

Does AI-Related Activity Improve Forecasts of US Industrial Production?

MJ Shin

October 20, 2026

1 Question

Investment in artificial intelligence (AI) has become a visible driver of US high-tech production, equipment orders, and electricity demand. This note asks a narrow, testable question. Do indicators of AI-related activity contain information about future US industrial production (IP) beyond what standard macroeconomic predictors already capture? And if they do, has that information appeared only recently, and under which economic conditions does it matter most?

I answer it with a pseudo-out-of-sample forecasting exercise. The exercise compares benchmark models with the same models augmented by an “AI block,” and evaluates them with tests designed for nested models, for conditional predictive ability, and for predictive ability that varies over time.

2 Target variable

Let IP_t be the level of the industrial production index in month t . The monthly annualized growth rate is

$$g_t = 1200 \Delta \ln IP_t, \quad (1)$$

and the h -month-ahead target is average annualized growth between t and $t + h$:

$$y_{t+h}^{(h)} = \frac{1200}{h} \ln \left(\frac{IP_{t+h}}{IP_t} \right), \quad h \in \{1, 3, 6, 12\}. \quad (2)$$

I use two targets: total IP (FRED: INDPRO), and IP excluding selected high-technology industries and motor vehicles (IPX5VHT2S). Semiconductors and computers are components of headline IP, so part of any gain on the total index is mechanical. A gain on the ex-high-tech index instead points to spillovers from AI-related activity to the rest of industry.

3 Forecasting models

All forecasts are *direct*: a separate regression is estimated for each horizon h . Each model takes the form

$$y_{t+h}^{(h)} = \alpha + \sum_{\ell=0}^{p-1} \phi_{\ell} g_{t-\ell} + \beta' \mathbf{F}_t + \gamma' \mathbf{z}_t + \varepsilon_{t+h}, \quad (3)$$

where $\mathbf{F}_t$ is a vector of k macroeconomic factors and $\mathbf{z}_t$ is the AI block. The models differ only in which regressors they include. AR is an autoregression in IP growth alone. DI is the diffusion-index

model of Stock and Watson (2002), which adds to the autoregression a small number of factors that summarize a large panel of macroeconomic series (Section 3.1). The suffix “+AI” means the AI block $\mathbf{z}_t$ is added to the model.

Model	Regressors	Role
AR	lags of g_t	simple benchmark
AR+AI	lags of $g_t, \mathbf{z}_t$	augmented
DI	lags of $g_t, \mathbf{F}_t$	main benchmark
DI+AI	lags of $g_t, \mathbf{F}_t, \mathbf{z}_t$	headline augmented model
DI+ z_i	lags of $g_t, \mathbf{F}_t, z_{i,t}$	one indicator at a time

Every augmented model nests its benchmark: setting $\boldsymbol{\gamma} = \mathbf{0}$ recovers the benchmark. The DI benchmark matters most. An AI indicator that beats a univariate AR may simply be standing in for the business cycle, so the interesting question is whether it adds information beyond a broad diffusion index. The baseline sets $p = 3$ and $k = 4$; Section 3.2 explains these choices.

3.1 Diffusion-index factors

The factors follow Stock and Watson (2002) and are estimated from the FRED-MD panel of McCracken and Ng (2016). Let $\mathbf{X}_t$ be the N -vector of FRED-MD series after each series’ recommended stationarity transformation. At each forecast origin, and using only observations inside the estimation window:

1. Keep the series that are fully observed in the window. Treat an observation as an outlier if $|X_{it} - \text{med}(X_i)| > 10 \cdot \text{IQR}(X_i)$, and set it to the series mean.
2. Standardize each series to mean zero and unit variance, giving the $n \times N$ matrix $\mathbf{Z}$.
3. Take the singular value decomposition $\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}'$. The factors are the first k principal components, $\hat{\mathbf{F}} = \mathbf{U}_k \mathbf{S}_k$.

Because the factors are re-estimated at every origin, the forecaster never uses information from after the origin.

3.2 The AI block

Each AI indicator $x_{i,t}$ enters as a three-month change, which smooths monthly noise. Production and orders series enter as annualized log changes,

$$z_{i,t} = \frac{1200}{K} \ln \left(\frac{x_{i,t}}{x_{i,t-K}} \right), \quad K = 3, \quad (4)$$

and capacity utilization enters as a level difference, $z_{i,t} = x_{i,t} - x_{i,t-3}$.

FRED id	Transform	Description
IPHITEK2S	$\Delta_3 \ln$	IP: high-technology industries
IPG3344S	$\Delta_3 \ln$	IP: semiconductors and other electronic components
CAPUTLG3344S	Δ_3	Capacity utilization: semiconductors
IPG2211S	$\Delta_3 \ln$	IP: electric power generation, transmission and distribution (proxy for data-center load)
A34SNO	$\Delta_3 \ln$	New orders: computers and electronic products (Census M3)
NEWORDER	$\Delta_3 \ln$	New orders: nondefense capital goods excluding aircraft (Census M3)

Some indicators are published later than IP. When an indicator is missing at the origin (the “ragged edge”), its last available value is carried forward, for at most two months.

Specification choices. The lag order $p = 3$, the number of factors $k = 4$ and the change horizon $K = 3$ were fixed before any results were computed and were not tuned on the evaluation sample. Three lags span one quarter of monthly data, a common choice for monthly IP. Four factors is a moderate number. Stock and Watson (2002) find that a few factors capture most of the predictive content of large panels for real activity. For the full FRED-MD sample, the PCp2 criterion of Bai and Ng (2002) selects eight factors (McCracken and Ng, 2016), though the number it selects varies across subsamples. The three-month change smooths the substantial month-to-month noise in the orders and sector production series, and it matches the quarterly span of the lags. Stock and Watson (2002) instead choose the lag order and the number of factors by information criteria at each origin. Fixing them in advance has one advantage: the headline results cannot reflect a search over specifications. The cost is that the results may depend on these particular values. One concern deserves emphasis. If four factors omit business-cycle information that a richer factor model would capture, the AI block can appear useful merely by filling that gap.

4 Out-of-sample design

Estimation scheme. Let T index forecast origins, running from T_0 (January 2005 in the baseline) to the last origin for which $y_{T+h}^{(h)}$ is observed. At each origin, model (3) is estimated by OLS on a *rolling* window of the $R = 120$ most recent usable observations,

$$t \in \{T - h - R + 1, \dots, T - h\}, \quad (5)$$

and the forecast is $\hat{y}_{T+h|T}^{(h)} = \hat{\alpha}_T + \sum_{\ell} \hat{\phi}_{\ell,T} g_{T-\ell} + \hat{\beta}'_T \hat{\mathbf{F}}_T + \hat{\gamma}'_T \mathbf{z}_T$. This yields P out-of-sample forecasts for each horizon. I use a rolling rather than an expanding window because the conditional and fluctuation tests below are derived under limited-memory estimation, which keeps estimation error from vanishing asymptotically.

Information sets. In the *pseudo-real-time* version, today’s revised data are truncated at each origin. In the *real-time* version, the forecaster at origin T sees the target and the factor panel only as published in that month’s FRED-MD vintage, and forecasts are evaluated against the latest vintage. This design is real-time only in part. The AI indicators and the conditioning variables enter at their latest (revised) vintage, which favors the AI block, so the real-time results are an upper bound on its value. The ex-high-tech target is not in FRED-MD, so the real-time exercise covers total IP only.

Data retrieval. All inputs are downloaded automatically. FRED series come from FRED’s public CSV endpoint. The current FRED-MD panel is the newest dated monthly file on the St. Louis Fed site; the undated `current.csv` link was found to lag by a year and is not used. The real-time exercise uses all 326 FRED-MD vintages from August 1999 to September 2026, assembled from the two historical archives (1999–2014 and 2015–2024) and the monthly files from 2025 onward. No API key is required.

5 Evaluation

Index the two competing models by B (benchmark) and A (augmented). The forecast errors are $e_{T+h}^j = y_{T+h}^{(h)} - \hat{y}_{T+h|T}^{(h),j}$ for $j \in \{A, B\}$, and the squared-error loss differential is

$$\Delta L_{T+h} = (e_{T+h}^B)^2 - (e_{T+h}^A)^2, \quad (6)$$

which is positive when the AI-augmented model forecasts better. In what follows, $\hat{\sigma}^2(\cdot)$ is a Newey–West long-run variance estimator with $h - 1$ lags, which accounts for the $\text{MA}(h - 1)$ structure of overlapping h -step errors.

5.1 Relative accuracy

$$\text{MSFE ratio} = \frac{\sum_T (e_{T+h}^A)^2}{\sum_T (e_{T+h}^B)^2}, \quad (7)$$

where a value below one means the AI block lowers mean squared forecast error.

5.2 Clark–West test (unconditional, nested)

Under the null that $\gamma = \mathbf{0}$, the augmented model estimates parameters whose population values are zero. That noise inflates its MSFE, so the Diebold–Mariano statistic is biased against the larger model and has a non-standard distribution. Clark and West (2007) correct for this noise term:

$$\hat{f}_{T+h} = (e_{T+h}^B)^2 - \left[(e_{T+h}^A)^2 - (\hat{y}_{T+h|T}^B - \hat{y}_{T+h|T}^A)^2 \right], \quad CW = \frac{\bar{f}}{\sqrt{\hat{\sigma}^2(\hat{f})/P}}. \quad (8)$$

H_0 : equal population MSFE; H_1 : the augmented model is more accurate. The test is one-sided against standard normal critical values. For reference I also report the Diebold and Mariano (1995) statistic with the small-sample correction of Harvey et al. (1997), while noting that it is conservative for nested models.

5.3 Giacomini–White test (conditional predictive ability)

The unconditional tests ask whether the AI block helps *on average*. Giacomini and White (2006) instead ask whether past information can predict *when* it helps. Let $\mathcal{G}_T$ be the information available at the origin and let $\mathbf{h}_T \in \mathcal{G}_T$ be a q -vector of test functions. The null of equal conditional predictive ability is

$$H_0 : \mathbb{E}[\Delta L_{T+h} \mid \mathcal{G}_T] = 0 \implies \mathbb{E}[\mathbf{h}_T \Delta L_{T+h}] = \mathbf{0}. \quad (9)$$

With $\mathbf{Z}_{T+h} = \mathbf{h}_T \Delta L_{T+h}$ and $\bar{\mathbf{Z}} = P^{-1} \sum_T \mathbf{Z}_{T+h}$, the statistic is

$$GW = P \bar{\mathbf{Z}}' \hat{\Omega}^{-1} \bar{\mathbf{Z}} \xrightarrow{d} \chi_q^2, \quad (10)$$

where $\hat{\Omega}$ is a Newey–West estimator of the long-run variance of $\mathbf{Z}_{T+h}$ with $h - 1$ lags. If the null is rejected, the regression $\Delta L_{T+h} = \boldsymbol{\delta}' \mathbf{h}_T + u_{T+h}$ gives a decision rule: use the augmented model whenever $\boldsymbol{\delta}' \mathbf{h}_T > 0$. I report the share of origins at which this rule selects the AI model.

Test functions. The baseline set is $\mathbf{h}_T = (1, \Delta L_{\tau(T)}, s_T)'$. Here $\Delta L_{\tau(T)}$ is the most recent loss differential already *realized* at the origin, meaning the largest target date $\tau + h \leq T$, and s_T is a state variable. Every state variable is observed at the origin; ex post objects such as NBER recession dates are excluded.

Variable	Source	Interpretation
$\Delta L_{\tau(T)}$	internal	persistence of relative performance
$400 \ln(IP_T/IP_{T-3})$	internal (real time)	own business-cycle state
NFCI	Chicago Fed (NFCI)	financial conditions (latest vintage; revised)
Term spread	T10Y3M	recession risk
Credit spread	BAA10Y	default risk, credit cycle
VIX, EPU	VIXCLS, USEPUINDXM	financial and policy uncertainty
NASDAQ momentum	NASDAQCOM	state of the AI/technology equity cycle

Each state variable is first tested on its own, as $(1, s_T)$, and then jointly with the others. Keeping q small preserves power.

Reading the results. Two points matter for interpretation. First, a rejection says that the *sign* of $\mathbb{E}[\Delta L_{T+h} | \mathcal{G}_T]$ is predictable, not that the augmented model is better on average: the GW null can be rejected even when the MSFE ratio exceeds one, in which case the decision rule selects the AI model only in a minority of states. The reported share of origins makes this explicit. Second, about six instrument sets are tried for each comparison and horizon, so an isolated p -value near 0.01–0.02 should be discounted for multiple testing. A Bonferroni bound over six sets, for example, requires $p < 0.0083$ for a 5% family-wise level.

5.4 Fluctuation test (time-varying predictive ability)

Strong AI-related capex is a recent development. Over a sample beginning in 2005, any predictive content may therefore be concentrated at the end, where full-sample tests have little power to detect it. The fluctuation test of Giacomini and Rossi (2010) tracks relative performance over rolling windows of m out-of-sample forecasts:

$$\mathcal{F}_{j,m} = \hat{\sigma}^{-1} m^{-1/2} \sum_{s=j-m+1}^j \hat{f}_{s+h}, \quad j = m, \dots, P, \quad (11)$$

with $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{f})$ estimated over the full out-of-sample period. The null of equal performance *at every point in time* is rejected when

$$\max_j \mathcal{F}_{j,m} > k_\alpha(\mu), \quad \mu = m/P, \quad (12)$$

where $k_\alpha(\mu)$ is tabulated in Giacomini and Rossi (2010). The baseline sets $\mu = 0.3$ and $\alpha = 5\%$, one-sided, which gives $k \approx 2.77$.

Checking the critical values. Under the null, $\mathcal{F}_{j,m}$ converges to $\sup_{\tau \in [\mu, 1]} [B(\tau) - B(\tau - \mu)] / \sqrt{\mu}$, where B is a standard Brownian motion. I verified the tabulated values by simulating this functional (40,000 replications). On a grid of $P = 260$ points, matching the sample size here, the simulated 5% values agree with the table to within 0.04 for every μ (at $\mu = 0.3$: 2.75 one-sided and 3.01 two-sided, against 2.77 and 3.01 tabulated). On a fine grid ($P = 4000$) the values are 0.03–0.09

higher (2.83 one-sided at $\mu = 0.3$), because the supremum is taken over more points. The table is therefore appropriate for samples of this size. Every rejection reported below also exceeds the fine-grid value, so the conclusions do not depend on which version is used.

I apply the test to the Clark–West adjusted differential $\hat{f}$ so that it is consistent with the nested comparison; this use should be read as approximate, since the original critical values are derived for non-nested comparisons. The date at which $\mathcal{F}_{j,m}$ first crosses $k_\alpha(\mu)$ estimates when the AI block began to carry predictive information. Because each window looks backward, that date refers to the window’s end.

5.5 Supplementary diagnostics

The cumulative sum of squared-error differences, $\text{CSSED}_J = \sum_{T \leq J} \Delta L_{T+h}$, gives a visual record of when the augmented model gains or loses ground. To ask which individual indicators survive, the single-indicator models $\text{DI}+z_i$ can be ranked with the Model Confidence Set of Hansen et al. (2011).

6 Preliminary results

The results below are *pseudo-real-time*. They use the revised data available in October 2026 and the September 2026 FRED-MD file, both targets (total IP and IP excluding high-tech industries and motor vehicles), the baseline settings ($p = 3$, $k = 4$, $R = 120$, $\mu = 0.3$), and forecast origins from January 2005 to the last month with a realized outcome ($P = 249$ to 260). Panel C repeats the total-IP exercise in real time (described below).

h	AR vs. AR+AI		DI vs. DI+AI		
	MSFE ratio	CW p	MSFE ratio	CW p	$\max_j \mathcal{F}_{j,m}$
<i>A. Total IP (INDPRO)</i>					
1	0.946	0.038	1.030	0.287	1.53
3	0.806	0.087	0.845	0.025	2.74
6	0.881	0.082	1.030	0.086	2.64
12	0.923	0.145	1.094	0.205	2.52
<i>B. IP excluding high-tech and motor vehicles (IPX5VHT2S)</i>					
1	0.949	0.019	1.017	0.097	2.53
3	0.882	0.028	0.934	0.004	3.65*
6	1.025	0.094	1.142	0.293	2.59
12	1.137	0.353	1.325	0.838	1.58
<i>C. Total IP, real-time target and factors (AI block revised)</i>					
1	0.972	0.038	1.021	0.171	2.45
3	0.807	0.046	0.851	0.003	2.86*
6	0.877	0.036	0.995	0.058	2.90*
12	0.951	0.102	1.008	0.071	3.09*

* Exceeds the one-sided 5% critical value of about 2.77.

The paragraphs below first describe the total-IP results (Panel A), then what changes for the ex-high-tech target (Panel B) and in real time (Panel C).

On average. Against the AR benchmark, the AI block lowers MSFE at every horizon, by 5–19%. Against the diffusion index, the only clear gain is at $h = 3$: MSFE falls by 15.5% and Clark–West

rejects at 5% ($p = 0.025$). At $h = 1$ and $h = 12$ the AI block adds nothing beyond the factors. So much of its advantage over the AR comes from proxying the business cycle.

Which indicators. Among single indicators added to DI, capital-goods new orders (NEWORDER) is the strongest at $h = 3$ (CW $p = 0.001$, MSFE ratio 0.966), followed by semiconductor production (IPG3344S, $p = 0.021$). The other indicators, including the electricity-load proxy, are insignificant or harmful at every horizon.

When. The fluctuation test does not reject for the full block DI+AI at any horizon; the largest statistic, 2.74 at $h = 3$, is just below the critical value of 2.77. It rejects for NEWORDER at $h = 3$ (maximum 5.10, first crossing for the window ending October 2020) and at $h = 6$ (3.30, from October 2025), and for semiconductor capacity utilization at $h = 12$ (2.96, from August 2015). The $h = 3$ crossing in late 2020 falls in a window that contains the pandemic collapse and rebound of IP. A few very large forecast errors could drive it, but the subsample check below shows that the short-horizon gain survives when 2020–21 is excluded. Only the $h = 6$ crossing for NEWORDER falls in the period of heavy AI investment.

Under what conditions. For DI vs. DI+AI, Giacomini–White rejects only at $h = 12$: with the term spread as instrument ($p = 0.013$) and with all instruments jointly ($p = 0.019$). The unconditional MSFE ratio at this horizon is 1.094, and the decision rule selects the AI model at only 31% of origins. The AI block therefore helps in a predictable minority of yield-curve states and hurts otherwise. Given the number of instrument sets tried, this evidence is suggestive rather than conclusive.

Ex-high-tech target. Removing semiconductors, computers, communications equipment and motor vehicles from the target rules out a mechanical gain, yet the three-month result survives and becomes statistically stronger. Against DI, the AI block lowers MSFE by 6.6% (smaller than the 15.5% for total IP), but Clark–West rejects at $p = 0.004$. The fluctuation test now rejects for the full block (3.65), first crossing in the window ending February 2021. The gain again comes mainly from capital-goods orders (NEWORDER: CW $p < 0.001$, fluctuation maximum 5.14). Semiconductor production (IPG3344S) also helps ($p = 0.030$), and since semiconductors are excluded from this target, that gain cannot be mechanical. However, its fluctuation statistic first crosses in late 2014, well before the recent AI investment cycle. At $h = 6$ and $h = 12$ the AI block is clearly harmful, raising MSFE by 14% and 33%, which points to over-fitting six extra regressors in a 120-month window at long horizons. Giacomini–White again rejects at $h = 12$ (term spread, $p = 0.039$; all instruments, $p = 0.023$), but the decision rule selects the AI model at no more than 5% of origins. Here the instruments predict when the AI model will do *worse*, not when it helps. As for total IP, most fluctuation crossings fall in windows ending in 2020–21 and so contain the pandemic; only NEWORDER at $h = 6$ crosses within the recent period (January 2026).

Real time. Panel C uses, at each origin, the FRED-MD vintage published in that month for both the target and the factors, and evaluates against the September 2026 vintage. As described in Section 4, the AI indicators remain today’s revised series. This mix favors the AI block, so Panel C is an upper bound on its real-time value. At seven origins (2013-09, 2018-12, 2019-01 and 2025-11 to 2026-02) the newly published vintage lacked too many series for the factors to be estimated. These months follow the federal government shutdowns of 2013, 2018–19 and 2025, and they drop out of the DI comparisons.

The three-month gain is essentially unchanged in real time (MSFE ratio 0.851, CW $p = 0.003$), and the fluctuation test now rejects for the full block at $h = 3, 6$ and 12. The main difference is in which indicators matter. In real time the AI-specific series carry much more of the signal: semiconductor production (IPG3344S: CW $p = 0.002$ at $h = 3$, $p = 0.022$ at $h = 6$) and high-tech production (IPHITEK2S: $p = 0.018$ and $p = 0.004$). A plausible reading is that these series help most when the macroeconomic factors are estimated from unrevised data, so their value is understated

in pseudo-real-time tests. Their fluctuation statistics, however, first cross in 2011, so they look like a long-standing leading indicator from the electronics cycle rather than a signal that emerged with recent AI investment. Giacomini–White rejects with the term spread at $h = 3, 6$ and 12 ($p = 0.033, 0.044, 0.034$), with the rule selecting the AI model at 100%, 55% and 52% of origins. Given the multiple instrument sets, these rejections are suggestive only.

Excluding the pandemic. To check whether 2020–21 drives these results, the pseudo-real-time exercise is repeated on two subsamples that leave it out of the evaluation. The first covers forecasts made from January 2005 to February 2019, so that even the 12-month targets end before March 2020 ($P = 170$). The second covers forecasts made from July 2021 on ($P = 51$ to 62). The models are still estimated on rolling 120-month windows, so the post-2021 coefficients are partly estimated on 2020 data.

h	Forecasts made 2005-01 to 2019-02				Forecasts made 2021-07 on			
	Total IP		Ex-high-tech		Total IP		Ex-high-tech	
	ratio	CW p	ratio	CW p	ratio	CW p	ratio	CW p
1	0.923*	<0.001	1.027*	0.071	0.943*	0.005	0.953*	0.008
3	0.910*	0.006	1.050	0.053	0.884*	0.001	0.884*	<0.001
6	0.957*	0.043	1.196	0.250	1.146	0.109	1.369	0.179
12	1.099	0.153	1.446	0.752	1.247	0.674	1.310	0.752

DI vs. DI+AI, pseudo-real time. Ratio is the MSFE ratio; * the fluctuation test rejects at 5%.

The short-horizon gain is not a pandemic artifact. For total IP, DI+AI beats DI at $h = 1$ and $h = 3$ in both subsamples (CW $p \leq 0.006$), and the fluctuation test rejects in both. For the ex-high-tech target the same holds after 2021, but not before 2020, when the full block is no better than DI.

What changes between the subsamples is which indicators carry the signal. Before 2020 it comes from semiconductor and high-tech production (IPG3344S: CW $p = 0.002$ at $h = 1$ and 0.027 at $h = 3$; IPHITEK2S: 0.001 and 0.012). For the ex-high-tech target, semiconductor production alone helps ($p = 0.008$ and 0.009) even though the full block does not. The fluctuation crossings in this subsample cluster in 2009, in windows that contain the Great Recession. After 2021 the two production series add nothing ($p \geq 0.20$). The gain comes instead from capital-goods orders (NEWORDER: $p = 0.011$ at $h = 3$ and 0.035 at $h = 6$ for total IP; $p = 0.004$ at $h = 3$ for the ex-high-tech target). The full block’s fluctuation statistic first crosses in mid- to late 2025, the first crossings in this note dated within the AI investment boom. With only 51–62 forecasts, the post-2021 tests have little power, and the Giacomini–White statistics at $h = 12$ (based on as few as 39 observations with up to six instruments) are unreliable and are not interpreted.

Summary. The evidence that AI-related activity improves IP forecasts beyond a broad factor model is clearest at short horizons of one to three months. The three-month gain holds for both targets, in the partly real-time exercise, and in subsamples before and after the pandemic. It is therefore neither a mechanical effect of high-tech output in headline IP nor a pandemic artifact.

Its source, however, changes over time. Before 2020, semiconductor and high-tech production carry the signal, mostly around the Great Recession, which is consistent with the electronics cycle leading the industrial cycle. After 2021, during the AI investment boom, the gain comes from capital-goods orders, a broad equipment-investment measure, while the AI-specific production series add nothing. The data therefore support the claim that investment activity helps forecast IP in the AI era more than the claim that AI-specific indicators do.

Three caveats remain. First, the post-2021 sample is short, so its tests have low power. Second, the real-time exercise uses revised AI data, so it is an upper bound, and it covers total IP only.

Third, at horizons of six months or more the AI block usually worsens forecasts.

7 Interpretation

The tests answer three questions in sequence:

1. **On average:** do AI indicators improve IP forecasts beyond a broad diffusion index (MSFE ratio, Clark–West)?
2. **When:** is the improvement recent, consistent with the post-2023 AI investment cycle (fluctuation test, CSSED)?
3. **Under what conditions:** is it concentrated in particular financial or cyclical states (Giacomini–White)?

A finding that holds for the ex-high-tech target is the strongest evidence that AI investment is transmitting to the broader industrial cycle rather than merely forecasting its own contribution to headline IP.

Validation. Before the code was applied to actual data, it was run on simulated data with a known structure. Three factors drive a 60-series panel and IP growth. From January 2018, two of the six AI indicators (in the high-tech and semiconductor IP slots) predict IP growth one month ahead, with an effect two-thirds larger when the simulated NFCI is negative. At $h = 1$, Clark–West rejects for both true indicators ($p < 0.001$ and $p = 0.013$) and for the full block against DI ($p < 0.001$). It also rejects for one indicator with no direct effect (capacity utilization, $p = 0.020$). That indicator shares the common factor with the true ones and so partly proxies for them, a reminder that single-indicator rejections identify useful *proxies*, not causal channels. Every fluctuation crossing occurs after the break, in windows ending between 2021 and 2025; the delay reflects the window length of about 78 forecasts. One indicator with no effect crosses marginally at $h = 12$ (2.87 against 2.77), about what chance predicts across 32 comparisons. Giacomini–White with NFCI as the instrument does not detect the state dependence in DI vs. DI+AI ($p = 0.23$). The test therefore has limited power against state effects of this size, and a non-rejection in the actual data should not be read as evidence that no state dependence exists.

References

- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.
- Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138(1), 291–311.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics* 25(4), 595–620.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74(6), 1545–1578.

- Hansen, P. R., A. Lunde and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- Harvey, D., S. Leybourne and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13(2), 281–291.
- McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34(4), 574–589.
- Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20(2), 147–162.