replacement to on-condition and "do nothing" maintenance turns a comfortable and predictable cycle of orders into an erratic and apparently random pattern of demands.

No surprises looked at strategies for managing maintenance inventories that allow "just in time" part ordering even in a world of unpredictable demand, but what should we do when those strategies don't apply?

Risk: how much do you want?

To satisfy random, unplanned demands, we have to hold stocks of the part. But working out the details from there can be far from easy.

- How many should be held?
- Should we hold stock on site or share stock with other sites?
- Would vendor stocking be cheaper than local stock and as effective?
- Do we reorder a part if one is used?
- How should we run down stocks at the end of equipment life?

Let's start by trying to answer the first question: how many parts should we hold, and what principles and tools can help us to make the decision robust and defensible?

Consequences

One of the reasons for RCM's success is that it does not try to prevent every failure. It only chooses a maintenance task if it is technically feasible (it can prevent or predict the failure) and *worth doing*.

"Worth doing" means that the task deals with the consequences of the failure. This is important because the emphasis is not on preventing failures, but on managing what happens when the failure occurs.

The maintenance review group begins by working out what would happen if nothing at all were done to prevent the failure, the *failure effects*. Using the description of failure effects, the RCM decision diagram categorises the failure consequences:

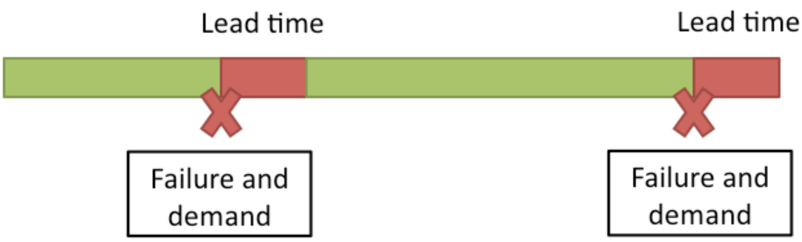
Hidden	There are no effects at all if the failure occurs under normal circumstances. This category applies primarily to protective systems such as alarms, trips and other emergency systems, when the objective is to manage the possible multiple failure
Safety and environmental	Failures which have a direct effect on safety (killing or injuring someone) or on the environment
Operational	Failures that cause economic loss (typically equipment downtime costs) in addition to the direct cost of failure.
Non-operational	The only effect is the cost of repairing the failure.

A maintenance task is only selected if it deals successfully with the consequences of the failure. For operational and non-operational failures, it means that a failure will be allowed to happen if the cost of carrying out a maintenance task to prevent it is greater than the cost of the failure itself. For example, if an inspection task could predict failures of a pump seal, but carrying out the inspection would cost more than letting the failure happen, RCM would select *No scheduled maintenance*: do nothing to prevent the failure.

What does this have to do with stock levels?

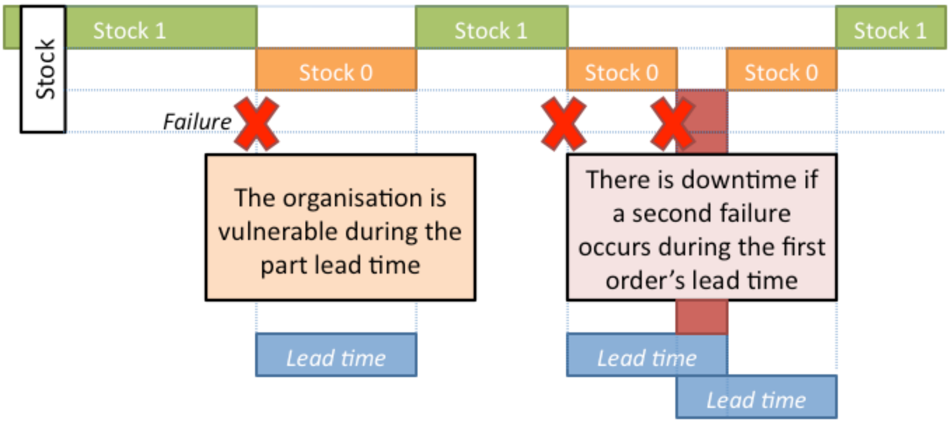
Whether or not your organisation has used a formal maintenance strategy development method like RCM, its core principles can be used to develop a robust MRO inventory policy.

First imagine that the store contains no engineering spare parts at all. A part fails and its equipment is taken out of operation. A demand is made for the part and a replacement is ordered, but the equipment remains out of service until it is delivered and fitted.



The period of enforced downtime—days, weeks or even months—is almost always unacceptable unless the organisation has standby equipment or some other way to take over the equipment's functions. The motivation for holding part stocks is to try to manage the risk of downtime by providing an immediate replacement part when it is needed.

If failures occur at random, or if the part is needed to support more than one equipment, it is possible that a second failure could occur while the store is waiting for the replacement spare to be delivered. Once again it would experience downtime until a part was delivered.



As more parts are added to stock, the probability of unfulfilled demands goes down and the expected downtime is reduced. So increasing stock reduces risk, but the key question is: how many parts are enough?

Different Types of Risk

The key question is: how many parts are enough?

Effective management of stock levels means balancing two factors. Higher stock levels reduce risk but increase cost; low stock levels reduce cost but increase risk. How is it possible to decide what level of risk is appropriate and what cost is appropriate for each item?

In exactly the same way as failures have different effects, it is possible for stock outages to cause a range of consequences from trivial to catastrophic. The effects fall into one of three broad categories.

Reduced protection	Downtime leads to reduced system availability
Safety or environmental	The stock out causes an event (or prevents the organisation from successfully managing an event) that affects safety or environmental integrity
Economic	Lack of stock leads to monetary loss

Reduced Protection

Lack of stock leads to *reduced protection* if the associated protected equipment continues to operate until parts are available to repair the protective device. Stock outs in this category usually involve parts used to support protective devices such as alarms, trips and standby equipment.

> A switch triggers an alarm if the level in a liquid storage tank drops too low. An independent alarm system shuts down the system if flow stops completely, so if there were no spare available to replace a failed switch, the organisation might accept the increased risk of a shutdown and continue to operate the tank until a spare was available.
>
> A routine test finds that a standby pump's motor has failed in a cooling water supply system and no spare is in stock. The organisation continues to run the system without a standby pump until the motor is available, accepting the risk that the duty pump might fail and shut down the process.

A stock out belongs in this category if:

1. The stock out has no direct downtime effects
2. The associated equipment would continue to operate until a spare part is available

Usually this category is only relevant if it is safe and operationally acceptable to continue running equipment without the associated protective device. That may be possible if there is standby equipment available, or if there are other devices that provide acceptable levels of protection. In many cases (perhaps most) the organisation prefers to shut down the equipment, or it may have no choice because of potential safety risks.

> A scheduled test discovers that a turbine overspeed device would not be able to trip the turbine successfully. Spare parts are not available, and the organisation shuts down the turbine until replacement parts are delivered.

> A vessel's high pressure trip switch is found to have failed and no spare part is available. The owner decides to shut down the process until the trip switch can be replaced.

In these cases, the stock out would be classified as *economic* because the asset's owner suffers extended downtime.

Safety or Environmental

A stock out is categorised as having *safety or environmental* consequences if:

- It could cause someone to be injured or killed; or
- It could lead to breaching an environmental regulation

Stockout events that belong in this category are extremely rare, but genuine examples fall into two main categories.

Spare parts are needed to manage an emergency	Drums of fire fighting foam Submarine escape oxygen candles Oil well "kill" fluids
Protected equipment cannot be shut down or made safe after a failure	Spare parts for a spacecraft's oxygen regeneration system[1]

It is important not to think about equipment failure here, but about the subsequent stock out event. Although the initial failure may have had widespread and disastrous effects, the stock out has no further impact on safety or the environment.

> A leaking valve caused about 250 barrels of crude oil to leak from an offshore oil platform into the sea. No spare valve was available, so production was shut down until a spare part could be obtained.

> A 14kV transformer failed during a storm and started a fire which destroyed a neighbouring store room. Fortunately no one was injured. No spare transformer was immediately available and production unit was shut down until an emergency power supply could be arranged.

In both examples above, the initial failure had environmental or safety consequences, but the stock out extended the downtime period. As a result, the stock out is not categorised as *safety or environmental*.

Economic

Stock shortages can cost money in several different ways.

---

[1] In practice the chance of serious problems is remote; spacecraft such as the International Space Station carry enough emergency oxygen to supply the crew through just about any conceivable contingency.

- Direct loss of a product or service, with a direct impact on profit
- Penalties as a result of not meeting customer contract performance requirements
- The cost of obtaining emergency parts
- Loss of customer confidence and damaged reputation

For economic consequences, holding higher stocks costs more in terms of holding costs, administration and maintenance in the store, but higher stock levels reduce the chance of incurring risked costs due to stock shortages. The core principle of stock optimisation is to find the number of spare parts that incurs the lowest overall cost to the business.

The overwhelming majority of stock decisions involve a balance between the cost of holding spare parts and the risked cost of stock outages. In the remainder of this paper, we will look in detail at these decisions.

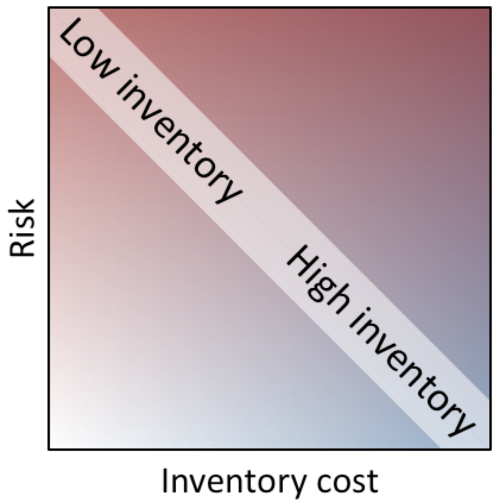
Point of Balance

"Holding and downtime costs are not the same type of expenditure"

Before going any further, we need to be clear about one thing: holding costs and downtime costs may both be measured in pounds or dollars or euros, but they are not the same type of expenditure.

Holding spares gives rise to definite costs: the cost of buying spare parts, the cost of maintaining and storing them, and replacement costs if they have a finite shelf life. These are all real, definite costs. Your business will pay for them with real money from a real bank account. As a result, they are usually fairly easy to estimate. There may be some uncertainty, but if you buy the spares, you will incur the associated costs.

The consequences of *not* having spares are different because your business only spends money if a spare is needed and the store doesn't have one. The consequences may be trivial or they may be severe; demands may be frequent or rare. Although part mean time between failure (MTBF) statistics can be used to work out the average rate of downtime losses, it is impossible to know exactly what will happen in practice, because the actual number of demands is random.



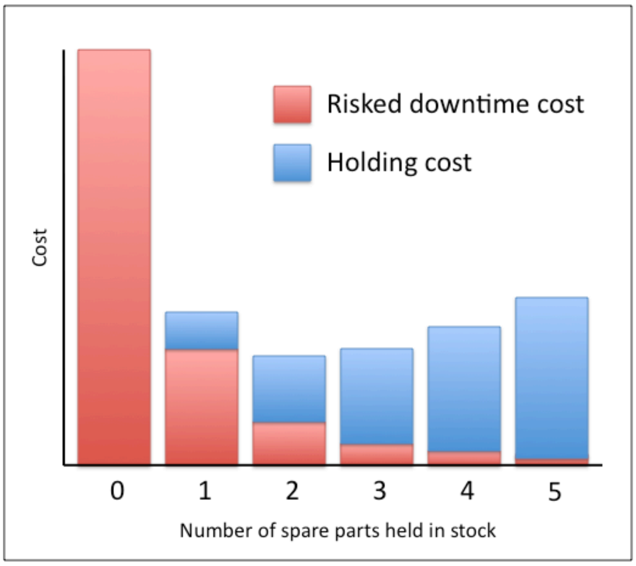
Setting a stock level means choosing a point of balance: more stock means higher costs but less risk; holding fewer parts means lower purchase and holding costs, but higher risk. The challenge is to understand the risk and to choose the right balance point.

In this paper we are going to consider stock shortages that result in extended equipment downtime and associated costs. The reason for this is that the vast majority of stock decisions are cost-based. Even if the initial failure has safety or environmental consequences, extended downtime waiting for spare parts costs money, so the decision comes down to a balance of costs.

Spare Parts as an Investment

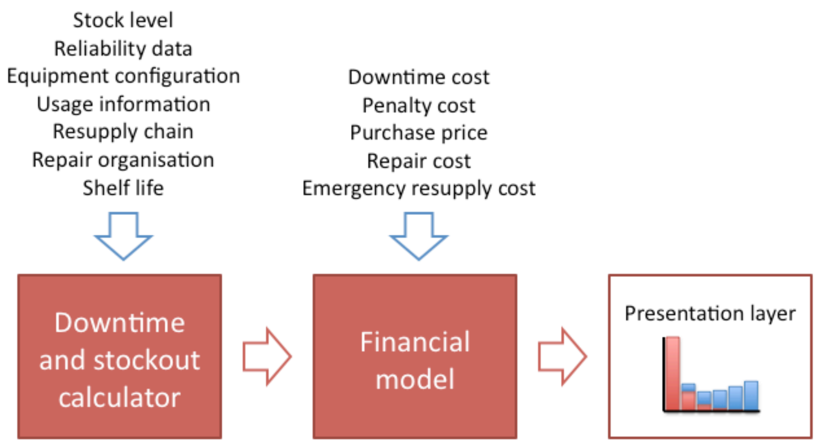
Calculating a part's optimum inventory seems an easy problem.

- Calculate the inventory holding cost per year
- Calculate the risked downtime cost per year
- Add the two costs together
- Repeat for all reasonable inventory levels
- Choose the inventory that results in the lowest overall cost



In very simple cases it is possible to write the total cost as a mathematical expression that could be implemented in a spreadsheet. Unfortunately the calculation becomes far more complex when dealing with realistic situations, and there is no option but to use a computer-based model.

A computer-based spare parts optimiser consists of three parts: the core reliability model, a financial model, and the presentation layer.



Core Model

At the lowest level there is a core model that uses demand data, equipment configuration, lead time and repair characteristics to calculate statistics including expected downtime, stock out rate and others.

The model can be as simple or complex as is needed to deal with the stores organisation, equipment and part characteristics. It might be a simple model for one part supporting one item of equipment; it could be a sophisticated statistical program to deal with multiple items of equipment served by a number of stores.

Financial Model

The financial modeller assigns costs to downtime, purchases, repairs and expired shelf life based on the statistics produced by the core model.

The output of the financial model may include:

- The total amount to be spent on buying new spare parts (if inventory is to increased) or the amount received through selling off existing inventory (for a decrease in inventory)
- The rate of expenditure on buying replacement items during equipment life
- The rate of expenditure on repairing parts
- The cost of replacing life-expired parts
- The cost of administration and maintenance of parts in the store

Presentation Layer

The presentation layer consolidates financial data for a range of spares holdings, finds the optimum inventory level and displays supporting information. It may also enable the user to explore the sensitivity of decisions to uncertainty in the analysis data and to answer "what if?" questions that reflect possible future trends.

Traditional Cost Models

"Traditional models are unrealistic"

During the design of the MRO Analytics software, we looked in detail at how to balance holding costs with risk. One of our conclusions was that the traditional models used by most optimisers are unrealistic.

Traditional cost-based spare parts optimisation is based on a balance of holding costs against risked costs. "Holding cost" is a catch-all term for a number of types of expenditure including the following.

- The cost of money tied up in on-the-shelf inventory that otherwise could have been used elsewhere in the business to generate increased profits
- A proportion of the overall cost of the warehouse
- A proportion of the cost of inventory personnel
- Part of the cost of the organisation's information infrastructure, purchasing and stores systems
- For some parts, the cost of carrying out maintenance while the part is in stock

The holding cost is usually expressed as percentage of the parts' purchase price per year as demonstrated in the example below.

> A chemical company is reviewing its stock of a mixer unit bearing. A new bearing costs $15000.
>
> The holding cost has been assessed as: 12% of the part's value as money tied up; 8% of its value to cover inventory personnel, heat, light and other warehouse costs; 2% for

maintenance while in stock; and 2% to include general administration, IT and other overheads.

The total holding cost is 24% of the part's value, or $3600 per year.

The holding cost is then calculated for each inventory level.

Inventory	Inventory Cost	Holding Cost/year
0	$0	$0
1	$15000	$3600
2	$30000	$7200
3	$45000	$10800
4	$60000	$14400
5	$75000	$18000

The core calculator uses reliability data to work out the expected equipment downtime per year. Details of the downtime calculation have been omitted here.

Inventory	Inventory Cost	Holding Cost/year	Downtime hr/yr
0	$0	$0	1240
1	$15000	$3600	110
2	$30000	$7200	5.97
3	$45000	$10800	0.24
4	$60000	$14400	0.008
5	$75000	$18000	0.00028

Downtime cost is calculated for each spares holding. For the mixer unit, the cost of downtime is $1000 per hour.

Inventory	Inventory Cost	Holding Cost/year	Downtime hr/yr	Downtime cost/year
0	$0	$0	1240	$1.24M
1	$15000	$3600	110	$110k
2	$30000	$7200	5.97	$5970
3	$45000	$10800	0.24	$240
4	$60000	$14400	0.008	$8
5	$75000	$18000	0.00028	$0.28

Finally the holding cost and downtime cost per year are added together to give an overall cost to the business.

Inventory	Holding Cost/year	Downtime cost/year	Total cost/year
0	$0	$1.24M	$1.24M
1	$3600	$110k	$113600
2	$7200	$5970	$13170
3	$10800	$240	$11040
4	$14400	$8	$14408
5	$18000	$0.28	$18000.28

The lowest cost to this business is obtained by holding three spare bearings.

Limitations of the Traditional Model

What is the problem if the traditional model seems to give a reasonable answer? The fundamental issue is that it asks the wrong question.

Current Stock Level

When someone decides to review inventory levels, it is not a theoretical exercise carried out in isolation from reality. The question being asked depends on several factors, the most basic of which is determined by the reason for the review.

- This is a new spare part. How many should we buy?
- We have a large stock of these items and the stock turnover is negligible. Should we dispose of some (or all) of the stock?
- Our operations recently suffered nearly a week's downtime because there was no spare item in stock. What stock should we be holding?

Here is the first problem: the traditional model takes no account of the *current* stock level, which may too high, or too low. The organisation is not interested in the optimum stock level; we want to know the right investment or disposal *given the current stock*.

Equipment Life

Any industrial equipment has a finite lifetime; at the end of its life, any spare parts that are held to support the equipment become obsolete as well.

The traditional holding cost optimisation assumes that the store will continue to incur holding costs into the far future. If you work out the optimum spares stock holding at 24 years into an equipment's 25 year life, the answer is the same as it was when it was first commissioned. Taking this to an extreme, the traditional approach still gives the same answer on the day before the equipment is due to be removed. The result for the business is potentially serious: when equipment reaches the end of its life, the store room contains a full set of near-worthless spare parts.

This approach is contrary to common sense. As equipment approaches the end of its life, any spare parts that are used should not be reordered, so that stocks run down to zero as gracefully as possible. Holding a full set of spares for obsolete equipment just wastes money that should be used elsewhere in the business.

Investment and Disposal

Buying spare parts is an investment of money now to buy protection against expensive downtime in the future. Reducing inventory can produce some income and a reduction in overheads, but at the expense of higher risked downtime.

The traditional approach treats investment and disposal as the exact inverse of each other: if adding a spare part increases holding costs by 24% of its value, then disposing of the same part reduces costs by the same amount.

Experience shows that principle is almost never true.

- Adding a spare part costs the full price at the time of purchase, not the holding cost every year
- Part disposal generates far less value than their initial cost. If a part is ordered and then returned the day after delivery, the vendor will probably apply a substantial restocking fee. The value of a part that has been in stock for more than a very short period is probably be no more than a small proportion of its purchase price.
- Removing a single part from stock reduces administrative and other overheads by a negligible incremental amount

Holding costs are therefore a poor reflection of the real business problem.

The Investment Model

"The decision is treated as an investment"

The model developed for the MRO Analytics inventory optimiser tries to reflect the real business decision.

- The financial model takes into account any existing stock when recommending an updated stock level rather than calculating an "optimum" level that is always based on zero stock
- The decision is treated as an investment: money spent or received now is compared with the change in risked downtime cost
- Disposals reflect the true value of redundant stock
- Risked costs, administrative costs and maintenance in stores are assessed over the remaining equipment life

Existing Stock and Investment

The immediate cost of a change in stock is assessed by calculating the difference between the proposed stock and the current stock level. If the stock level increases, the business must spend money buying new parts; if stock decreases, the business may receive revenue by selling unwanted parts. This is an immediate cash flow.

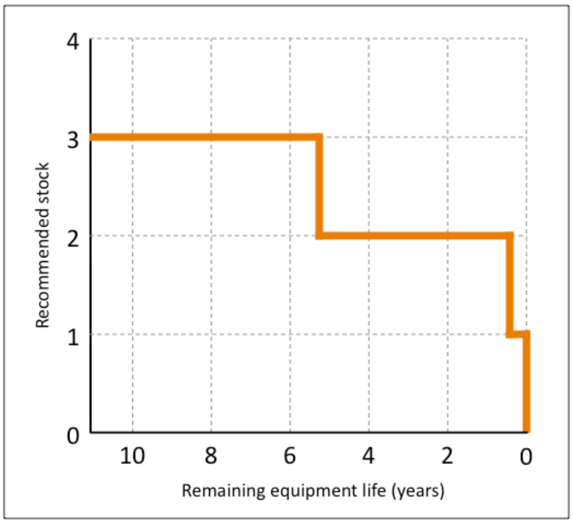
Through-Life Cost

A number of costs recur throughout the time that parts are held in stock:

- Purchases to replace failed parts
- The cost of refurbishing repairable parts
- The cost of replacing time-expired parts that have reached the end of their shelf life
- The cost of administration and maintenance in stores
- Risked downtime costs

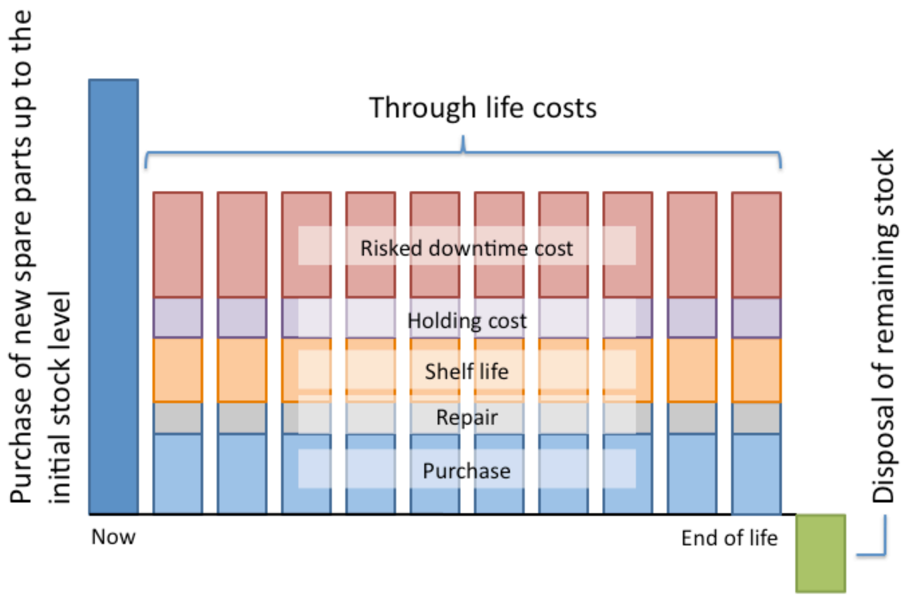
Finite Life

The model is applied over a fixed period of time, usually the full lifetime of the equipment that the parts support. The recommended stock level depends on the equipment's life: long life means a higher stock recommendation, but the model tends to recommend reducing stocks as the end of equipment life approaches.

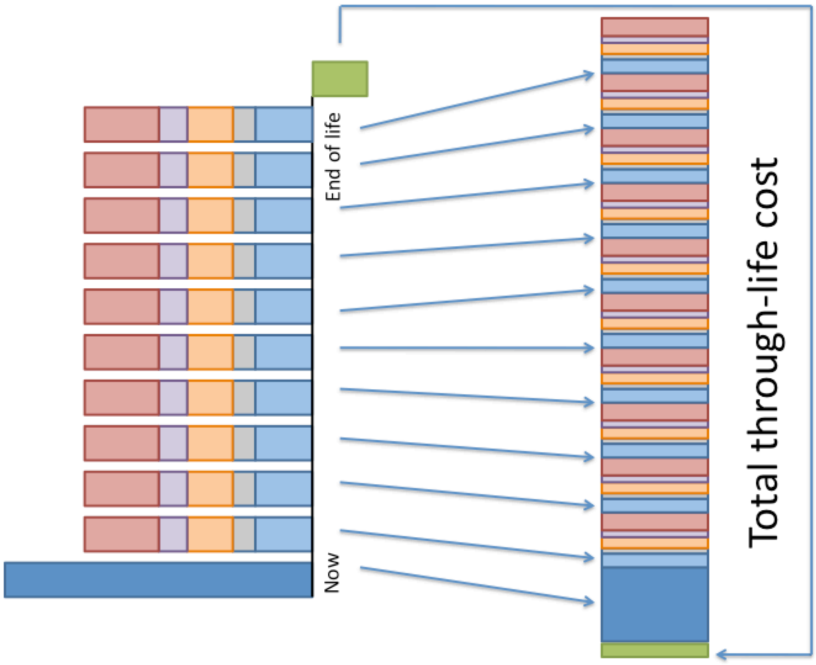


Total Investment Value

Recurring expenditure is calculated for each period during the equipment's lifetime. Then initial costs, recurring expenditure and final disposal value are brought together to give a single value for each spare part holding.



At this point, the total represents the sum of all income and expenditure over the stock's lifetime.



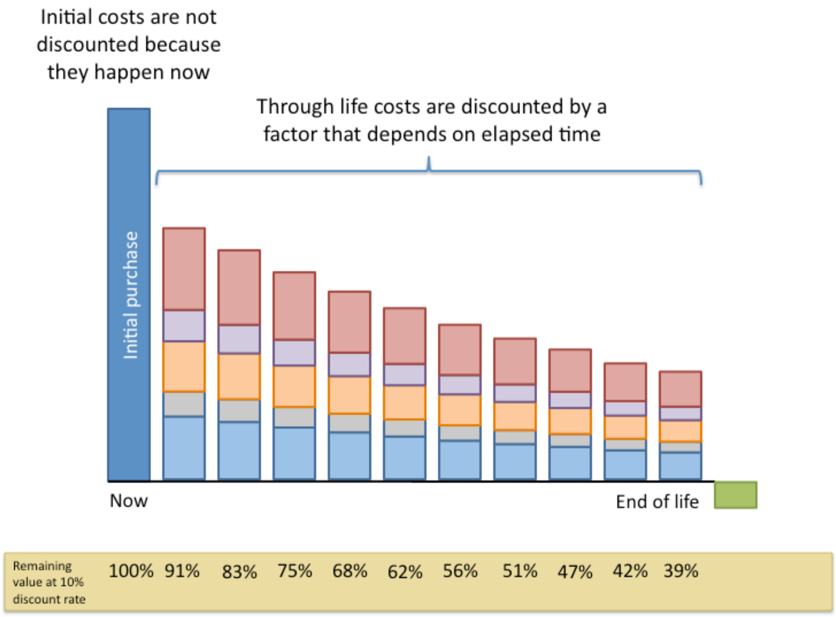
Discounting Expenditure

There is one final step in the model: recognising the time value of money. So far, expenditure or income in year 20 of the model has been given exactly the same weight as the same expenditure today, but that isn't how we evaluate most investments.

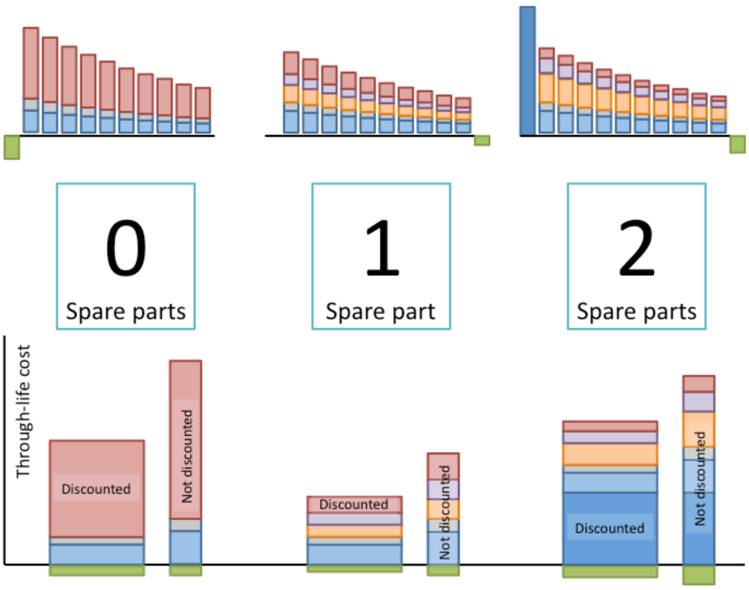
Most long-term business investment decisions involve evaluating future returns on an investment that is made now. Given a choice, the business would prefer to receive income earlier rather than later; even if inflation is removed from the decision, a million dollars received now or in a few months' time is preferred to the same amount in ten years' time. On the other hand, we would choose to postpone expenditure as far as possible, so spending a million pounds ten years from now is preferred over the same expenditure today.

This preference for early income and deferred expenditure is modelled by applying a *discount rate* to any amounts that are spent or received in the future. A discount rate of 10% per year means that $1100 received in a year's time is "worth" the same as $1000 now, and that we would treat $1210 spent two years from now in the same way as spending $1000 immediately.

The final calculation in the financial model is to discount all future income and expenditure.



Then the discounted values are added together to give a single figure: the *net present value* (*NPV*) for that stock level[2].



The optimiser calculates the overall net present value for each possible stock level; the recommended holding is the one that produces the highest NPV overall. Most of the values that contribute to the NPV are expenditure (negative income), so the NPV is usually negative; the optimum stock level is the one that gives the least negative value.

Dealing with Risk

How well does the investment model deal with MRO inventory decisions?

---

[2] Unlike the total cost without discounting, this number does not represent real money in the sense of being the sum of all the cash moving into and out of the business bank account. The net present value is just a number that is used to choose the best business decision.

It answers the real question

First, and most importantly, it answers the real question: should I use the company's hard-earned money to buy spare parts when it could otherwise be invested somewhere else instead? Probably more important is that it answers the question using real information: purchase costs, risked downtime costs, demand, repair costs and so on. Every one of these numbers can be justified, and the yearly discounting factor is the same one that would be used for any other investment decision. Taken together, it means that money spent on critical spare parts can be ranked alongside alternative investments elsewhere in the business.

It can sometimes be possible to make the traditional model produce similar results by choosing the holding cost carefully. And right there is the problem: "holding cost" is something slightly arbitrary, a percentage of the purchase price that can be varied, perhaps to give a recommendation that the analyst thinks is acceptable. The traditional model is to some extent subjective; the investment model is robust.

Disposals are rational

The use of a substantial holding cost means disposals tend to be overvalued. In the traditional model, removing a spare part from stock is the exact inverse of adding a part. The change is just the addition or subtraction of its associated holding cost, which is often between 15% and 30% of the part's purchase price.

Treating buying a new spare part and disposing of one as exact opposites is irrational: we know from experience that the average return on obsolete stock is no more than a few percent of its value. The investment model treats disposals more realistically. It recommends reducing stock only if the risk of increased future downtime is offset by the expected disposal income.

End of life without a full stock room

Evaluating inventory costs over the life of the equipment that it supports makes the model dynamic. As the remaining life diminishes and risked costs decrease, the recommended level of spare parts will eventually drop; even if stock is not immediately sold, opportunities can be taken not to replace any parts that are used. By contrast, the traditional model has an infinite horizon, and the recommended level of spares today is the same as it will be on the day before the supported equipment shuts down.

Conclusion

Choosing the right inventory level for an MRO part depends on finding the right balance between inventory costs and risked downtime costs. Using net present value in an investment model to compare possible decisions provides a robust framework for business decisions, together with the flexibility to model inventory levels accurately as equipment becomes obsolescent.

Terms of use and Copyright