



BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA

Bank Negara Malaysia Working Papers

WP5/2026

Predicting Inventory Build-ups and Drawdowns in Malaysia: A Pseudo-Real-Time Evaluation

Tan Ling Yan, Liew Tien How, Helmi Ramlee

October 2026

Working papers describe research in progress by the authors and are published to elicit comments and further debate. The authors are grateful to Dr Mohamad Hasni bin Sha'ari, Dr James Yetman, Akmal bin Zulkarnain and Muhammad Hazrul Akmal bin Mohd Hazarudin for their valuable comments and suggestions, which helped improve this paper. Any remaining errors or omissions are the sole responsibility of the authors. The views expressed are solely those of the authors and should not be taken to represent those of Bank Negara Malaysia.

Abstract

Inventory investment is among the most volatile and difficult-to-observe components of expenditure-side gross domestic product (GDP). This paper develops a pseudo-real-time framework to assess whether real inventory investment in Malaysia is positive or negative in the target quarter. The baseline approach employs binary-response models, namely probit and complementary log-log specifications, using hard and soft indicators covering production, demand, inventory conditions, trade activity and business expectations. An ordinary least squares (OLS) forecast of real inventory investment provides a continuous benchmark. The analysis is extended through a mixed-frequency machine-learning framework incorporating temporal convolutional networks (TCNs), gated recurrent units (GRUs), attention networks, the FT-Transformer and a wide-and-deep architecture. Forecasts are generated under one-month, two-month and full-quarter information sets using expanding estimation windows to approximate a pseudo-real-time forecasting environment. Within the six-quarter exploratory evaluation, the binary-response models record superior directional predictive accuracy than the continuous OLS benchmark in most comparable specifications. The results also highlight the operational importance of aligning the information set used for estimation with that available at the forecast origin. The machine-learning extension provides a supplementary nonlinear comparison of whether temporal dynamics and complex interactions across a wider set of indicators add predictive content. Given the short evaluation window and limited effective training sample, the paper does not seek to identify a single best-performing model. Instead, it demonstrates how vintage-consistent binary models can be operationalised as a supplementary signal for inventory assessment in Malaysia.

Keywords: inventories; nowcasting; binary-response models; machine learning; ragged-edge nowcasting; Malaysia

JEL classification: C25; C45; C53; E27; E32

1 Introduction

Changes in inventories reflect the net addition to or withdrawal from the stock of goods that businesses hold for future sale, production or use. Under the expenditure approach to gross domestic product (GDP), changes in inventories are recorded as inventory investment because they represent output produced but not yet sold or consumed. Positive inventory investment adds to the level of GDP, while negative inventory investment subtracts from it. Its contribution to GDP growth, however, depends on the change in inventory investment relative to the preceding period. As inventory investment is highly volatile and responsive to shifts in demand expectations, production decisions and supply-chain conditions, it remains among the most challenging GDP components to forecast.

Despite its importance, inventory forecasting has received limited attention in the empirical literature, particularly in Malaysia, where dedicated forecasting frameworks remain scarce. To address this gap, this paper examines binary-response models for nowcasting whether real inventory investment is positive, indicating aggregate inventory accumulation, or negative, indicating aggregate destocking. Using a broad set of hard- and soft-data indicators related to inventory conditions, production and demand, the models generate directional signals that complement existing GDP forecasts and provide a systematic approach to tracking inventory developments. The value of this exercise extends beyond forecasting a single GDP component. Given the balancing role that inventories often play within the expenditure framework, a better understanding of inventory dynamics can strengthen the assessment and finalisation of headline GDP forecasts while providing insights into underlying demand-supply conditions and business-cycle developments.

This paper makes four contributions. First, it provides a dedicated pseudo-real-time assessment of the sign of inventory investment in Malaysia. Second, it compares conventional continuous forecasting with probability-based classification approaches. Third,

it illustrates the operational importance of aligning the information set used for estimation with that available at the forecast origin. Fourth, it integrates machine learning with ragged-edge information sets, enabling direct comparison with traditional binary-response models.

The remainder of the paper is organised as follows. Section 2 reviews the literature on inventory dynamics, binary-response forecasting and machine-learning-based now-casting. Section 3 presents key stylised facts on inventory investment in Malaysia. Section 4 outlines the empirical methodology, including indicator selection, forecasting models and evaluation procedures. Section 5 reports the exploratory recursive evaluation results, while Section 6 discusses their economic and operational implications. Section 7 addresses limitations and sensitivity considerations, and Section 8 concludes.

2 Literature Review

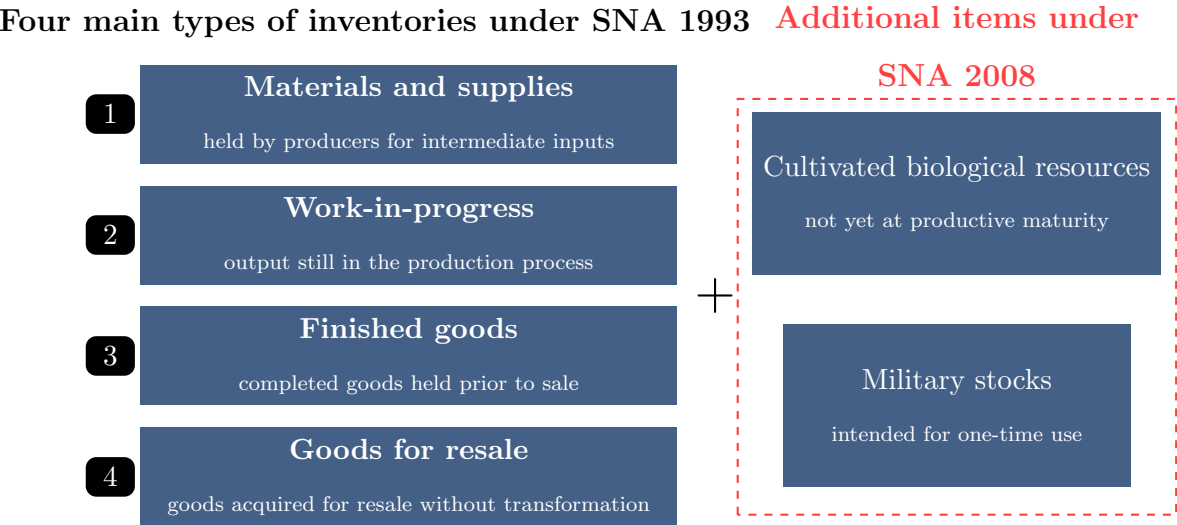
2.1 Inventories and the Business Cycle

Inventories have long been recognised as an important but relatively underexplored component of business-cycle dynamics. The classical inventory-cycle literature views inventory accumulation and destocking as mechanisms through which firms respond to changes in demand, production conditions and expectations. Under the United Nations System of National Accounts (SNA), inventories are broadly classified into four categories¹: materials and supplies, work in progress, finished goods, and goods for resale. These categories span multiple stages of the production and distribution chain.

¹Under the 2008 SNA, two additional items previously treated as fixed assets were reclassified as inventories: cultivated biological resources that have not reached productive maturity and military inventories intended for one-time use.

Materials and supplies are inputs held for use in the production process, such as raw materials, components and operating supplies. They are intended for intermediate consumption rather than resale. Work-in-progress (WIP) goods have entered the production process but are not completed by the end of the accounting period. Finished goods are completed products held by manufacturers before sale, with no further processing intended before supply to customers. Goods for resale are purchased by wholesalers and retailers for resale without substantial transformation, apart from activities such as sorting, packaging or transportation.

Figure 1: Main Types of Inventories under System of National Accounts



Reference: SNA 1993, SNA 2008, Shrestha, Manik L. and Fassler, S. (2003), Changes in Inventories in the National Accounts. IMF Working Paper No. 03/120

Inventory investment is highly procyclical and is among the most volatile components of GDP, often accounting for a disproportionate share of output fluctuations during downturns and recoveries. Inventory investment exhibits substantial cyclical variability and moves closely with broader economic activity (Khan and Thomas, 2002). However, the complex relationship between inventories and broader demand and supply dynamics remains insufficiently understood, making inventory movements and their contribution

to GDP difficult to predict.

Inventory movements are influenced not only by prevailing demand and production conditions but also by firms' expectations of future demand. Inventory accumulation may occur when realised demand falls short of prior expectations, resulting in a larger-than-desired stock of unsold goods. Accumulation can also arise when firms anticipate stronger future demand and increase or retain inventories in preparation for expected orders. Conversely, destocking may occur when realised demand exceeds expectations, prompting firms to satisfy sales through existing inventories. Expectations of weaker future demand may similarly encourage firms to reduce inventory holdings and limit accumulation.

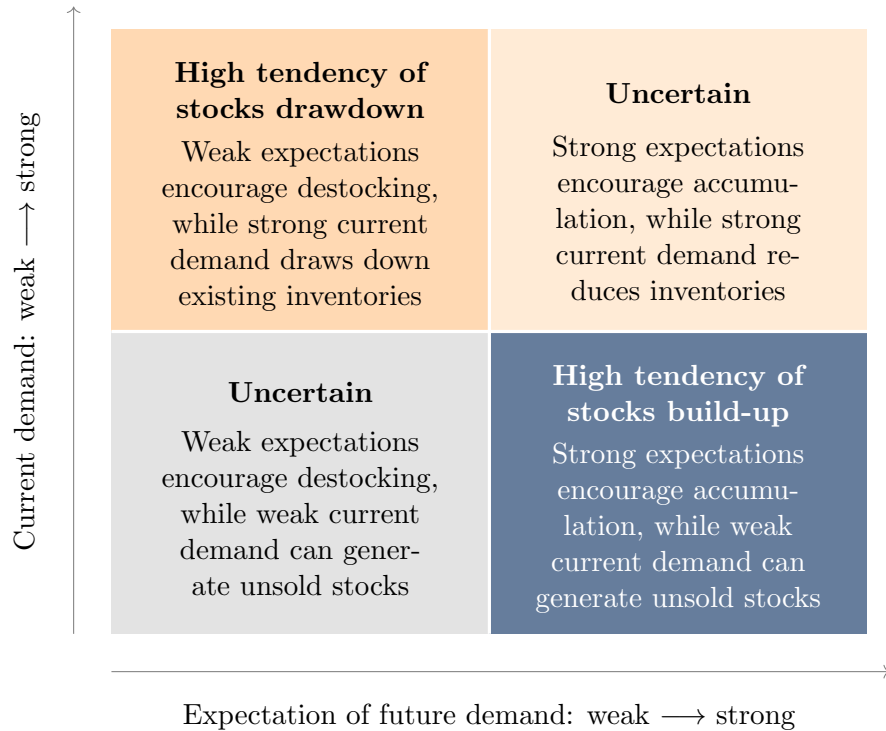
Current demand conditions and expectations of future demand can therefore exert opposing effects on inventory behaviour. Strong current demand may contribute to inventory depletion, while optimistic expectations of future demand may simultaneously encourage accumulation. This interaction complicates the interpretation of inventory movements because observed changes reflect both contemporaneous demand dynamics and forward-looking business expectations. Figure 2 provides a conceptual illustration of these relationships.

Inventory dynamics are also shaped by factors beyond economic activity, including uncertainty-driven precautionary stocking, shifts towards inventory-intensive business models and geopolitical developments. Economic activity alone may therefore have limited predictive power for inventory fluctuations.

2.2 Measurement of inventories and real-time forecasting constraints

Forecasting and interpreting inventory behaviour is inherently challenging. First, comprehensive and timely hard data on economy-wide inventory conditions are limited.

Figure 2: A stylised framework of inventory assessment



Source: Authors' illustration

Central banks and multilateral organisations therefore use high-frequency surveys to infer inventory dynamics in real time. The European Central Bank (ECB), for example, incorporates survey-based indicators into its inventory projections (Andersson & Breton, 2022). Purchasing managers' surveys, including the Institute for Supply Management's (ISM) Manufacturing Purchasing Managers' Index (PMI) and S&P Global Purchasing Managers' Index (PMI), are also widely used as timely indicators of inventory accumulation and depletion. Nevertheless, such indicators are often industry-specific and survey-based and therefore do not always provide a clear signal of aggregate inventory movements. For example, the PMI Finished Goods Inventory Index exhibits a weak negative correlation with the sign of inventory investment, with a correlation

coefficient of -0.45^2 , while the annual change in the index shows virtually no correlation with inventory dynamics.

Second, the measurement of inventory accumulation and destocking is not straightforward. Inventory investment contributes to GDP growth through changes in the pace of accumulation or destocking rather than through the level of the inventory stock itself. For example, inventories can make a positive contribution to GDP growth during a period of destocking if the pace of destocking slows. Hard- and soft-data indicators therefore cannot be translated directly into signals of aggregate inventory accumulation or destocking.

Finally, the measured inventory component may capture factors beyond underlying stock-volume movements. Changes in inventories may reflect valuation effects (Lal, 2006), as well as coverage and reconciliation issues inherent in the compilation of expenditure-side national accounts (Fassler and Shrestha, 2003), weakening their relationship with observable macroeconomic indicators.

Given these measurement challenges, identifying the likely sign of inventory investment may be more tractable than forecasting its magnitude. A binary modelling framework provides a practical alternative by estimating whether real inventory investment is positive or negative in a given quarter. This approach combines information from multiple hard- and soft-data indicators into a probabilistic assessment that can supplement the broader GDP forecasting process.

The broader macroeconomic forecasting literature has widely employed binary-response models, including logit and probit specifications, to predict turning points and regime shifts. Estrella and Mishkin (1998), for example, estimate recession probabilities using financial and macroeconomic indicators, while Borio, Drehmann and Xia (2019) apply panel probit models to assess recession risks across countries. These applications

²Authors' calculations based on S&P Global PMI and Department of Statistics Malaysia (DOSM) national accounts data for the quarterly sample from 2015Q1 to 2026Q2.

motivate the use of a probability-based framework for directional inventory assessment when the magnitude of inventory investment is difficult to estimate.

2.3 Machine learning and mixed-frequency nowcasting

Recent advances in machine learning have expanded the set of tools available for economic forecasting and nowcasting, particularly in environments characterised by nonlinear relationships, high-dimensional predictors and ragged-edge data availability. Machine-learning architectures can capture complex interactions among indicators without requiring the researcher to specify the complete functional form. This flexibility is potentially relevant for inventory dynamics, which reflect interactions among production, demand, expectations and supply-chain conditions.

Temporal convolutional networks (TCNs) were evaluated by Bai, Kolter and Koltun (2018), who show that purely convolutional architectures can outperform canonical recurrent networks across a broad range of sequence-modelling tasks while exhibiting long effective memory. TCNs use causal and dilated convolutions to capture short- and long-term dependencies while allowing parallel computation. For inventory nowcasting, lagged movements in production, trade and stock indicators may contain information about future inventory regimes.

Gated recurrent units (GRUs) were introduced by Cho et al. (2014) as a computationally efficient alternative to long short-term memory (LSTM) networks. Their gating mechanism enables the network to retain or discard historical information selectively, mitigating the vanishing-gradient problem that affects traditional recurrent architectures. A subsequent evaluation by Chung et al. (2014) found that GRUs achieved performance comparable to LSTMs while requiring fewer parameters and shorter training times. GRUs therefore provide a natural framework for modelling persistence in inventory-adjustment behaviour.

Vaswani et al. (2017) introduced the Transformer architecture and its self-attention mechanism, allowing the model to assess the relative importance of observations within a sequence without recurrent processing. Attention-based models can capture long-range dependencies that may be relevant for inventory dynamics when adjustment decisions and shocks unfold over extended periods.

For tabular applications, Gorishniy et al. (2021) proposed the FT-Transformer, an architecture adapted to heterogeneous numerical features. Because inventory nowcasting relies on macroeconomic, survey and industry-specific indicators, the FT-Transformer provides a flexible framework for learning interactions among predictors.

The wide-and-deep architecture introduced by Cheng et al. (2016) combines a linear “wide” component with a nonlinear “deep” component. In this application, the wide component can capture stable relationships, while the deep component can accommodate nonlinear responses to shifts in demand, production and supply-chain conditions.

The five architectures considered in this paper provide a diverse set of approaches to sequence and tabular-data prediction. Flexibility, however, does not imply superior economic forecasts. The relatively small sample of Malaysian quarterly observation heighten the risk of overfitting, while hyperparameter tuning adds another layer of model-selection uncertainty. The machine-learning exercise is therefore treated as a supplementary nonlinear forecasting comparison rather than as a presumption that more complex models are preferable.

2.4 Contribution to the literature

The study contributes to the literature through a Malaysian pseudo-real-time application to an under-examined GDP component. It compares directional and continuous forecasts, evaluates partial-quarter information sets and investigates both parsimonious econometric and flexible machine-learning classifiers. The ragged-edge design preserves

the operational question faced by forecasters: how does the inventory assessment change as information accumulates within the quarter? Agreement across structurally different models may also provide useful corroborating evidence when the evaluation window is too short to establish stable model rankings.

Before presenting the modelling framework, the next section documents selected stylised facts on inventory investment in Malaysia. These historical patterns provide context for the subsequent analysis and help motivate the indicator set and modelling approach.

3 Stylised facts on inventory investment in Malaysia

In Malaysia, inventory investment is considerably more volatile than aggregate economic activity, reflecting its role in absorbing economic shocks and facilitating adjustments during downturns and recoveries. Real inventory investment has fluctuated sharply between accumulation and destocking, particularly around major shocks such as the Asian Financial Crisis (1997–1998), the dot-com slowdown (2001), the Global Financial Crisis (2008–2009), the oil-price shock (2014–2016) and the COVID-19 pandemic (2020–2021).

Despite the conceptual link between inventories and the business cycle, the historical relationship between inventory investment and GDP growth appears relatively weak in Malaysia. Contemporaneous correlations are close to zero at both annual (0.15) and quarterly (0.09) frequencies, while lead correlations remain modest. This suggests that inventory movements are influenced by factors beyond current economic activity, including firms’ expectations, stock-management decisions, valuation effects and statistical balancing adjustments within the national accounts.

Inventory cycles can also be persistent. Table 1 reports the longest accumulation and destocking episodes, together with the largest historical movements and selected corre-

lations. The table carries the detailed dates and magnitudes to avoid duplicating the numerical information in the text.

Although inventory investment is generally small relative to GDP, changes in inventory investment can make sizeable percentage-point contributions to GDP growth during periods of economic disruption. This episodic importance, combined with weak observability and high volatility, motivates the directional forecasting approach used in this paper.

Overall, the analysis focuses on predicting the sign rather than the magnitude of real inventory investment. Reframing the problem as classification allows information from multiple indicators to be combined while reducing sensitivity to the large level fluctuations observed in the national accounts.

Table 1: Annual and quarterly properties of stocks

Property	Annual Basis	Quarterly Basis
Level change	Contemporaneous: 0.15	Contemporaneous: 0.09
correlation with	Lead 1-year: 0.15	Lead 1-quarter: 0.04
GDP (1992-)		Lead 2-quarter: 0.10
		Lead 3-quarter: 0.02
		Lead 4-quarter: 0.16
Largest Build-up	RM30.7 billion (2023)	RM16.5 billion (2Q21)
Level		
Longest	Longest: 7 years (2002-2008)	Longest: 11 quarters
Consecutive	Second longest: current 5 years	(2Q21-4Q23)
Build-up period	(2021-2025)	Second longest: 6 quarters (2Q10-3Q11)

Table 1: Annual and quarterly properties of stocks

Property	Annual Basis	Quarterly Basis
Largest drawdown Level	-RM13.9 billion (2019)	-RM10.6 billion (1Q09)
Longest Consecutive Drawdown Period	3 years (1996-1998, 2013-2015, 2018-2020)	7 quarters (3Q18-1Q20, 3Q13-1Q15)
Threshold* for Share of GDP	< 1.4% of GDP (except 2021-2023: 1.5%; 1.6%; 2.0%)	< 4.3% of GDP except (3Q95, 1Q09, 2Q21: 5.4%, 5.4%, 4.9%)
Threshold* for Percentage-point (ppt) contribution to GDP	< 1.5 ppt to GDP (except 2010, 2021: 2.0%; 1.9%)	< 4.3 ppt to GDP except (3Q95, 3Q96, 1Q09: 4.8%, 5.4%, 5.4%)

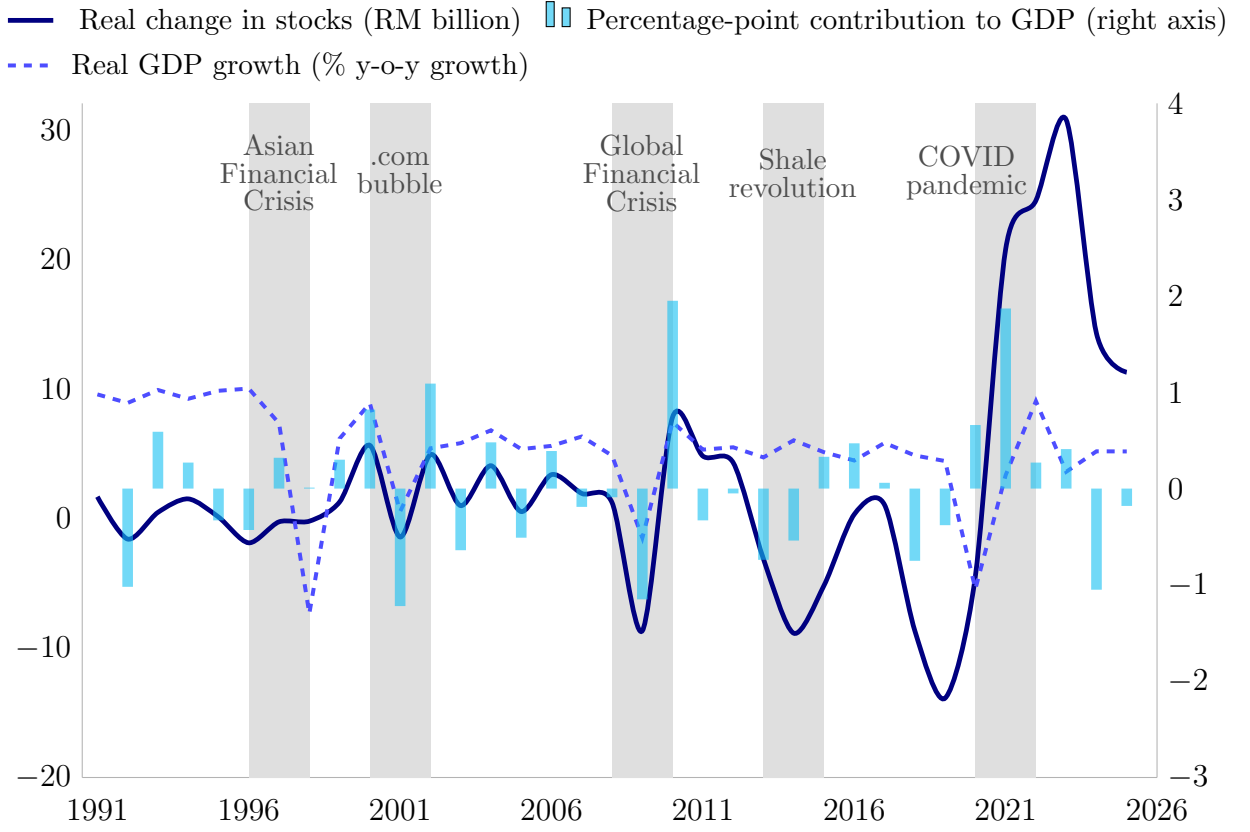
**Note: The thresholds are defined using 2 standard deviations for annual and 2.5 standard deviations for quarterly respectively.*

Source: Department of Statistics Malaysia; authors' calculations.

4 Methodology

This section describes the methodology used to predict the sign of real inventory investment in the target quarter. The framework combines indicators of production activity, demand conditions, inventory expectations and supply-chain developments. The analysis proceeds through construction of the binary target, indicator compilation,

Figure 3: Overview of GDP growth and change in inventories



Source: Department of Statistics Malaysia; authors' calculations.

preliminary screening, elastic-net variable selection, binary-response modelling and recursive evaluation. A continuous OLS forecast of real inventory investment provides the benchmark.

i. Construction of the dependent variable

Let I_t denote real inventory investment, measured by the national-accounts change-in-inventories component in quarter t . The binary target is defined as

$$Y_t = \begin{cases} 1, & \text{if } I_t > 0 \quad (\text{inventory accumulation}), \\ 0, & \text{if } I_t < 0 \quad (\text{inventory destocking}). \end{cases}$$

The classification therefore predicts the sign of real inventory investment. It does not predict whether inventory investment increased relative to the preceding quarter, nor does it directly predict inventories' contribution to GDP growth.

ii. Identification and compilation of indicators

Potential explanatory variables are selected based on the inventory-cycle literature, available business surveys and practical considerations of timeliness and availability. The indicator set spans production activity, demand conditions, inventory expectations and supply-chain developments. No individual indicator is expected to provide a comprehensive signal; the objective is to combine information from different sectors and stages of the production process.

Table 2: Description and rationale for variable selection

Variable	Description	Economic Rationale	Impact	Source
PPI-CPI growth gap (<i>p_gap</i>)	Difference between producer price index (PPI) year-on-year growth and consumer price index (CPI) year-on-year growth. <i>Frequency: Monthly</i>	A widening producer-consumer price gap may signal cost pressures, weaker final demand absorption or accumulation of unsold goods. Conversely, stronger consumer prices relative to producer prices may indicate faster inventory clearance.	Predictor of current demand relative to expectations, Positive	DOSM
Manufacturing IPI (<i>ipi_mfg</i>)	Manufacturing Industrial Production Index (IPI) year-on-year growth <i>Frequency: Monthly</i>	Strong manufacturing production increases the likelihood of inventory accumulation if output growth exceeds final demand.	Predictor of current production, Positive	DOSM

Table 2: Description and rationale for variable selection

Variable	Description	Economic Rationale	Impact	Source
Survey Expected Finished Retail Goods (<i>bts_retail</i>)	Business Tendency Survey (BTS) indicates higher, same or lower expectation of stock of retail goods for the upcoming quarter.	Expectation of stock of finished goods and end period, firms expecting higher retail inventories indicate anticipated stock accumulation.	Soft indicator for stocks, Positive	DOSM
<i>Frequency: Quarterly</i>				
Survey Expected Finished Manufacturing Goods (<i>bts_mfg</i>)	Business Tendency Survey (BTS) indicator of manufacturers' expectations regarding inventory levels in the upcoming quarter (higher, unchanged, or lower)	Expectation of stock of finished goods and end period, direct forward-looking indicator of manufacturers' inventory intentions.	Soft indicator for stocks, Positive	DOSM
<i>Frequency: Quarterly</i>				

Table 2: Description and rationale for variable selection

Variable	Description	Economic Rationale	Impact	Source
Intermediate Imports / Manufacturing Exports Ratio y-o-y change (<i>ig_xmfg</i>)	Year-on-year change of nominal intermediate imported goods to nominal manufacturing exported goods ratio. <i>Frequency: Monthly</i>	A rise in imported intermediate inputs relative to exports may signal inventory stockpiling ahead of production or weaker external demand.	Predictor of current demand relative to expected demand, Positive	DOSM
Palm Oil and Palm-Based Product Stocks (<i>palm</i>)	Year-on-year difference in sum of stock of palm oil, palm kernel and palm kernel cake in metric tonnes. <i>Frequency: Monthly</i>	Palm oil inventories constitute a significant share of commodity stockholding in Malaysia and may provide information on broader inventory conditions within the commodity sector.	Hard indicator for stocks, Positive	MPOB

Table 2: Description and rationale for variable selection

Variable	Description	Economic Rationale	Impact	Source
Motor Vehicle Production less Sales (<i>net_mv</i>)	Difference between motor vehicle production and motor vehicle sales in units. <i>Frequency: Monthly</i>	Inventories accumulate when vehicle production outpaces sales and decline when sales exceed production.	Predictor of current demand relative to expected demand, Positive	MAA
PMI Finished Goods Inventory Index (<i>pmi_fg</i>)	Purchasing Managers' Index indication of stocks of manufacturing finished goods, seasonally adjusted. <i>Frequency: Monthly</i>	Measures changes in manufacturers' finished goods inventories and provides a timely assessment of inventory conditions.	Soft indicator for stocks, Positive	S&P Global
Crude Oil Storage Capacity Utilisation (<i>oil_cu</i>)	Crude Oil levels in million barrels divided by the available capacity captured by Kpler satellite data. <i>Frequency: Monthly</i>	Higher utilisation signifies more crude oil stocks inventory.	Hard indicator for stocks, Positive	Kpler

The mixing parameter is fixed at $\alpha = 0.5$, assigning equal weight to the lasso and ridge components. This provides a transparent compromise between sparse selection and stabilisation of correlated predictors while avoiding an additional tuning dimension in a small quarterly sample. The regularisation parameter λ is selected through cross-validation. Variables with non-zero coefficients form the initial screening set for the binary-response models. Elastic-net selection is used as a predictive screening device and is not interpreted as evidence of structural importance.

v. Binary-response models

The selected variables are included in two binary-response specifications.

Probit model

$$P(Y_t = 1 \mid X_t) = \Phi(X_t\beta),$$

where Φ denotes the standard normal cumulative distribution function.

Complementary log-log model

$$P(Y_t = 1 \mid X_t) = 1 - \exp\{-\exp(X_t\beta)\}.$$

The complementary log-log specification permits an asymmetric mapping between the linear predictor and the probability of positive inventory investment, unlike the symmetric probit link. It is treated as an alternative probability link rather than as evidence of regime-specific structural coefficients.

A predicted probability above 0.5 indicates inventory accumulation, while a probability below 0.5 indicates destocking. The 0.5 threshold is retained as a transparent convention rather than an estimated optimum.

Target-quarter assessments are generated recursively using expanding estimation windows under one-month, two-month and full-quarter information sets. The evaluation

covers 2025Q1–2026Q2, providing six target quarters. Because this development window partly informed final specification refinement, the results are interpreted as an exploratory recursive evaluation rather than a fully untouched out-of-sample test. The matched-vintage exercises assess whether relationships estimated using partial-quarter information transfer to comparable ragged-edge forecasting environments.

vi. **Continuous forecasting benchmark**

As a benchmark, a second modelling strategy forecasts real inventory investment, measured in RM million, rather than its sign. Elastic net with $\alpha = 0.5$ is used for variable selection, after which the selected predictors are estimated using ordinary least squares:

$$I_t = \beta_0 + \beta X_t + \varepsilon_t.$$

The resulting forecast of inventory investment is converted into an implied accumulation or destocking signal and compared with the binary-model predictions.

vii. **Machine-learning models**

Five machine-learning architectures are estimated: TCN, GRU, an attention model, FT-Transformer and a wide-and-deep model. The architectures reviewed in Section 2 are implemented using the same vintage-specific predictor sets, and each model returns a sigmoid probability of positive inventory investment.

Hyperparameters are selected once through Bayesian optimisation conducted within the training sample. The search covers network width, filters or recurrent units, attention heads, feed-forward dimensions, dropout, learning rate and batch size. Hyperparameters are selected to maximise the area under the receiver operating characteristic curve using five-fold stratified cross-validation. Scaling parameters are fitted within each

training fold, and early stopping limits unnecessary training. The selected hyperparameters are then held fixed during the recursive evaluation.

For each target quarter, model and vintage combination, feature scalers and network weights are re-estimated using the expanding training sample. Forecasts are averaged across three random seeds to reduce sensitivity to initialisation.

viii. Model comparison and evaluation

The final stage compares the probit model, complementary log-log model, continuous OLS benchmark, TCN, GRU, attention model, FT-Transformer and wide-and-deep model. The specifications are compared descriptively using directional classification accuracy and recursive performance under reconstructed one-month, two-month and full-quarter information sets. The continuous benchmark provides a direct comparison between sign inference from a level forecast and direct binary classification.

ix. Sensitivity analysis

The models are first compared with two naïve reference classifiers: (i) always predicting accumulation, which historically occurs in around 54% of quarters; and (ii) predicting the same sign as in the previous quarter. The reference classifiers provide context for the reported six-quarter success rates.

As a separate sensitivity exercise, *bts_retail* and *bts_mfg*, which are not included in the reported parsimonious specifications, are added and the models are re-estimated. The exercise assesses whether the reported directional classifications remain stable after the additional BTS indicators are included; it is not interpreted as a formal statistical robustness test.

5 Results and Findings

Figures 4–6 show the target-quarter recursive classifications for the parsimonious and machine-learning models from 2025Q1 to 2026Q2, using expanding estimation windows beginning in 2017Q1.

5.1 Sensitivity to additional BTS indicators

Within the six-quarter evaluation, the parsimonious specifications generally record higher success rates than most machine-learning variants. As a sensitivity exercise, *bts_retail* and *bts_mfg* are added to the parsimonious specifications that exceed the naïve reference classifiers. The complementary log-log classifications are unchanged, while the success rates of the partial-quarter probit specifications decline. These results, reported in Table 3, indicate classification stability for the complementary log-log specifications under this specific perturbation, but they do not establish broader parameter or probability-calibration robustness.

Table 3: Sensitivity to the inclusion of BTS variables

Reported specification	Six-quarter success rate	Success rate with additional BTS indicators
3m probit	83.3%	83.3%
3m cloglog	100%	100%
2m–2m probit	83.3%	66.7%
2m–2m cloglog	83.3%	83.3%
1m–1m probit	66.7%	50.0%
1m–1m cloglog	66.7%	66.7%

Figure 4: Recursive success rate of the parsimonious models

Selected Variables		Year	2025				2026		Recursive success rate (6Q)		
		Quarter	1	2	3	4	1	2			
Actual		-	+	-	+	-	+	Dist.	50.0%	50.0%	
ig_xmfg, net_mv, palm, p_gap, oil_cu	Train full Q, forecast with full Q	3m Probit	-	-	-	+	-	+	83.3%	+	-
										66.7%	100.0%
		3m cloglog	-	+	-	+	-	+	100.0%	100.0%	100.0%
	3m Reg	-	-	-	-	-	-	50.0%	0.0%	100.0%	
	Train full Q, forecast with 2m	2m Probit	+	-	-	-	+	+	33.3%	33.3%	33.3%
		2m cloglog	+	-	-	-	+	+	33.3%	33.3%	33.3%
	2m reg	+	-	-	-	-	-	33.3%	0.0%	66.7%	
Train full Q, forecast with 1m	1m Probit	+	-	+	-	+	-	0.0%	0.0%	0.0%	
	1m cloglog	+	-	+	-	+	-	0.0%	0.0%	0.0%	
1m reg	+	-	-	-	-	-	33.3%	0.0%	66.7%		
pmi_fg, ig_xmfg, net_mv, palm, p_gap, oil_cu, ipi_mfg	Train 2m, forecast with 2m	Probit 2m2m	-	+	-	+	+	+	83.3%	100.0%	66.7%
	cloglog 2m2m	-	+	-	+	+	+	83.3%	100.0%	66.7%	
	Train 1m, forecast with 1m	Probit 1m1m	-	+	-	+	+	-	66.7%	66.7%	66.7%
cloglog 1m1m	-	+	-	+	+	-	66.7%	66.7%	66.7%		

Note: Green highlighted accuracy rate indicates model outperforms a naïve classifier of predicting build-ups all the time, while green font indicates model outperforms a naïve classifier of predicting the same direction of the previous quarter.

+	Stock buildup		Accurate
-	Stock drawdown		Inaccurate

Figure 5: Recursive success rate of the machine-learning models

ML Model	Year Quarter	2025				2026		Recursive success rate (6Q)		
		1	2	3	4	1	2	<i>Dist.</i>		
TCN	Actual	-	+	-	+	-	+	50.0%	50.0%	50.0%
	<i>3m</i>	+	+	+	-	-	+	50.0%	66.7%	33.3%
	<i>2m</i>	+	+	+	-	-	+	50.0%	66.7%	33.3%
	<i>1m</i>	+	+	+	-	-	+	50.0%	66.7%	33.3%
	<i>3m</i>	+	-	+	-	-	+	33.3%	33.3%	33.3%
	<i>2m</i>	+	+	+	-	-	+	50.0%	66.7%	33.3%
GRU	<i>3m</i>	+	+	+	-	-	-	33.3%	33.3%	33.3%
	<i>2m</i>	+	+	+	-	-	+	33.3%	33.3%	33.3%
	<i>1m</i>	+	+	+	-	-	-	16.7%	0%	33.3%
Attention	<i>3m</i>	+	+	+	-	+	+	33.3%	66.7%	0%
	<i>2m</i>	+	+	+	-	+	-	16.7%	33.3%	0%
	<i>1m</i>	-	+	+	+	-	+	83.3%	100.0%	66.7%
FT-Transformer	<i>3m</i>	+	+	+	-	-	+	50.0%	66.7%	33.3%
	<i>2m</i>	+	+	+	-	+	+	33.3%	66.7%	0%
	<i>1m</i>	-	+	+	+	-	+	83.3%	100.0%	66.7%
Wide-and-deep	<i>3m</i>	+	+	+	-	-	-	33.3%	33.3%	33.3%
	<i>2m</i>	+	+	+	-	+	-	16.7%	33.3%	0%
	<i>1m</i>	+	+	+	-	+	-	16.7%	33.3%	0%

Note: Green highlighted accuracy rate indicates model outperforms a naïve classifier of predicting build-ups all the time, while green font indicates model outperforms a naïve classifier of predicting the same direction as the previous quarter.

+	Stock buildup	Accurate
-	Stock drawdown	Inaccurate

Figure 6: Recursive success rate of the machine-learning models trained on matched partial-quarter information

ML Model	Year Quarter	2025				2026		Recursive success rate (6Q)		
		1	2	3	4	1	2			
TCN	Actual	-	+	-	+	-	+	<i>Dist.</i>	50.0%	50.0%
	<i>2m</i>	+	-	+	-	-	+		+	-
	<i>vintage</i>							33.3%	33.3%	33.3%
	<i>1m</i>	+	+	+	-	-	+			
GRU	<i>2m</i>	+	+	+	-	-	-			
	<i>vintage</i>							33.3%	33.3%	33.3%
	<i>1m</i>	+	+	+	-	-	-			
	<i>vintage</i>							33.3%	33.3%	33.3%
Attention	<i>2m</i>	+	-	+	-	-	+			
	<i>vintage</i>							33.3%	33.3%	33.3%
	<i>1m</i>	+	+	+	-	-	+			
	<i>vintage</i>							50.0%	66.7%	33.3%
FT-Transformer	<i>2m</i>	-	+	+	+	-	+			
	<i>vintage</i>							83.3%	100.0%	66.7%
	<i>1m</i>	-	+	+	+	-	+			
	<i>vintage</i>							83.3%	100.0%	66.7%
Wide-and-deep	<i>2m</i>	-	+	+	+	-	+			
	<i>vintage</i>							83.3%	100.0%	66.7%
	<i>1m</i>	-	+	+	+	-	+			
	<i>vintage</i>							83.3%	100.0%	66.7%

Note: Green highlighted accuracy rate indicates model outperforms a naïve classifier of predicting build-ups all the time, while green font indicates model outperforms a naïve classifier of predicting the same direction as the previous quarter.

+	Stock buildup		Accurate
-	Stock drawdown		Inaccurate

6 Discussion

All findings in this section describe the six-quarter development window and are not interpreted as population model rankings.

First, classifications are sensitive to the choice of explanatory variables and the information set available at the forecast origin. The reported full-quarter models use the intermediate-imports-to-manufactured-exports ratio (*ig_xmfg*), motor-vehicle production less sales (*net_mv*), palm-related stocks (*palm*), the PPI–CPI growth gap (*p_gap*) and crude-oil storage capacity utilisation (*oil_cu*). These hard-data indicators generally exhibit the expected positive association with inventory accumulation. The partial-quarter specifications additionally include finished-goods PMI (*pmi_fg*) and manufacturing IPI (*ipi_mfg*). Their estimated coefficients are less stable in sign and statistical precision, so their role is interpreted as incremental predictive content rather than as evidence of structural relationships.

Second, direct binary classification records more correct directional classifications than the displayed OLS benchmark in most comparable specifications. With full-quarter information, the complementary log-log model correctly classifies six of six evaluation quarters, the probit model five of six, and the OLS benchmark three of six. These results suggest that modelling the sign of inventory investment directly may be useful when level forecasts are noisy, although the short evaluation window precludes a stable ranking of model performance.

Third, the OLS benchmark shows a tendency to classify observations as destocking, leading to higher accuracy during realised destocking episodes than during accumulation episodes. Given the small number of observations in each class, this pattern is treated as a feature of the present evaluation rather than as evidence of a general forecast bias.

Fourth, applying full-quarter relationships mechanically to partial-quarter inputs performs poorly in the current evaluation. Models estimated on full-quarter data record

success rates of 33.3% and 0.0% when applied to two-month and one-month inputs, respectively. In contrast, matched two-month probit and complementary log-log specifications classify five of six quarters correctly, while matched one-month models classify four of six. Selected machine-learning models show a similar benefit from vintage alignment: the FT-Transformer and wide-and-deep models each record an accuracy rate of 83.3% under both matched one-month and matched two-month information sets. The result supports aligning estimation and forecast-origin information sets, but it does not establish a universal performance gain.

Finally, machine-learning performance is model-specific and does not display a uniform decline as the information set becomes more limited. The FT-Transformer and wide-and-deep models record the strongest success rates within the current evaluation, while the TCN, GRU and attention models are less consistent. The narrower operational conclusion is therefore that information-set alignment matters; the results do not establish a monotonic information gradient from one-month to full-quarter forecasts.

7 Limitations

This nowcasting exercise is subject to several limitations. First, the effective training sample is constrained by the *oil_cu*, which is only available from 2017. This is particularly restrictive for the neural-network models because Bayesian hyperparameter optimisation, five-fold stratified cross-validation and multiple random seeds create substantial model flexibility relative to the information contained in the sample. Simpler regularised classifiers, tree-based models and Bayesian shrinkage specifications therefore remain relevant benchmarks for future work.

Second, the evaluation window contains only six quarters from 2025Q1 to 2026Q2. A single classification error changes accuracy by 16.7 percentage points, and the area under the receiver operating characteristic curve (AUC) may be unstable with few observations

in either class. Reported success rates should therefore be regarded as preliminary development-sample evidence rather than population performance estimates.

Third, machine-learning hyperparameters are selected using stratified random folds. This preserves class balance but not temporal order, allowing later observations in the training sample to influence model selection for earlier validation folds. A rolling-origin or blocked time-series validation scheme would better match the forecasting problem, but such approaches were not implemented because they would further reduce the size of each validation sample.

Fourth, the 0.5 classification threshold is a transparent convention rather than an estimated optimum. If the policy cost of missing an accumulation differs from that of issuing a false accumulation signal, the classification threshold should reflect the relevant loss function. The 0.5 threshold is retained here for consistency across models and because the six-quarter evaluation contains an equal number of positive and negative observations.

Fifth, interpretability differs across model classes. Coefficient signs in probit and complementary log-log models can be related to economic channels, whereas neural-network predictions require post-estimation tools such as permutation importance or SHAP-type explanations. With correlated predictors, such diagnostics should be interpreted as predictive attribution rather than causal effects.

Finally, the target is directional. A correct sign does not identify the magnitude of inventory investment or inventories' contribution to GDP growth. The framework therefore complements, rather than replaces, judgemental or model-based estimates of inventory magnitude. The measured inventory series may also contain valuation and balancing components that cannot be forecast using the included indicators.

8 Conclusion

This paper develops a pseudo-real-time framework for classifying the sign of real inventory investment in Malaysia. Probit and complementary log-log specifications are compared with a continuous OLS benchmark under one-month, two-month and full-quarter information sets, while a machine-learning extension examines whether nonlinear and dynamic representations provide corroborating directional signals. The exercise uses expanding estimation windows over a six-quarter development period from 2025Q1 to 2026Q2.

The results provide preliminary evidence that partial-quarter indicators contain information about the sign of inventory investment. Under matched information definitions, the binary-response specifications correctly classify four or five of the six evaluation quarters. More consistently, models estimated using full-quarter data transfer poorly when applied mechanically to partial-quarter inputs, whereas models estimated using one-month or two-month information record higher success rates when evaluated under comparable information constraints. This finding highlights the practical importance of aligning the information set used for estimation with that available at the forecast origin.

The conclusions remain qualified by the short evaluation window, the small effective training sample and the directional nature of the target. Inventory adjustment can also be driven by idiosyncratic firm-level factors, sudden demand shocks and supply disruptions not fully captured by the included indicators. Future forecast rounds should therefore extend the prospective evaluation, archive quarter-level probabilities and forecast-origin information, and assess model stability as additional observations become available.

Subject to these qualifications, vintage-consistent binary-response models provide a structured supplementary signal for monitoring inventory accumulation and destocking

in Malaysia. For policy analysis and nowcasting, this signal should be used alongside model-based estimates of magnitude and expert judgement rather than as a stand-alone forecast of inventories' contribution to GDP growth.

A Appendix

Model specification for full-quarter training sample

Explanatory variable	Probit 3m	Complementary log-log 3m
IG_XMFG	0.6771 [†] (0.5023)	0.9093 (0.6338)
NET_MV	0.000226*** (0.000076)	0.000320*** (0.000109)
PALM	0.000997* (0.000531)	0.001535** (0.000728)
P_GAP	0.0980 [†] (0.0622)	0.1150 [†] (0.0823)
OIL_CU	0.0153** (0.0060)	0.0313*** (0.0101)
Observations	37	37
Log-likelihood	−15.061	−14.405
Akaike information criterion	1.114	1.049
Schwarz information criterion	1.332	1.267

Model specification for partial-quarter training sample

Explanatory variable	Probit	Cloglog	Probit	Cloglog
	2m	2m	1m	1m
PALM	0.000728* (0.000442)	0.000939* (0.000561)	0.000541 (0.000435)	0.000650 (0.000562)
P_GAP	0.1050† (0.0654)	0.1081† (0.0743)	0.1236* (0.0678)	0.1336* (0.0790)
OIL_CU	0.0534 (0.0437)	0.0580 (0.0492)	0.0153 (0.0407)	0.0133 (0.0506)
NET_MV	−0.000032 (0.000047)	−0.000050 (0.000062)	−0.000059† (0.000040)	−0.000079† (0.000055)
PMI_FG	−0.0612 (0.0517)	−0.0582 (0.0569)	−0.0141 (0.0489)	−0.0017 (0.0607)
IPL_MFG	0.0016 (0.0277)	0.0002 (0.0314)	−0.0146 (0.0410)	−0.0231 (0.0444)
IG_XMFG	0.4625† (0.3583)	0.4989 (0.3908)	0.5253† (0.3434)	0.5764† (0.3859)
Observations	37	37	37	37
Log-likelihood	−21.569	−21.824	−20.569	−20.652
Akaike information criterion	1.544	1.558	1.490	1.495
Schwarz information criterion	1.849	1.863	1.795	1.799

Notes: The dependent variable, *d_stocks*, equals one when real inventory investment is positive and zero when it is negative. Standard errors are reported in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively; † indicates the 20% level and is reported for transparency only. The 20% threshold is not interpreted as strong statistical evidence. The estimation sample is 2017Q1–2026Q1.

References

- Andersson, M. & Le Breton, G. (2022), ‘The role of the inventory cycle in the current recovery’, ECB Economic Bulletin (2).
URL: <https://www.ecb.europa.eu/press/economic-bulletin/html/eb202202.en.html>
- Bai, S., Kolter, J. Z. & Koltun, V. (2018), ‘An empirical evaluation of generic convolutional and recurrent networks for sequence modeling’, arXiv preprint arXiv:1803.01271 .
URL: <https://arxiv.org/abs/1803.01271>
- Borio, C., Drehmann, M. & Xia, D. (2019), Predicting recessions: Financial cycle versus term spread, BIS Working Paper 818, Bank for International Settlements.
URL: <https://www.bis.org/publ/work818.pdf>
- Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M. et al. (2016), Wide & deep learning for recommender systems, in ‘Proceedings of the 1st Workshop on Deep Learning for Recommender Systems’, pp. 7–10.
URL: <https://arxiv.org/abs/1606.07792>
- Cho, K., Van Merriënboer, B., Bahdanau, D. & Bengio, Y. (2014), Learning phrase representations using rnn encoder-decoder for statistical machine translation, in ‘Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing’, pp. 1724–1734.
URL: <https://arxiv.org/abs/1406.1078>
- Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. (2014), ‘Empirical evaluation of gated recurrent neural networks on sequence modeling’, arXiv preprint arXiv:1412.3555 .
URL: <https://arxiv.org/abs/1412.3555>
- Estrella, A. & Mishkin, F. S. (1998), ‘Predicting u.s. recessions: Financial variables as leading indicators’, Review of Economics and Statistics **80**(1), 45–61.

European Central Bank (2003), ‘The contribution of inventory changes to business cycle fluctuations’, ECB Monthly Bulletin .

URL: <https://www.ecb.europa.eu/pub/pdf/other/mb200306focus06.en.pdf>

European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations & World Bank (2009), System of National Accounts 2008, United Nations.

URL: <https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf>

Fassler, S. & Shrestha, M. L. (2003), Changes in inventories in the national accounts, IMF Working Paper 03/120, International Monetary Fund.

Gorishniy, Y., Rubachev, I., Khrulkov, V. & Babenko, A. (2021), Revisiting deep learning models for tabular data, in ‘Advances in Neural Information Processing Systems’, Vol. 34.

URL: <https://arxiv.org/abs/2106.11959>

Khan, A. & Thomas, J. K. (2002), Inventories and the business cycle: An equilibrium analysis of (s,s) policies, Working Paper 02-20, Federal Reserve Bank of Philadelphia.

URL: <https://www.philadelphiafed.org>

Lal, K. (2006), Recording of change in inventories in the sna and in the business accounts: A case study of canadian practices, Technical report, Statistics Canada.

URL: <https://www150.statcan.gc.ca>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017), Attention is all you need, in ‘Advances in Neural Information Processing Systems’, Vol. 30.

URL: <https://arxiv.org/abs/1706.03762>