

PERSPECTIVE · A COMPANION TO THE TRILLION DOLLAR FORECAST SERIES

Mining for Gold at the Intersection of Push and Pull

How to use forecast-driven push and demand-driven pull together, each for the job it does best

A perspective for Boards, CFOs, and Senior Supply Chain Leaders



Why this perspective

Our earlier papers, the Trillion Dollar Forecast series, made a clear case. Forecast-driven MRP is built on assumptions that are no longer true. A pull-based operating model works better. It replaces what actual consumption uses. It uses buffers to absorb variation. It frees up working capital, improves service, and ends the daily firefight.

That case still stands. But a strong case can be read the wrong way. Some readers decide that the forecast is the enemy, and that pull is always the answer. It is not. Every supply chain has decisions that pull cannot make, because the consumption signal does not exist yet. This paper is about those decisions. It shows the value you gain when you use push and pull together, on purpose.



Erik Bush

EVP – Internal Operations, Algo

1. The Wrong Debate

For decades, planners have argued the same point. One side says: plan ahead. Forecast demand, build inventory, and run the schedule from the plan. The other side says: wait and react. Wait for the consumption signal, replace what was used, and let buffers cover the rest.

This is the push-versus-pull debate. It is also forecast-driven versus demand-driven, and MRP versus DDMRP. Each side can point to the other side's failures. Push gives you the forecast that whipsawed the plant. Pull gives you the empty buffer when the once-a-year order arrives. Each side is right about the other and wrong about itself.

Both methods work, and both fail, but they do not fail in the same places. They are two tools that answer two different questions, and neither one fights for the whole supply chain. So the useful question is where the gold lies, and the answer is the seam between push and pull, where each is used only where it is stronger.

2. What Each Method Is For

Pull answers a question about what you can see.

When a part is used, a unit is sold, or a case is picked, that is real demand. It is the one signal that cannot be wrong. Pull replaces what was used. When demand varies and the forecast is weak, nothing beats pacing supply to what really happened. Pull is the engine of a healthy supply chain. It should run most of it, most of the time.

Push answers a question about what you cannot see yet.

Some decisions come before any consumption signal exists. A new product has never shipped, so it sends no signal. New capacity can take eighteen months to install, so it cannot wait for the order that needs it. A pre-buy ahead of a tariff or a price rise commits before any order arrives. Each of these reaches for demand that has left no trace. Only a forecast can guide them. This is the one thing push does that pull cannot.

So every decision comes down to one question: which tool fits here?

//

**"Pull replaces what has already been used. Push prepares for what has not.
The mistake is to make one tool do both jobs."**

3. The Forecast Prepares and Tests

Here is the main idea of this paper. The forecast has two jobs: to prepare, and to test the operating model. It should never drive replenishment.

MRP makes the first mistake. It lets the forecast drive replenishment. It creates the purchase, production, and distribution orders from a number that is wrong at the item level, and more wrong the finer you cut it. This is push doing pull's job. It is the cause of the excess stock, the poor service, and the firefight.

Pull's supporters can make the opposite mistake. They throw the forecast away. That works until you reach the decisions consumption cannot see: the new product, the long-lead capacity, the build before the shutdown. Those decisions then get made badly, or not at all.

The value sits between the two mistakes. Use pull for every decision that a reliable consumption signal can pace. Use the forecast only where no such signal exists yet, and only to prepare and to test, never to drive replenishment. One question settles the old debate for any decision in the chain.

//

"Is there a reliable consumption signal at the horizon of this decision? If yes, pull. If no, push to prepare, then hand the decision to pull as soon as the signal arrives."

4. Where Forecast and Pull Work Together

If the forecast does not drive replenishment, where does it belong? Many assume it belongs at the front of the flow. People often say that long-lead items are pushed to a plan and short-lead items are pulled. That mixes up lead time with method, and it is wrong. An item with proven demand is buffered and pulled, however long its lead time. In fact, a long lead time is the main reason to use a buffer. So the forecast's place does not depend on where the item sits in the flow. It depends on one thing: whether a reliable consumption signal exists yet. That gives the forecast two clear jobs.

The forecast sizes the buffer.

The most common case is a repeating season. When demand rises and falls on a known yearly pattern, use the forecast daily usage to size the buffer rather than the past average. Raise the buffer for the busy season, then lower it again as the season ends. Service stays high through the peak, and you do not strand stock in the quiet months. A planned promotion works the same way. It is a short, planned rise in demand. You raise the buffer before the promotion and lower it after. In both cases, the forecast sizes the expected rise. It does not create the orders that draw the buffer down. Replenishment stays driven by consumption. The planner sets a demand rate; the planner does not chase single numbers.

Some events are too sharp for a buffer to absorb after the fact. A one-time spike (Black Friday, a launch tie-in, or one large order) packs weeks of demand into a day. This is well inside the replenishment lead time. Pull cannot catch up: by the time you see the demand, the event is over. Here the forecast does more. You build ahead to a planned event quantity. You build the stock up before the spike, sell it down during the event, then return to normal. This is the clearest case of forecast-based preparation, short of a new product. Like a new product, it is limited, planned, and handed back to pull as soon as the event ends.

The forecast prepares for the launch.

This is the job pull cannot do at all. A new product has no sales history, so there is nothing to pace against. A buffer sized from zero demand is only a guess. The only sound way to prepare for a launch is a forecast. You build it from similar products, from shared product features, and from informed judgment about the launch, the channel, and the promotion behind it. Then, as real sales come in, the launch forecast matters less and the consumption signal matters more. The item moves to pull.

Every vendor sells a launch forecast. The value is in the handover you design: the point at which an item moves from forecast to consumption, and, at end of life, back to a planned run-down. Most companies never design this. They either keep the launch forecast too long and carry the excess for a year, or they trust a weak early signal and run out of stock.

5. When the Signal Exists But Is Wrong

The launch case has no signal yet. The opposite case has a signal, but you cannot trust the forecast of it. The automotive supplier is the clearest example. A tier supplier gets EDI releases from its carmaker customers. Each release has a near-term call-off and a forward forecast. That forward forecast is very unstable, and by many accounts it has grown worse over time.

Carmakers often forecast on the high side, and their programs keep changing. If the supplier treats these releases as a firm plan, it pulls the carmaker's swings straight into its own plant. It then passes them on to its own suppliers. The result is the bullwhip: torn-up schedules every week, expedite freight as the norm, and stock that is both too high and too low at once.

The fix is to change the forecast from driver to advisor. Actual releases and actual use drive replenishment, through buffers sized for the demand rate and its variation. The plant runs to what is really used. You keep the forward forecast, but you do not use it to create orders. You use it to test whether the model can absorb what the forecast implies.

6. How to Stress-Test the Model

The goal of Demand-Driven S&OP (DDS&OP) is to use forecasts to validate that the operating model can support the variation and volatility that may come over the tactical horizon. You run the forecast against the operating model to answer one question: will the current buffers and capacity hold through what is coming? A baseline forecast is the “most likely” case. But how do you stress-test the operating model?

Hold the demand level and vary the noise. Buffers are sized for the variation in demand, not only for the demand rate. So the test should introduce levels of variation while holding the level of demand constant. Our method sets a target rate of expected demand. Then it applies a range of probabilities for how noisy and erratic the signal is. It tests, at the confidence level you need, whether the buffers absorb the likely swings around the expected rate. This is a test the operating model can pass or fail on its own terms.

This is also what lets you take your hands off the wheel. If the model absorbs the modeled noise at the confidence you need, and delivers the service, inventory, and capacity targets you set, the team can run to consumption day to day and step in only by exception. They have proof that the model will handle the swings that are likely to come. The forecast has done its real job. It does not tell the supply chain what to make. It shows that the operating model is ready for what may be coming.



“The useful test keeps the demand level and studies the noise around the rate you really expect.”

The decision	Reliable consumption signal?	Tool	The forecast's job
Replace a mature, actively selling item	Yes. It exists at the horizon.	Pull	None. Consumption sets the pace.
Replace a long-lead item with proven demand	Yes. The buffer covers the lead time.	Pull (buffered)	None. The buffer is there to cover the long lead time.
Launch a product with no history	Not yet.	Push, then hand to pull	Prepares for the launch (similar products, features, judgment).
Size buffers for a season or promotion	A change in demand rate is expected.	Forecast sizes, pull replaces	Sets the buffer from forecast daily usage; consumption still refills it.
Prepare for a one-time spike (Black Friday)	The event is too sharp for pull to catch.	Push (build-ahead), then pull	Sizes a build-ahead for the event; real sales draw it down.
Invest in long-lead capacity	No order-specific signal at the horizon.	Push	Prepares the commitment ahead of demand.
Run a plant against unstable carmaker releases	It exists, but the forecast of it is wrong.	Pull	Tests whether the buffers are adequate; flags big rate changes.

One question decides each: is there a reliable consumption signal at this horizon? The answer selects the tool.

7. Designing the Combination

None of this happens on its own. Left alone, a supply chain collects push and pull habits; it does not combine them. The forecast drives replenishment because it always has. A buffer keeps a size that no one has revisited since it was first set. A launch forecast stays in place because no one set a date to remove it.

To design the combination, you name a few things clearly: which decisions the forecast sizes or prepares for, and which ones consumption paces; how the handover is triggered; what the forecast may do (prepare, size buffers, and test); and what it may not do (create the replenishment orders that belong to pull). You also set the DDS&OP cadence that tests and tunes the model.

Do this, and pull delivers the working capital, the service, and the calm the series promised. Push, kept to the decisions only it can make, delivers launches that are planned rather than guessed, capacity that is ready rather than late, and seasonal moves that are staged rather than rushed. The gold is a clear division of labor, so each tool does only what it does best.

8. Other Decisions in the Seam

The seam between push and pull is wider than the demand-side cases in this paper. The same test applies every time: is there a reliable consumption signal at the horizon of this decision? Where there is not, or where a known change has not yet reached demand, the forecast prepares, sizes, or tests, and pull paces the actual draw. The same test applies in these places too.

Supply-side seasonality and known constraints. The forecast is not only a demand signal. Known supply events (factory shutdowns, hurricane season, tariff or allocation changes) are just as predictable. Prepare buffers ahead of the event; pull runs normally the rest of the year.

New-supplier and dual-sourcing onboarding. A new supplier has no delivery history, so you cannot size its lead-time buffer from data. Prepare the buffer from an analog supplier, then move to pull as real reliability data accrues.

Product transitions (phase-in and phase-out). The cutover between an old item and its replacement is a seam of its own. The forecast prepares the changeover (build-ahead of the new item, run-down of the old, and the cannibalization between them), while pull governs steady state on each side.

Capacity, labor, and shift planning. Rough-cut capacity and seasonal labor are prepared from the forecast, including the sprint-capacity buffer. Pull then schedules to actual demand within that envelope.

Contract structure: flex and commitments. Flex agreements, minimum and maximum volumes, capacity reservations, and vendor-managed inventory are the seam made contractual. The forecast prepares the commitment envelope; pull calls off within it.

Machine learning and demand sensing. Better forecasting never earns the right to drive orders. That is the old MRP mistake. It earns its place by sharpening the two real jobs: better buffer sizing, earlier detection of a real step change, and a better noise model for the stress test.

Each of these is the same idea in a new place. Prepare for what the signal cannot yet show, and let consumption pace the rest.

About this perspective

This paper is a companion to the four-paper Trillion Dollar Forecast series on supply chain operating model design. That series argued for a shift from forecast-driven MRP to a pull-based operating model. This paper covers the decisions pull cannot make alone. It shows how to combine forecast-driven preparation with demand-driven execution by design rather than by habit.