The Impact of the Financial Sector Development Program 2030 on the Saudi Economy

أثر برنامج تطوير القطاع المالي 2030 على الاقتصاد السعودي

Fawaz Sudan Alotaibi[1]
Department of Economics- King Khalid Military Academy- Saudi Arabia
f.s.alotaibi@hotmail.com

Received: 27/09/2026 *Accepted: 14/10/2026* *Published: 11/11/2026*

Abstract

This study examines whether Saudi non-oil growth changed around the introduction of the Financial Sector Development Program (FSDP), whether banking and capital-market activity are associated with subsequent growth, and whether constrained machine learning improves prediction. Using official data for 2010Q1–2025Q4, an interrupted time-series model with a 2018Q2 introduction date, macroeconomic controls and pandemic indicators is estimated with HAC inference, block bootstraps and sensitivity checks. The 62-observation baseline estimates a trend change of 0.094 percentage points per quarter (95% HAC interval [0.047, 0.141]); the immediate shift is imprecise, and evidence weakens with an additional lag, shorter samples and aggregate GDP. Credit and turnover specifications do not establish stable positive associations. In retrospective forecast comparisons, ridge regression and random forests do not reliably outperform simple benchmarks, and financial predictors add no consistent accuracy. The findings support a conditional program-period change in the non-oil growth trajectory, without isolating an FSDP-caused gain.

Key words: *Saudi Arabia, Financial Sector Development Program, non-oil growth, interrupted time series, machine learning*

ملخص

تبحث الدراسة ما إذا كان نمو القطاع غير النفطي السعودي تغيّر بالتزامن مع إطلاق برنامج تطوير القطاع المالي، وما إذا كان نشاط القطاع المصرفي وسوق رأس المال يرتبطان بالنمو اللاحق، وقدرة نماذج التعلم الآلي المقيّدة على تحسين التنبؤ. وباستخدام بيانات رسمية للفترة 2010–2025، يُقدَّر نموذج للسلاسل الزمنية المتقطعة يعتمد الربع الثاني من 2018 تاريخًا لإطلاق البرنامج، مع ضوابط كلية واختبارات الحساسية. ويقدّر نموذج الأساس (62 مشاهدة) تغيرًا Block Bootstrap وأسلوب HAC ومؤشرات للجائحة، واستدلال في اتجاه النمو قدره 0.094 نقطة مئوية لكل ربع (فاصل ثقة 95%: [0.047، 0.141])، بينما يفتقر التحول الفوري إلى الدقة، وتضعف الأدلة مع فترة إبطاء إضافية والعينات الأقصر وإجمالي الناتج. ولا تُثبت مواصفات الائتمان ودوران السوق ارتباطات إيجابية والغابات العشوائية بثبات على النماذج المرجعية البسيطة، ولا Ridge مستقرة. وفي المقارنات التنبؤية الاسترجاعية، لا يتفوق نموذجا تضيف المتغيرات المالية دقة مستقرة. وتدعم النتائج تغيرًا مشروطًا في مسار النمو غير النفطي خلال البرنامج، دون عزل أثر سببي له.

الكلمات المفتاحية: المملكة العربية السعودية، برنامج تطوير القطاع المالي، النمو غير النفطي، السلاسل الزمنية المتقطعة، التعلم الآلي.

[1] - *Corresponding author: Fawaz Sudan Alotaibi, e-mail f.s.alotaibi@hotmail.com.*

1. INTRODUCTION

Saudi Arabia is trying to diversify its economic activity away from hydrocarbons, with financial development being key. The policy issue is whether a more capable financial system facilitates productive investment and non-oil output or just adds to financial assets. Within Vision 2030, in 2018, a new program was launched, the Financial Sector Development Program (FSDP), which offers an institutional framework for exploring this question. It aims to link financial institutions and markets with economic diversification and financial participation [1]. In addition to the other reforms and external shocks, there is also a national program functioning. Therefore, it is important to make a clear distinction between documented implementation and changes in economic performance and effects that can reasonably be attributed to policy when assessing its economic impact, as done by the International Monetary Fund (IMF) [2].

Measurable financial change is shown in the original regulatory statistics. According to the Saudi Central Bank (SAMA) [3], electronic payments accounted for 79% of retail payments in 2024, up from 70% in 2023. The total value of electronic non-cash transactions rose from 10.8 billion to 12.6 billion during the same time. These observations indicate a payment adoption increase, but they cannot give an indication of the extra output that the FSDP produced. Transactions can rise as a result of substitution away from cash, changes in consumption or bettering in the way that it is recorded. Figure 1 is therefore intended to provide the institutional context rather than an estimate of program effectiveness.

The primary outcome is non-oil real gross domestic product (GDP), since it is more closely related to diversification than is aggregate GDP. However, non-oil activity is not economically independent from hydrocarbons. The ability for oil income to affect domestic demand is through public spending, investment, and liquidity in the banking sector. Alabdulwahab [4] finds a long-run correlation between oil rent and non-oil GDP, which justifies the consideration of the oil environment in the interpretation of diversification. Aggregate GDP continues to be a second-order measure and allows the analysis to separate out financial transactions with the non-oil sector of the economy.

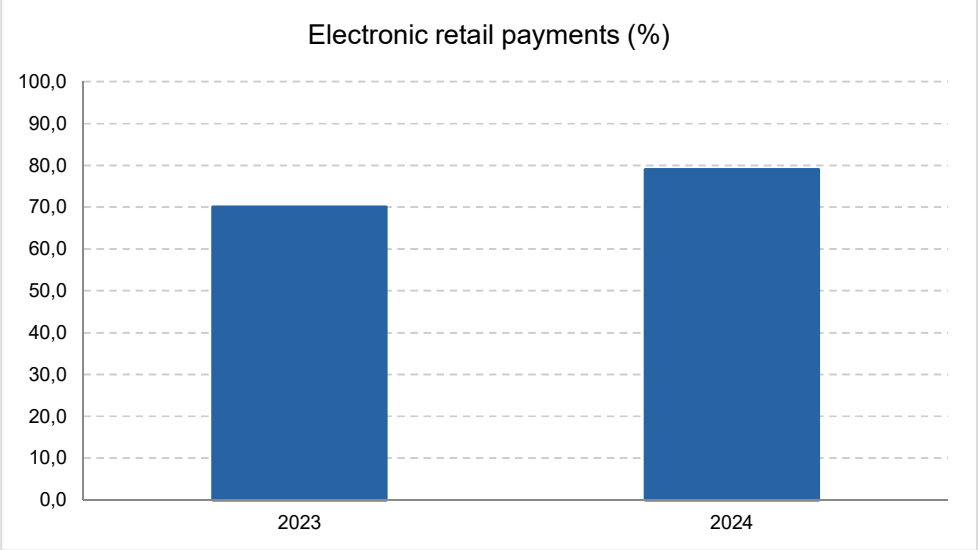
There is a need to analyze financial channels separately in view of the existing Saudi research and the reasons it offers. Haque [5] finds that there are varying kinds of linkages between private sector output and bank credit as well as money supply and stock-market capitalization. Alzyadat [6] also discovers that, at the sectoral level, the growth of credit relations is different for non-oil sectors at various time horizons. The results here warn against equating all rises in financial transactions with economic growth. They also raise an interesting empirical issue: whether the processes and patterns of relationships and growth evolved beyond the first few years after the implementation of FSDP.

This study has three questions. Was there a level or a trend change in non-oil growth before and after the introduction of FSDP? Do bank lending and capital-market activity link to the growth of non-oil activities in the following year, once selected macroeconomic conditions are considered? Are prediction tasks better achieved using carefully constrained machine-learning models than using

clear statistical models? The contribution is comprised of a coordinated design which links policy timing, separate financial channels and chronological prediction tests. Does not require asserting that machine learning automatically provides causal identification or that the methodologies have not been used in Saudi Arabia before.

This study provides a replicable measure that relates policy timing, financial-channel associations and chronological prediction. The quarterly time window collected is 2010Q1-2025Q4, where the number of usable model samples is based on source coverage and lags. The baseline marks a scenario of a change in growth trend, and financial-channel estimates are not precise, and gains from machine learning are not reliable. These results are made up of institutional progress and program attribution. The word 'impact' in the title refers to the issue of policy, and the conclusions are based on the aggregate observational evidence. The literature, institutional setting, and methods implemented are first discussed, followed by the results and policy implications.

Figure 1. Electronic payments as a share of Saudi retail payments



Source: SAMA [3]. Original observations: 2023, 70%; 2024, 79%.

2. LITERATURE REVIEW AND CONEPTUAL FRAMEWORK

2.1 Financial intermediation and productive activity

Financing constraints form the basis of the conceptual argument. Companies need short-term and medium-term capital to invest in equipment, hire labor and expand into new markets. Reliable screening and monitoring can help allocate resources to projects that are likely to be viable, and reliable payment systems can smooth the way of transactions. Financial development could boost productive activity in terms of access, allocation, and efficiency by means of these mechanisms. They also suggest that a lack of financial depth does not necessarily correspond to a lack of productive capacity, because there may be more money that is loaned out that isn't used for a new

expansion of productive capacity but instead for additional consumption.

Cross-country and regional studies generally report a positive but heterogeneous finance–growth relationship, with results depending on inclusion thresholds, governance and financial-system efficiency [7–11]. Digital-finance research similarly links broader access to efficiency and inclusion while noting stability and allocation trade-offs [12–13]. Fintech may change screening, competition and service delivery without eliminating the core economic functions or risks of banking [14–15].

Using annual observations from 1985 to 2018 and a cointegration approach, Haque [5] investigates Saudi private-sector development. This is in contrast to the reported negative money-supply relationship and the insignificantly sized market-capitalization relationship. This separation is an incentive for a banking policy in the current study and a disincentive to replace broad money with wider money. A sample with an FSDP introduction year provides little information regarding subsequent implementation, however. Furthermore, a long-run statistical correlation does not identify which part of a policy package caused it.

2.2 Sectoral finance and diversification

Alzyadat [6] investigates sectoral bank credit and Saudi non-oil growth using annual data for 1970–2019. The study reports an overall positive credit relationship, with differences between sectors and between short- and long-run responses. Its relevance is the allocation question: the economic contribution of finance depends partly on the activities receiving it. The present design therefore treats aggregate private credit as an initial indicator and reserves sectoral decompositions for a limited extension where classifications are consistent. It avoids combining numerous sectoral series in a small quarterly model merely because they are available.

Related Saudi evidence connects inclusive finance with socio-economic development [16], and links ICT diffusion and financial development with women's labor-market participation [17]. Recent research on the sustainability of Saudi financial development also emphasizes institutional and macroeconomic determinants alongside financial depth [18].

The type of investments is also significant. Belloumi and Alshehry [19] analyze domestic investment, foreign direct investment and growth for the period of 1970-2015, based on cointegration analysis and alternative estimation techniques. They examine investment relationships in a broader development framework that is Saudi's and predates the FSDP. It allows for a study of finance as a means of funding investments instead of taking financial growth for granted as a direct source of output. Investment is thus a possible transmission factor for the current program evaluation. When controlling for investment automatically, a portion of the pathway being studied could be eliminated.

2.3 Oil dependence and financial stability

Alabdulwahab [4] highlights the limitations of distinguishing between oil and non-oil accounting categories to present a picture of economic independence from oil. These reported cointegration and short-run dynamics are consistent with ongoing linkages. For this reason, non-oil

growth is a suitable measure of the study but not sufficient to measure successful diversification. Several factors in the oil market, financial conditions and fiscal changes offer conflicting interpretations of the movements observed. Care needs to be taken with their inclusion as some domestic financial and fiscal responses may be affected by the broader reform program.

According to the financial stability assessment done by the IMF [2], the overall condition of the banking system is fairly good, but funding concentration and increased credit and real estate exposures are listed as monitoring issues. This evidence brings in a new qualification to the growth mechanism. The vulnerabilities resulting from the maturity structure, funding base, and sectoral concentration of finance that supports investment can also build up. The study thus considers the composition of credit growth, where possible, in addition to the overall speed of growth, and does not equate the success of a program solely with an increase in credit. The financial stability indicators give context, not another big bundle of regressors.

2.4 Prediction and program evaluation

Athey and Imbens [20] differentiate between prediction and economic questions that call for causal identification. A flexible algorithm can produce accurate estimates of conditional relationships without dealing with omitted variables or reverse causality issues. Goulet Coulombe et al. [21] demonstrate that nonlinearities can be beneficial in macroeconomic forecasting, especially during varying financial circumstances. These contributions inspire a limited machine learning extension, but not a complex algorithm that will profit from a short Saudi quarterly dataset. Predictive value has to be proven using observations not used to train the simpler models.

Broader forecasting evidence is deliberately mixed: large comparisons warn that machine learning does not automatically outperform statistical methods [22], whereas data-rich macroeconomic applications show gains in some settings [23]. If tree models are interpreted using SHAP, the resulting importance explanations describe model behavior rather than policy causality [24].

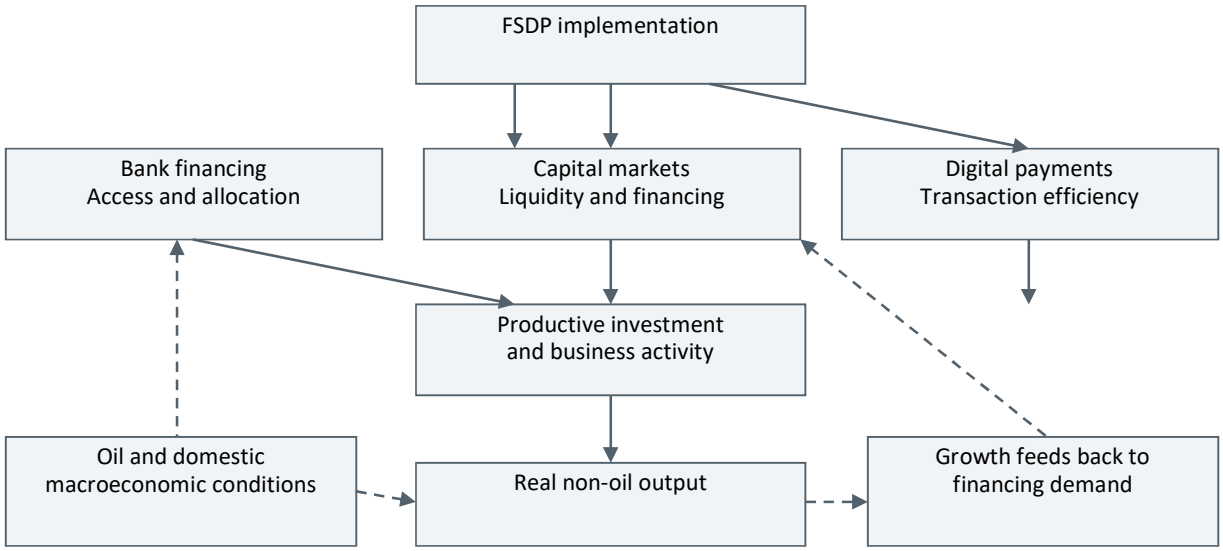
Validation is especially important when observations are ordered in time. Cerqueira et al. [25] find that preserving temporal order is valuable when assessing forecasting performance in non-stationary settings. Accordingly, the comparison uses expanding training windows and subsequent observations for evaluation, with no random train–test split. Even a forecasting improvement would establish additional predictive information, not an FSDP treatment effect. Likewise, variable importance measures describe model behavior and cannot determine the contribution of a government program to GDP. The analysis therefore concentrates on observed forecast errors and the incremental value of financial predictors rather than causal interpretations of algorithmic importance.

2.5 Research framework

Program implementation is linked to banking access and access to capital markets, as well as the efficiency of payments, which can facilitate investment and non-oil activity, as shown in Figure 2. It also cares for macroeconomic conditions that impact finance and output and feedback from

growth to financing demand. The framework gives rise to three hypotheses: H1 – that the growth of real private-credit increases is positively correlated with the growth of non-oil in the future; H2 – that an increase in capital-market trading activity is positively correlated with the growth of non-oil in the future; and H3 – that the conditional growth pattern changes upon the introduction of the FSDP. Predictive improvement using machine learning is not a given, but rather an evaluation question. These mechanisms are used to select the variables, but they are not all clearly expressed in the aggregated data.

Figure 2. Financial transmission framework



Source: Author's framework.

3. INSTITUTIONAL BACKGROUD

The FSDP is examined not as a single, sudden treatment, but as a program that is carried out with various financial measures. The decision to approve and introduce was announced by the SAMA on 9 May 2018. The announcement enumerates three pillars, namely financial institutions, capital markets and financial planning, in addition to 42 initiatives [1]. Adoption of payments services and development of capital markets can affect the ability to finance at different rates; investment or productivity might lag behind. A time indicator can thus reflect around the time of implementation rather than indicating policy intensity or the cost of the policy implementation.

The electronic share of retail payments grew by nine percentage points from 2023 to 2024, according to SAMA's payment announcement. That corresponds to an increase in transactions of about 16.7%, based on the published data numbers for 10.8 and 12.6 billion observations. These are two different aspects: one is the composition of the payment and the other is the frequency of the payment. Both are not a monetary amount. The percentage definition is maintained in Figure 1, as

does the data within it that is embedded in the figure and can be edited. They represent a program-relevant channel but are not sufficient in number to be an evaluation time series [3].

Complementary information on market activity and development is available from the Capital Market Authority's annual reports and statistical bulletins [26–27]. The sale of securities in the secondary market does not equal the market capitalization and funds raised. Trading reflects liquidity and participation, while new issuance is a more direct measure of financing raised. The analysis is based on a fixed definition of turnover ratio for the quarters available in the history and not on the value of trading as a source of new financing. Explicit sensitivity checks are performed on large listings and for changes in denominator. Mechanical changes in market size may otherwise be confused with general progress in financing access.

Interpretation also involves keeping an eye on the dynamic shifts in tax, public investment, labor-market conditions, and pandemic restrictions. They cannot be individually assigned to any one country for a single-country break. The design thus reports on conditional changes around the FSDP, tests alternative timing assumptions and looks at sensitivity to major disruptions. Defensible comparison units or differential policy exposure would be needed for a stronger causal extension. Here, following Abadie [28], it is not assumed that it is feasible to create a synthetic-control design after establishing comparable outcome definitions, credible pre-intervention fit, and a suitable donor pool.

4. DATA AND METHODOLOGY

4.1 Design and source selection

The study is a quantitative, longitudinal study. The window collected covers the period 2010Q1-2025Q4, or a total of 64 calendar quarters before transformations and lags. This is the reference restriction, NOT the dates of the historical observations; data prior to 2018 will be required for comparison. Model samples will employ the common comparable set of series they are required to model. There are a total of 62 observations in the policy baseline, 50 in the lagged-credit sample, and 41 in the market-level comparison; exact dates are reported in Table 2. No annual values are added to make up for missing observations.

GDP and consumer prices are obtained directly from the General Authority for Statistics (GASTAT) [29–31], while market measures are obtained from CMA's original statistical files [27]. Banking and repo series were created by SAMA and obtained from the KAPSARC data portal [32–33], which is listed as the distributor and not the series producer. The U.S. Energy Information Administration (EIA) provides the daily Brent quotation [34]. GDP observations for 2023–2025 are preliminary, whereas CPI is based on the July 2026 vintage. The definitions and sources of ownership implemented are shown in Table 1.

Table 1. Implemented variables and original source ownership

Measure	Operational definition	Original source
Non-oil growth	Quarterly log change in seasonally adjusted real non-oil GDP; primary outcome.	GASTAT national accounts
Private credit	Quarter-end total private claims deflated by quarter-end CPI; quarterly log change.	SAMA, via KAPSARC; GASTAT CPI
Market turnover	Quarterly traded value divided by end-quarter Main Market capitalization, multiplied by 100.	CMA statistical bulletins
Oil conditions	Quarterly mean of available daily Brent prices; quarterly log change.	U.S. EIA Brent daily series
Interest rate	Mean of three month-end repo rates; quarterly change in percentage points.	SAMA, via KAPSARC
Policy timing	D = 1 from 2018Q2; R = quarters since introduction; specified date sensitivities.	SAMA announcement, 9 May 2018
Payment adoption	Annual electronic retail-payment share and transaction count; descriptive only.	SAMA announcement, 15 April 2025

Source: Author's compilation from GASTAT [29]–[31], SAMA [1], [3], SAMA via KAPSARC [32], [33], CMA [27], and EIA [34].
Notes: KAPSARC is the distributor of SAMA-produced series, not the producer.

Original files, table & cell references, frequency, units, release and retrieval dates, and revision status are preserved. It has an extraction report and variable sheets to record the extraction, exclusions and transformations. Realized observations do not include published targets and forecasts. The gaps between financial models are not filled synthetically, and a consistent history of the current CPI can be found starting in 2013, while usable CMA market observations start in 2015Q3. SAMA payment observations are not contextualized. Sources and model inputs are also stored with the statistical scripts, which enables every reported estimate to be identified as coming from a given series.

4.2 Measurement and harmonization

The main result is the growth of the officially seasonally adjusted real non-oil activities measured at the quarterly level, $g(t) = 100 \times [\ln Y(t) - \ln Y(t-1)]$. This is an approximate 'no-annualization' growth rate. Government activities are not included in the non-oil component, but rather reported separately. Volume components are not rebuilt by subtracting one from the other, as additivity is not guaranteed. One vintage of the national accounts has been revised to cover the historical sample and does not involve a splice between old and revised GDP series [29–30]. Outcome sensitivities are based on conventional measures of percentage growth and aggregate GDP.

The banking measure is SAMA's liabilities to the private sector and includes bank credit and investments in private securities. Quarter-end nominal claims are deflated by the corresponding quarter-end CPI, and the log difference between quarter ends is the difference in the level of real claims. Bank credit excluding such securities is a separate sensitivity. Turnover is calculated as the value of the trades during the quarter divided by the end-quarter capitalization in the Main Market multiplied by 100. The denominator is the one used instead of the originally intended mean month-

end capitalization due to the availability of usable sources of coverage. It is based on the proportion of trading to the market size, rather than the capital raised by the new investment; a more common denominator is free-float capitalization.

Quarter-end values are used for monthly stocks. The repo regressor is a change in the percentage points on the average of three published policy-repo rates per quarter and does not represent an average of the policy rates over the course of the day. Brent is the mean of available daily quotations, as a log change, quarterly. The dates were checked for alignment in the calendar, duplicate dates, and missing values and units prior to the estimation. All the days on the full calendar are marked as Lags (before any exclusion). Seasonal diagnostics create a motivation for seasonal indicators in the principal credit interpretation. Observations of COVID-19 continue to be visible, and all available within-experiment comparison dates are the same.

4.3 Policy timing and the estimand

The main intervention date is 2018Q2, which is the date of the program announcement by SAMA [1] (9 May 2018). The value of D(t) is 0 before that quarter and 1 after that quarter. R(t) is 0 before and during the introduction quarter. Timing sensitivities are for 2018Q3 and 2019Q1, to consider implementation delays. These dates have been a part of the design prior to checking coefficient significance. Tests are two-sided because the positive program objective does not warrant including zero or adverse estimated changes. The indicator does not represent expenditure, intensity of implementation or exposure level per firm.

The key estimand is a conditional change of the non-oil growth path around the introduction. It must be understood as causal (as opposed to just another piece of evidence in a story of change), which implies that no other concurrent change accounts for the estimated break and that the pre-intervention relationship is still a valid comparison. In this context these assumptions are pushing. It is therefore not an assessment of experimentally identified impact but an interrupted time-series assessment. The initial specification does not include financial indicators since they can mediate the program effects. A second specification looks at financial associations, with a conditional interpretation of its policy coefficient.

4.4 Econometric specification

The baseline growth equation is:

$$g(t) = \alpha + \varphi g(t-1) + \beta t + \delta D(t) + \kappa R(t) + \theta' Z(t) + \varepsilon(t)$$

The terms are: Z(t) is a set of variables including Brent log growth and the difference between the quarterly mean repo rate, along with separate indicators for 2020Q2 and 2020Q3. The movements of oil prices are not taken as exogenous to Saudi activity in full. It does not therefore constitute identification to condition on these. The coefficient d indicates the conditional immediate change in the quarterly growth, and the coefficient κ describes the trend in the change of the quarterly growth, but the change itself does not describe a gain in the GDP level. Reported recursive comparison is a difference in growing path with pointwise uncertainty, assuming the macro controls to be constant and extrapolating the earlier trend. There is no assumption or implication of cumulative GDP gain.

A joint test of $\delta = \kappa = 0$ assesses the absence of a program-period break. Dynamic responses also depend on lagged growth, so fitted trajectories are recursive. Financial models add lagged real claims growth and turnover separately and jointly. An exploratory post-introduction interaction with lagged credit assesses a change in association. AR(1) and AR(2) are compared by BIC on common observations. ADF and KPSS tests, residual-correlation, heteroskedasticity, stability and collinearity diagnostics guide interpretation. Following those checks, the principal credit model includes seasons. An exploratory market reformulation differences turnover and removes the ramp while retaining trend, post level and seasonal controls; it changes the estimand and is not a retest of the original policy-break hypothesis.

Inference is based on the HAC intervals which are calculated using a Bartlett kernel, finite-sample covariance correction, and t critical values, with bandwidth equal to 4. An approximate F test is used for joint restrictions. The sensitivity intervals are based on 1,999 circular centered-residual draws of length 4, where the draws are based on recursive updates of the growth to take into account lagged growth and exogenous variables are fixed. These are intervals with fixed and conditional models. The focus is on interval size, interval height/sensitivity and not just significance. It is reported transparently what is an alternative outcome, when it is experienced, how long there was a lag, when it was seasonal, and what the bandwidth was during the pandemic. Many exploratory checks are not treated as independent checks, and the p-values for many exploratory checks are not adjusted for multiple testing.

4.5 Machine-learning extension

Machine learning supplies a supplementary test of predictive information. The ridge regression and shallow random forest are compared against AR(2) and historic mean forecast. The coefficient instability caused by correlated predictors is restricted by ridge shrinkage, and modest nonlinearities are allowed by a restricted forest. Excluding deep neural networks is a result of the limited basis for complexity that is provided by 64 calendar quarters. The design is similar to that of Athey and Imbens [20], which separates prediction from identification. It questions whether these specific algorithms and available financial measures will help forecast, rather than take it for granted that flexibility is good or that a forecast gain is program effectiveness.

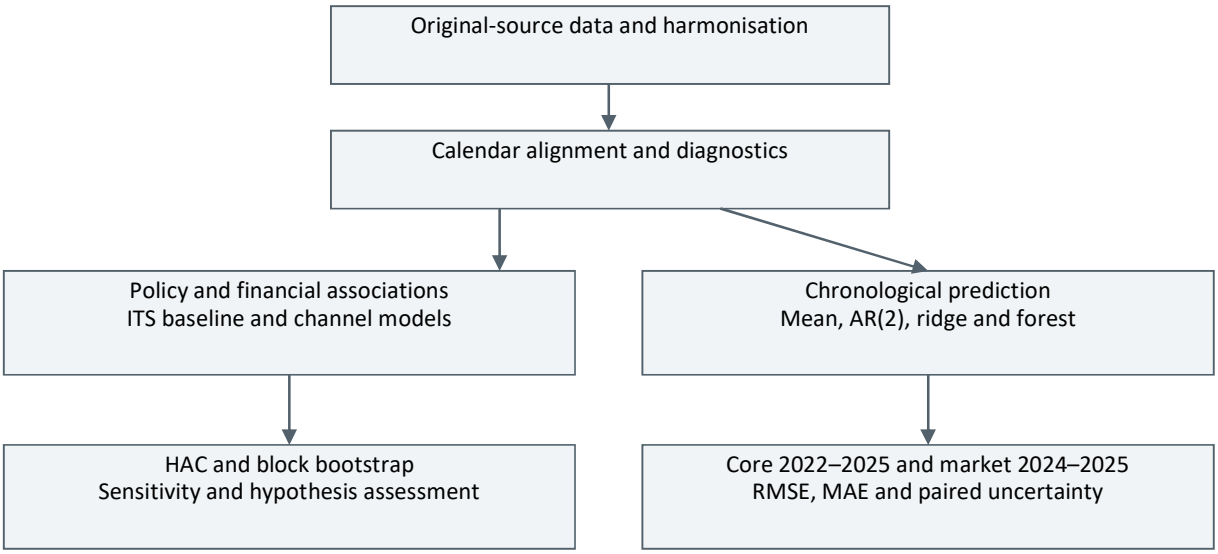
In the macro feature set, growth in t-1 and t-2 are included, as is one-quarter lag of Brent growth and repo changes. Lagged real claims growth and lagged turnover for credit adds. They are set as ridge penalties, which are selected from the set {0.01, 0.1, 1, 10, 100} and standardized within each training window. Forest candidates have 500 trees, a depth of 2 or 3, and at least 8 or 12 observations per tree. The values for the selected settings are fixed at validation time and fixed in the test time. Features are specified without full-sample outcome screening. All comparisons of the forest seed are based on a fixed year of 20302018, and common observations are used for all benchmark comparisons.

Both forecasting experiments begin with 28 observations available for training. The targets will be covered by core validation for 6 targets during 2020Q3-2021Q4 and 16 test targets during 2022Q1-2025Q4. The 2021Q4 schedule is infeasible because there are only 25 observations

available. Its revision is based on coverage with 28 initial rows up to 2022Q3, 5 validation targets from 2022Q4 to 2023Q4 and 8 test targets from 2024Q1 to 2025Q4. The calendar used was fixed prior to the evaluation of the forecast. Only previous targets are used for each expansion of the refit. The latter one is a market test of an exploratory nature, and the magnitudes of its errors are not directly compared with the various core periods [25].

RMSE and MAE are computed for predictions made on the same dates in the experiments. Growth can be either zero or negative, with no percentage errors. Comparisons with and without financial features are used for incremental information. Five hundred ninety-nine (4,999) paired circular block resamples of test errors are used for the differences with RMSE against AR(2) with block length 2. Their intervals will still be descriptive with the brief tests. This is not real-time forecasting and will not be updated to include historical release vintages. There are also publication delays that are not modelled, which further restrict operational interpretation. The analysis presents overall accuracy, paths of forecast, absolute errors and feature comparisons within the experiments without any identifying a policy treatment effect.

Figure 3. Implemented estimation and prediction workflow



Source: Author's implemented design.

4.6 Robustness and reproducibility

Robustness checks include delayed intervention dates, pandemic-quarter exclusions, alternative definitions of the credit and turnover, seasonality and lag order. Lags do not fill calendar gaps when building them, and exclusion regressions did not fill the calendar gaps. Outcome sensitivity is measured using an aggregate GDP model. Attention is paid to shifts in the market denominator as well as large listings and influential observations with baseline estimates that exclude the latter. Changes in the market denominator, large listings and influential observations are treated separately, including leave-one-observation-out baseline estimates. The results whose

dependence relies on a definition or a window are considered fragile. Sectoral decompositions and synthetic control were deferred because consistent sectoral histories and a defensible donor pool were unavailable [28].

Published aggregate data (not participant recruitment or identifiable personal information) is used. The reproducibility package includes the original data files, the source workbook, variable definitions, extraction and transformation records, scripts, random seeds, model outputs and software requirements. Python, using statsmodels and scikit-learn, is used for statistical estimation, and derived scores are independently compared with the forecast records. Original observations and author transformations and model estimates are distinguishable. The final design allows for an unambiguous evaluation of conditional changes over the course of the program period and maintains a more robust causal language for policy exposure and a plausible counterfactual.

5. RESULTS AND ANALYSIS

The analysis reveals a conditional shift in the growth path of the non-oil sector starting around the time of the introduction of FSDP, without revealing a causal FSDP program effect. The preponderance of baseline evidence is in the change in trend, not in a step increase in growth. Financial-channel estimates are far from accurate, and the machine-learning comparison isn't convincing of a significant benefit over basic forecasts. The three research questions are individually answered in three policy exercises, association exercises and prediction exercises, where the distinctions made in the methodology are maintained.

5.1 Implemented design and usable observations

The collected dataset contains 64 calendar quarters from 2010Q1 to 2025Q4. GDP uses GASTAT's Q2 2026 national-accounts vintage, including its preliminary 2023–2025 observations; CPI uses the July 2026 historical series. The analysis therefore examines revised historical outcomes rather than information available to forecasters at each historical date. Non-oil activities exclude the separately reported government component. Credit and repo observations originate from SAMA and were retrieved through KAPSARC. Market observations come directly from CMA, and Brent quotations from the U.S. EIA. The original producers, retrieval routes, source cells and vintages are documented in the dataset [27, 30–34].

Three measurement decisions qualify the original plan. First, a consistent current CPI history begins in 2013, shortening the real-credit sample. Second, CMA's usable quarterly market history begins in 2015Q3. Turnover is consequently defined as quarterly traded value divided by end-quarter market capitalization, rather than the proposed mean-month-end denominator. Third, the policy-rate regressor averages three published month-end repo rates, while Brent is the average of available daily quotations. These are explicit operational definitions, not interchangeable measures. Missing values were not interpolated, and lags were constructed on the full calendar before any exclusions.

Table 2. Samples used in estimation and prediction

Exercise	Usable dates	N	Before / from 2018Q2
Policy baseline	2010Q3–2025Q4	62	31 / 31
Credit channel	2013Q3–2025Q4	50	19 / 31
Market levels comparison	2015Q4–2025Q4	41	10 / 31
Market changes reformulation	2016Q1–2025Q4	40	9 / 31
Core forecast test	2022Q1–2025Q4	16	N/A
Market forecast test	2024Q1–2025Q4	8	N/A

Source: Author's compilation from GASTAT [29]–[31], SAMA via KAPSARC [32], [33], CMA [27], and EIA [34] data. Notes: Usable dates reflect data availability after lags and differencing.

Table 2 establishes the actual information available for each question. The policy model retains the longest sample because it initially excludes possible financial mediators. Financial models use smaller common samples, and their coefficients cannot be compared with the full baseline as if only one regressor had changed. Inference uses heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors with a Bartlett kernel, bandwidth four, a finite-sample covariance correction and t-based intervals. Reported p values are two-sided and unadjusted for multiple exploratory checks. Circular residual block bootstraps use 1,999 draws and four-observation blocks, recursively updating lagged growth while holding the observed exogenous variables fixed.

5.2 Descriptive patterns

Table 3. Descriptive statistics and period means

Variable	N	Mean	SD	Before mean	From mean
Non-oil log growth (%)	63	1.42	1.81	1.45	1.40
Real claims log growth (%)	51	1.75	1.75	1.10	2.17
Turnover (%)	42	8.37	5.88	16.15	5.61
Brent log growth (%)	63	-0.29	14.95	-0.42	-0.16
Repo change (pp)	63	0.04	0.34	0.00	0.08

Source: Author's calculations using GASTAT [29]–[31], SAMA via KAPSARC [32], [33], CMA [27], and EIA [34] data. Notes: Before and from refer to 2018Q2. Period samples differ by variable: growth, Brent and repo 32/31; real claims 20/31; turnover 11/31. Percent log changes are not annualized. pp denotes percentage points.

Average non-oil growth is 1.425% per quarter, with a standard deviation of 1.811 percentage points. Its pre-program mean of 1.450% is slightly higher than the subsequent mean of 1.399%. A simple before–after average therefore does not demonstrate acceleration. The post-introduction period is substantially more volatile, including the 2020Q2 contraction of 7.610% and 2020Q3 rebound of 7.650% in log-growth terms. Figure 4 makes these disruptions visible rather than suppressing them through winsorization.

Figure 4. Observed quarterly non-oil growth



Source: Author calculations using GASTAT real non-oil activities. Values are 100 times the quarterly log change, without annualization. FSDP introduction is coded from 2018Q2. Observed movements are not treatment effects.

Real private-sector claims growth rises from a pre-introduction average of 1.104% to 2.169% afterwards. This expansion concerns the total claims measure, which includes investments in private securities as well as bank credit. It does not establish that the additional finance reached productive investment. Turnover falls from 16.145% to 5.615%, but the comparison is especially sensitive to the short pre-period and capitalization denominator. Market capitalization jumps from approximately SAR 1.91 trillion in 2019Q3 to SAR 9.03 trillion in 2019Q4. A lower turnover ratio can therefore coexist with a much larger market and must not be interpreted directly as deteriorating financial access. Table 3 provides context for estimation, not a causal comparison.

5.3 Research question one and the program-period growth break

The baseline interrupted time-series model includes lagged non-oil growth, a linear trend, the post-2018Q2 indicator, a post-introduction ramp, contemporaneous Brent and repo changes, and separate 2020Q2 and 2020Q3 indicators. A common-sample BIC comparison favors one growth lag: 203.009 for AR(1) against 204.804 for AR(2), both using 61 observations. The selected AR(1) specification then uses all 62 available observations. This comparison selects a parsimonious equation; it does not make its break date or causal interpretation empirically proven.

Table 4. Baseline interrupted time-series estimates

Term	Estimate	HAC SE	95% interval	p value
Constant	2.689	0.449	[1.788, 3.589]	<0.001
Lagged non-oil growth	0.038	0.090	[-0.143, 0.219]	0.674
Pre-introduction trend	-0.071	0.017	[-0.106, -0.037]	<0.001
Immediate growth shift	0.704	0.487	[-0.273, 1.681]	0.154
Change in growth trend	0.094	0.023	[0.047, 0.141]	<0.001
Brent growth	0.028	0.019	[-0.011, 0.066]	0.153
Repo-rate change	0.997	0.387	[0.221, 1.773]	0.013
2020Q2 indicator	-6.369	1.309	[-8.995, -3.742]	<0.001
2020Q3 indicator	5.757	1.088	[3.574, 7.940]	<0.001

Source: Author's estimations using GASTAT [29], [30], SAMA via KAPSARC [33], and EIA [34] data. Notes: N = 62; R^2 = 0.732; adjusted R^2 = 0.692. Joint test of immediate shift and trend change: F(2,53) = 9.571, p = 0.000283. Growth and oil variables are in percent log-change units; the repo variable is in percentage points.

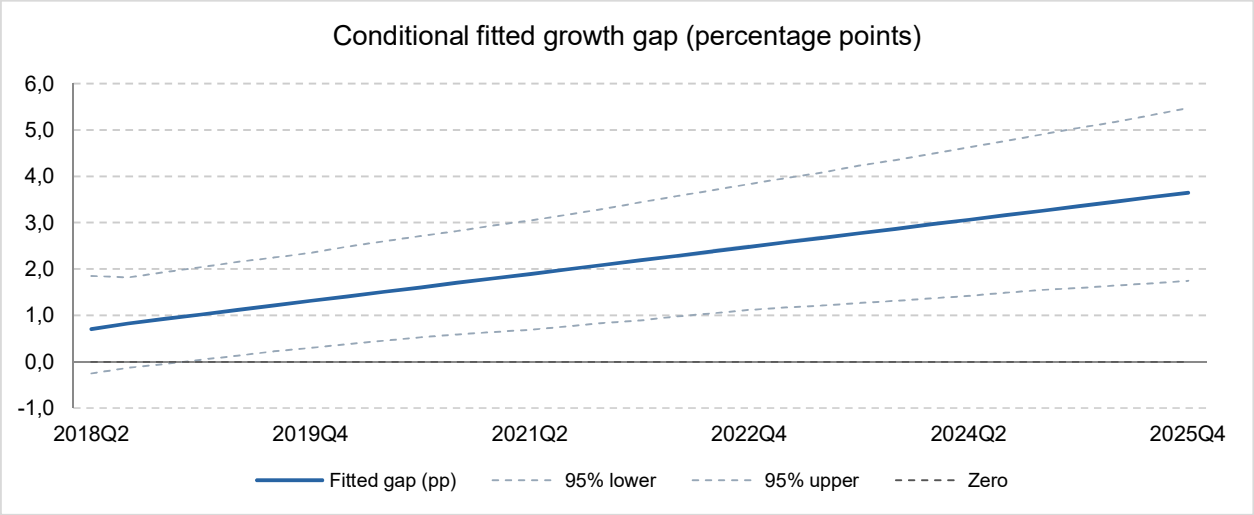
Table 4 gives an immediate conditional shift of 0.704 percentage points, with a 95% interval from −0.273 to 1.681 and p = 0.154. The data do not establish an immediate increase. The ramp coefficient is 0.0937 percentage points per elapsed quarter, with an interval from 0.0467 to 0.1407. It measures a change in the slope of quarterly growth, not 9.37% annual growth or a direct GDP-level gain. The pre-introduction slope is −0.0715, while the sum of pre-trend and ramp coefficients is approximately 0.0222. The model describes an attenuation of the earlier declining growth trajectory.

The joint null of no immediate shift and no trend change is rejected in the selected baseline. H3 consequently receives conditional support in this specification. Bootstrap intervals reinforce the distinction: the immediate shift interval is [−0.253, 1.854], whereas the ramp interval is [0.043, 0.160]. The positive repo coefficient of 0.997 is an association conditional on the equation. It is not evidence that monetary tightening raises output; contemporaneous policy rates may move with macroeconomic conditions that are incompletely captured here. The same caution applies to the oil-price control.

Figure 5 translates the break coefficients into a recursive growth-path difference. With the initial difference set to zero, the path incorporates the estimated growth persistence, immediate shift and ramp. The fitted gap reaches 3.650 percentage points in 2025Q4, with a pointwise conditional bootstrap interval of [1.744, 5.479]. This large difference arises partly because the comparison extrapolates the earlier negative trend for another 31 quarters. It is a model-dependent contrast between fitted growth paths, not observed additional GDP attributable to FSDP. The intervals exclude uncertainty about omitted reforms, alternative counterfactual trends and model selection. No cumulative output-gain claim is made.

$$\Delta g(h) = \varphi \Delta g(h-1) + \delta + \kappa h, \qquad h = 0, \ldots, 30;\ \Delta g(-1) = 0$$

Figure 5. Conditional fitted growth-path difference



Source: Author model simulation. Solid line: baseline fitted difference. Dashed bounds: pointwise 95% residual block-bootstrap intervals. The comparison holds macro controls fixed and continues the earlier trend. It is not a causally identified FSDP effect.

5.4 Diagnostics and admissible interpretations

Augmented Dickey–Fuller tests, with at most four lags selected by BIC, reject a unit root for non-oil growth, real-claims growth and Brent growth (all $p < 0.001$). Their KPSS statistics do not reject level stationarity at the available 10% bound. Repo-rate changes are less decisive: ADF $p = 0.106$ and KPSS $p \geq 0.100$. These mixed results do not justify declaring every series stationary or estimating an unrestricted long-run system. The analysis retains the already differenced rate as a conditional control and avoids cointegration or long-run multiplier claims.

Turnover presents a more serious qualification. The intercept-only ADF test rejects a unit root ($p = 0.020$), but KPSS rejects level stationarity ($p \leq 0.010$). With a deterministic trend, ADF $p = 0.093$ and KPSS still rejects trend stationarity. These conflicting diagnostics, together with the capitalization break, make a stable turnover-level relationship doubtful. Credit also displays seasonal variation after trend and post-period controls (joint HAC $p = 0.000111$), whereas turnover's seasonal test is not decisive ($p = 0.358$). Seasonal indicators are therefore retained in the principal credit interpretation and in the market reformulation.

The full policy baseline shows no rejection in the reported residual tests: Breusch–Godfrey $p = 0.653$ at lag two and 0.507 at lag four; Breusch–Pagan $p = 0.399$; and CUSUM $p = 0.999$. These diagnostics provide no obvious residual warning within this specification, but cannot verify the policy counterfactual. The seasonal credit model has lag-four residual-correlation $p = 0.141$ and mild non-normality (Jarque–Bera $p = 0.048$), reinforcing the value of HAC and bootstrap comparisons.

The joint market-level model is not treated as a reliable confirmatory equation. Its trend and ramp variance-inflation factors are approximately 131 and 85 because only ten pre-introduction observations remain. Adding seasonal controls produces lag-four residual-correlation $p = 0.037$; a

second growth lag does not resolve that warning. Following the methodology's requirement to reformulate rather than assert a stable relationship, an exploratory market equation uses lagged turnover changes, retains trend and post level, and omits the ramp. It also retains credit, macroeconomic and seasonal controls. This reduces the maximum VIF to 3.20, with lag-four residual-correlation p = 0.081. The altered equation addresses short-run financial associations and cannot be used as another test of the original H3 estimand.

5.5 Research question two and financial channels

Table 5. Lagged financial associations with non-oil growth

Specification and term	N	Estimate and 95% HAC interval	p value
Credit without seasons	50	0.032 [-0.137, 0.201]	0.703
Credit with seasons	50	0.097 [-0.141, 0.335]	0.415
Turnover level only	41	-0.008 [-0.200, 0.184]	0.932
Joint levels: credit	41	0.065 [-0.075, 0.206]	0.351
Joint levels: turnover	41	-0.004 [-0.203, 0.195]	0.968
Reformulated: credit	40	0.133 [-0.023, 0.290]	0.092
Reformulated: turnover change	40	0.058 [-0.108, 0.224]	0.481

Source: Author's estimations using GASTAT [29]–[31], SAMA via KAPSARC [32], and CMA [27] data. Notes: Credit and turnover lead growth by one quarter. Turnover is in percentage points, so coefficients reflect a one-point change in the ratio. Joint-level rows are diagnostic sensitivities.

The original credit specification produces a coefficient of 0.032, but the seasonal specification is more appropriate after the seasonality test. Its coefficient is 0.097: one percentage point faster real-claims growth in the previous quarter is associated with approximately 0.097 percentage points higher current non-oil growth. The interval [−0.141, 0.335] is wide and includes negative values, and p = 0.415. Its bootstrap interval [−0.112, 0.272] also crosses zero. H1 is therefore not established. This conclusion means that the available aggregate sample does not distinguish the hypothesized relationship precisely; it does not demonstrate that credit has no economic role.

On the common 41-quarter market sample, the turnover-only coefficient is −0.008, with an interval [−0.200, 0.184]. Joint inclusion gives credit 0.065 and turnover −0.004, also imprecisely estimated. Table 5 retains these results to show why a positive market-channel claim is not justified, while the diagnostics prevent using the joint equation as the principal evidence. The reformulated model yields a turnover-change coefficient of 0.058, interval [−0.108, 0.224], p = 0.481. Its bootstrap interval [−0.019, 0.143] likewise crosses zero. Since that coefficient concerns a change rather than the turnover level, it is complementary evidence rather than a literal retest of H2. The positive turnover-level hypothesis remains unsupported.

The reformulated credit coefficient is 0.133 (p = 0.092), but its 95% HAC interval [−0.023, 0.290] and bootstrap interval [−0.037, 0.303] include zero. A second-lag version reduces it to 0.016 and reintroduces residual-correlation concerns. There is no stable basis for elevating the favorable single estimate above the other results. The seasonal post-introduction credit interaction is also inconclusive: 0.040, interval [−0.399, 0.479], p = 0.855. Thus, the data do not establish that FSDP

changed the credit–growth association. None of these regressions identifies the investment, sectoral allocation or payment-efficiency mechanisms directly.

5.6 Robustness and sensitivity of the policy interpretation

Table 6. Policy-break sensitivity

Specification	N	Immediate shift	Trend change	Joint p
Baseline AR(1)	62	0.704	0.094	<0.001
Introduction 2018Q3	62	0.828	0.090	<0.001
Introduction 2019Q1	62	1.158	0.078	<0.001
Exclude all 2020	58	1.071	0.082	0.003
Add seasonal indicators	62	0.691	0.094	<0.001
Two growth lags	61	0.529	0.070	0.082
Credit common sample	50	0.748	0.102	0.009
Market common sample	41	0.602	0.063	0.616
Conventional GDP growth	62	0.728	0.095	<0.001
Aggregate GDP outcome	62	0.704	0.095	0.235

Source: Author's estimations using GASTAT [29], [30], SAMA via KAPSARC [32], and CMA [27] data. Notes: Joint p tests the immediate shift and trend change together. Common-sample rows restrict estimation to periods with available credit or market data.

Delayed introduction dates retain positive ramp estimates and significant joint tests, as do seasonal controls and conventional percentage growth. Excluding all of 2020 gives an immediate shift of 1.071, ramp 0.0815 and joint p = 0.0029. Excluding only 2020Q2 and 2020Q3 leaves the other coefficients unchanged because their separate indicators already absorb those observations in the baseline. That exact agreement is a mechanical property of the specification, not an independent replication. HAC bandwidths of two and eight leave the point estimates unchanged; the immediate-shift interval still crosses zero, while the ramp interval remains positive.

Other checks materially weaken the claim. With two growth lags, joint p rises to 0.082 even though the ramp's individual p value is 0.036. The baseline restricted to the credit sample has joint p = 0.009, but restriction to the much shorter market sample gives p = 0.616 before financial variables are added. Loss of evidence in the market equations therefore cannot be interpreted as financial mediation of an original policy effect. Aggregate GDP gives joint p = 0.235 and a ramp interval spanning zero. The non-oil finding does not automatically extend to the entire Saudi economy. Leave-one-observation-out baseline estimates keep the ramp positive, between 0.068 and 0.108, but this does not resolve dependence on lag order or sample window.

Measurement sensitivities give no consistent positive financial-channel result. Replacing total claims with bank credit excluding private-securities investments yields 0.033 (p = 0.695) in the unseasonalized credit specification. A free-float-capitalization denominator gives turnover 0.008 (p = 0.598). Excluding 2019Q4 and 2020Q1 from the joint-level model produces a negative turnover coefficient, −0.113 (p = 0.037), illustrating instability rather than establishing a harmful market

effect. These numerous exploratory checks are not independent confirmations and were not used to select the most favorable coefficient.

5.7 Research question three and predictive performance

The prediction exercise operationalizes next-quarter forecasting by predicting growth in quarter t from growth in t−1 and t−2, and one-quarter lags of Brent growth, repo changes and the relevant financial measures. No contemporaneous target-quarter predictor enters the forecast. Historical-mean and AR(2) benchmarks are compared with ridge and shallow random forests on identical target dates. Ridge scaling is fitted inside each training window, and hyperparameters are fixed before the test period. Models are then refitted as earlier target observations become available. This preserves chronological evaluation, consistent with the concerns discussed by Cerqueira et al. [25].

The core experiment uses 28 initial training observations, then six validation targets from 2020Q3 through 2021Q4. It retains all 16 planned test targets in 2022Q1–2025Q4. The market experiment has only 25 complete observations by 2021Q4, making the original chronology infeasible with the required initial window. Its explicitly revised sequence uses 28 observations through 2022Q3, five validation targets in 2022Q4–2023Q4 and eight final test targets in 2024Q1–2025Q4. This revision was determined from coverage before forecast evaluation. The later exercise is exploratory, and its smaller errors cannot be compared with the core exercise as evidence that market predictors improve forecasting.

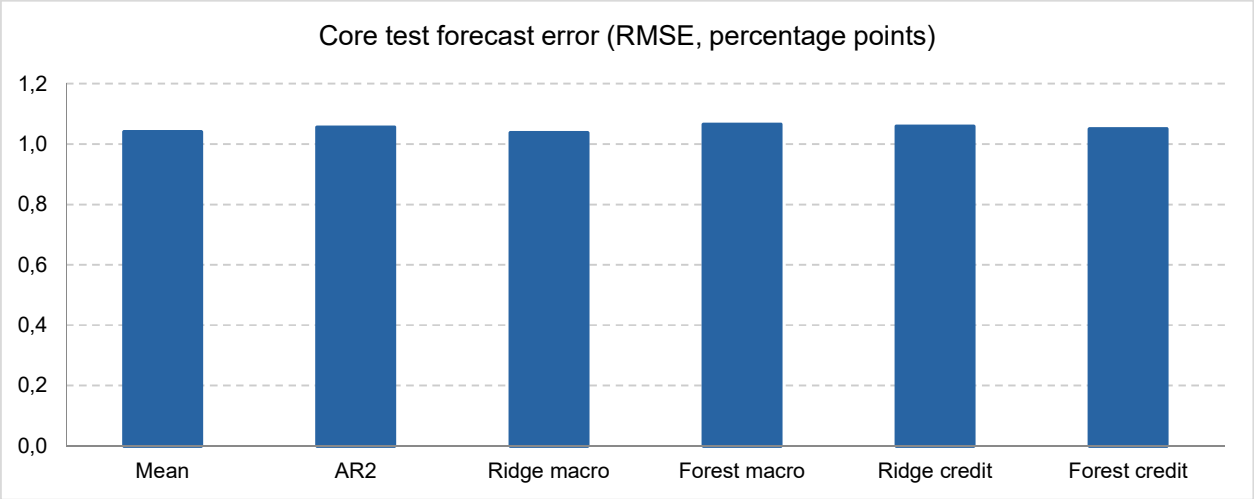
Table 7. Forecast errors on common held-out quarters

Model	Core RMSE	Core MAE	Market RMSE	Market MAE
Historical mean	1.041	0.683	0.512	0.416
AR(2)	1.055	0.691	0.519	0.416
Ridge, macro	1.037	0.688	0.524	0.431
Forest, macro	1.065	0.792	0.637	0.473
Ridge, credit	1.059	0.702	0.530	0.441
Forest, credit	1.050	0.781	0.645	0.565
Ridge, full	—	—	0.525	0.426
Forest, full	—	—	0.616	0.455

Source: Author's retrospective expanding-window forecasts using GASTAT [29]–[31], SAMA via KAPSARC [32], [33], CMA [27], and EIA [34] data. Notes: Errors are percentage points of quarterly log growth. Core test N = 16; market test N = 8. Macro contains two growth lags plus oil and repo lags.

Table 7 shows that ridge with macroeconomic predictors has the lowest core RMSE, 1.037, compared with 1.055 for AR(2) and 1.041 for the historical mean. The gain over AR(2) is only approximately 1.74%. Its paired block-bootstrap RMSE difference is −0.0184, with a 95% interval [−0.0332, 0.0037], crossing zero. The historical mean also has a slightly lower MAE than this ridge model. Adding credit worsens ridge RMSE to 1.059. The credit forest achieves RMSE 1.050, but its difference from AR(2) is −0.0052 with interval [−0.0958, 0.1771]. Figure 6 shows how close these RMSE values are; small point differences do not establish a reliable ranking.

Figure 6. Forecast error in the core test period



Source: Author's calculations using GASTAT [29]–[31], SAMA via KAPSARC [32], [33], and EIA [34] data. Notes: 2022Q1–2025Q4, N = 16 common targets. Lower RMSE is better. Macro uses lagged GDP growth, Brent growth and repo changes; credit additionally includes lagged real-claims growth.

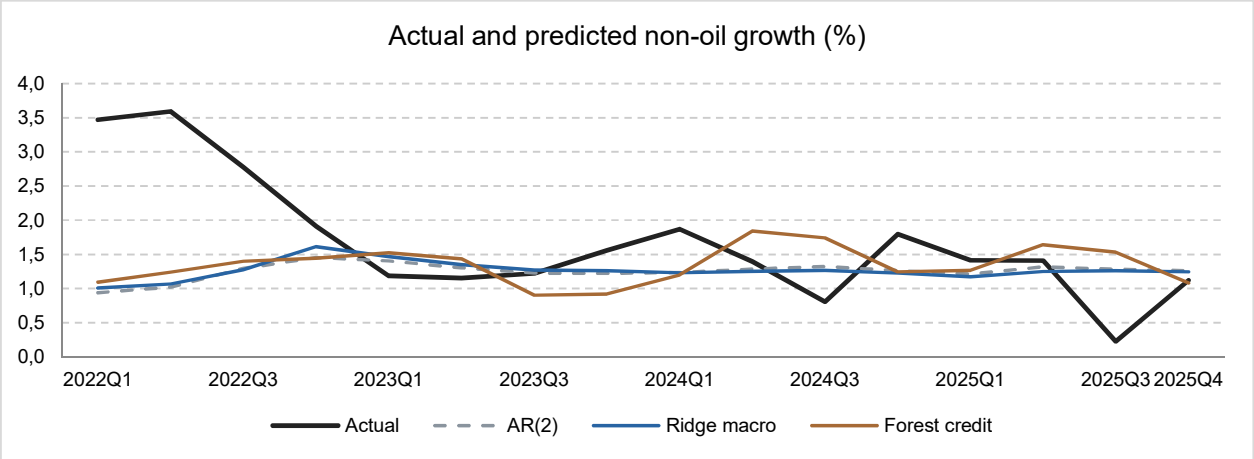
In the market experiment, the historical mean is best on both reported measures: RMSE 0.512 and MAE 0.416. Full ridge has RMSE 0.525 and full forest 0.616. The full-ridge difference from AR(2) is 0.0062 with interval [−0.0295, 0.0361]; there is no demonstrated advantage. Ridge selects penalty 100 in every model. Forests select depth two throughout, with minimum leaf size twelve except for the market credit forest, which selects eight. These settings and the forecast scores are consistent with limited useful complexity in this short sample, not with a need for a deeper model.

Forecast-error uncertainty uses 4,999 paired circular block resamples, length two. With only sixteen or eight test quarters, the intervals are descriptive and cannot guarantee future model rankings. The core validation window also includes pandemic recovery, which may influence settings. GDP revisions and unmodelled publication delays mean these are latest-vintage retrospective comparisons, not real-time forecasts. Accordingly, RQ3 is answered negatively in evidential terms: the tested machine-learning models do not demonstrate a dependable gain over transparent benchmarks, and additional financial predictors do not yield consistent incremental value. Predictive performance would in any case not identify the program's economic impact [20].

5.8 Machine learning forecast patterns and financial information

Figure 7 displays the actual growth series with AR(2), macro ridge and the credit forest on the same sixteen held-out targets. These are illustrative representatives of the benchmark, regularized linear and nonlinear model classes; Table 7 retains every tested specification. The forecasts are considerably smoother than the realized series and underpredict the strong growth observed at the beginning of 2022. The plot therefore locates a source of error concealed by a single average score. It does not constitute a new training exercise, and the representative curves are not selected to assert a statistically reliable winning model.

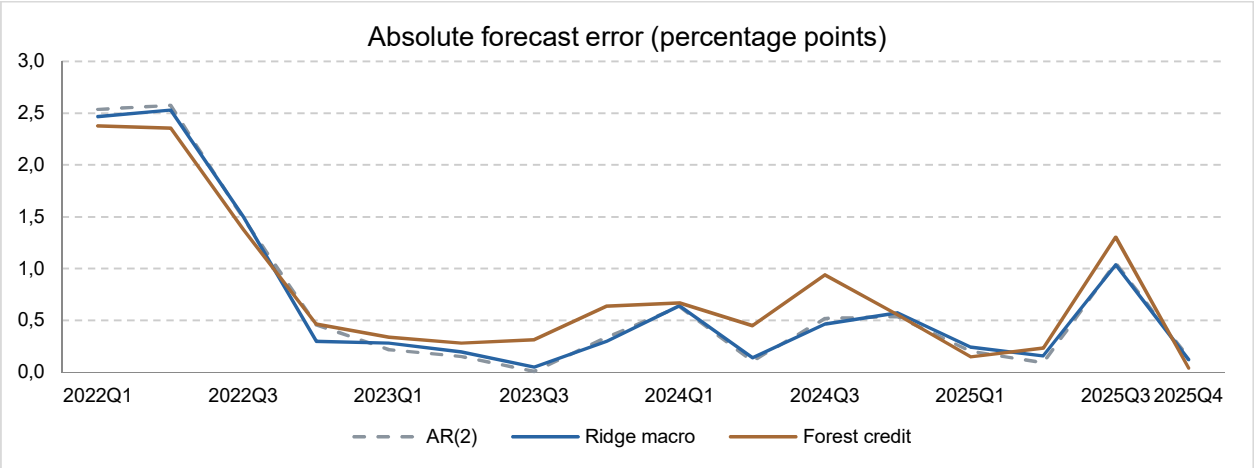
Figure 7. Actual growth and representative machine learning forecasts



Source: Author expanding-window forecasts, core test 2022Q1–2025Q4, N = 16. Growth equals 100 times the quarterly log change. Black is observed growth; dashed grey is AR(2); blue is macro ridge; brown is the credit forest. Predictions use earlier target observations and the latest data vintage.

Figure 8 plots the absolute errors for the same three models. Large errors are concentrated in the early test quarters, while later rankings vary by quarter. A model can fit some fluctuations more closely yet have no dependable aggregate advantage. Because RMSE gives greater weight to larger errors than MAE, those early misses help explain why both criteria are reported. This visual check supports the paired uncertainty analysis rather than replacing it: sixteen observations supply limited evidence about performance across future economic conditions.*/86+

Figure 8. Absolute forecast errors across the core test quarters

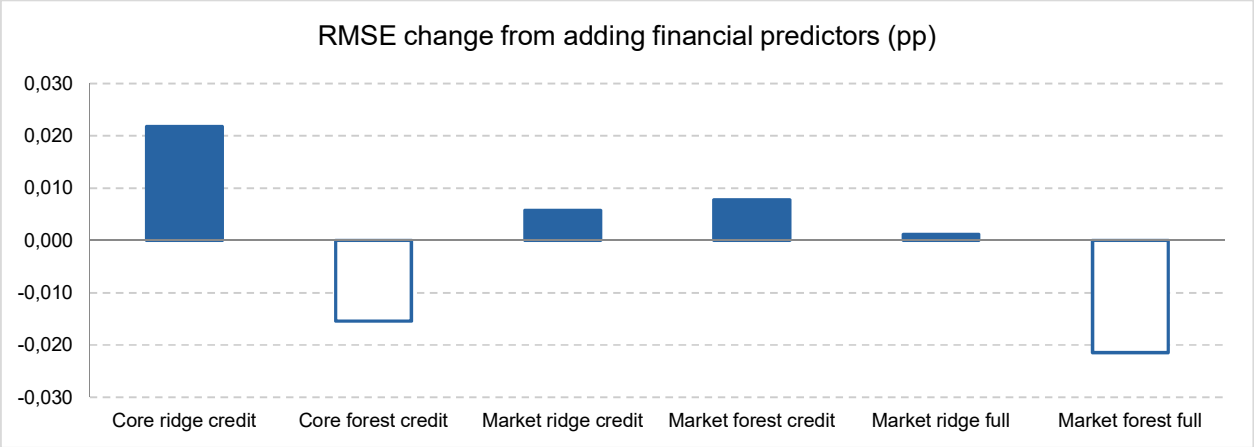


Source: Author calculations from Figure 7 predictions. Absolute error = |actual − predicted|, in percentage points. All sixteen test quarters are shown without deleting difficult targets. Colors and model definitions match Figure 7.

Figure 9 isolates the point change in RMSE from adding financial predictors to each algorithm's macro model. Credit increases core ridge error by 0.0218 percentage points but lowers core forest error by 0.0154. In the market experiment, credit raises ridge and forest errors by 0.0058

and 0.0078 respectively; including both credit and turnover changes them by 0.0011 and −0.0215. These mixed signs support the conclusion that additional financial information has no consistent predictive benefit. Each bar is a within-experiment comparison; differences between the core and market test periods are not interpreted as predictor effects. These bars are descriptive point differences, not significance tests or the earlier bootstrap comparisons against AR(2).

Figure 9. Incremental forecast error from financial predictors



Source: Author calculations from Table 7. Each bar equals financial-model RMSE minus the same algorithm's macro-model RMSE. Negative means a lower error. "Credit" adds real claims growth; "full" adds claims growth and turnover. Core N = 16; market N = 8.

6. DISCUSSION AND ANSWER TO THE RESEARCH OBJECTIVE

The first OBJECTIVE was to determine if non-oil growth was different prior to and following FSDP introduction. The chosen interrupted time-series model allows for a change from the previous trajectory of growth, but the meaning is more restricted than a program-induced expansion. It is unclear what happens immediately; the raw period mean does not increase, and evidence weakens with a second lag and shorter windows. These results can co-exist because a conditional trend contrast addresses a different question than an unconditional before-after mean. Their policy question in the title is thus answered on the basis of conditional evidence, not on the basis of a uniquely identified causal estimate.

The second goal was to explore whether financial and/or market activity is correlated to the growth of non-oil activity after the fact. The average credit point estimates are mostly positive, but they are not sufficiently precise and specific in order to set up H1. The consistency of turnover levels is not great, and the correlation is not sufficient to prove H2. Furthermore, total claims may mask variations in lending allocation, and secondary market trading does not reflect new finance raised. The conceptual framework does not explicitly account for aggregate data constraints, such as investment, payment efficiency and firm financing. These regressions do not prove that those are the channels, but they are still viable options to explore in future research.

The banking estimates are not precise and therefore must be taken with the preceding Saudi literature in mind. Haque [5] and Alzyadat [6] discuss longer-term histories and find sector- and

instrument-specific correlations. These are the current quarterly coefficients of subsequent short-run growth without controls under the program. They do not therefore directly contradict earlier findings in the long run. Rather, from every specification/frequency one cannot deduce a positive financing story for the aggregates. The difference also helps to support the investment-allocation mechanism mentioned by Belloumi and Alshehry [19]: the number of claims does not identify the effectiveness of the projects that are being funded.

The third goal was to see if there is any predictive information in constrained machine learning. There is little improvement in the common-date comparisons, and the simpler historical mean comparison is very competitive. The proper implication is to hold on to the parsimonious forecasting benchmarks in this set. These features, horizons, vintages and short evaluation periods are not a reflection of the overall lack of effectiveness of machine learning in macroeconomics. There is also no justification for generalizing to a nonlinear structure based on tuning parameters on 5 or 6 validation quarters.

Although it is argued by Goulet Coulombe et al. [21] that there are reasons for macroeconomic forecasting to use nonlinear models, the usefulness of them remains an empirical question. In these settings, the forest constraint is more restrictive, and there is little benefit to complexity because of low ridge shrinkage and a short validation period. The practical implication: smooth forecast paths (figures 7 to 9); hard quarters are the main drivers of errors, and financial additions yield mixed changes. This is consistent with the prediction-identification distinction of Athey and Imbens [20]. Neither accurate predictions nor a model's apparent use of credit would establish that FSDP caused the associated activity.

Table 8. Explicit answers to the research questions and hypotheses

Question and objective	Evidence-based answer	Hypothesis outcome
RQ1: Change in non-oil growth level or trend around introduction	Baseline supports a conditional trend break. Immediate increase is uncertain; sample and lag sensitivities weaken generalization.	H3 conditionally supported, not a causal effect
RQ2: Financial activity and subsequent non-oil growth	Seasonal credit interval crosses zero. Turnover levels are unstable, and the reformulated change association remains imprecise.	H1 and H2 not established
RQ3: ML improvement over transparent benchmarks	Small RMSE differences do not demonstrate a dependable gain. Financial features add no consistent improvement.	Predictive improvement not demonstrated

The limits of completion are clearly expressed in Table 8. All of the research questions are addressed, and when the answer is inconclusive and/or unfavorable to the proposed research hypothesis, they are addressed. This is the case for the payment adoption, which remains descriptive context as planned, with the two available annual observations not being converted to a quarterly treatment-intensity series. The remaining deferred extensions are sectoral decompositions and synthetic control. The other deferred extensions are sectoral decompositions and synthetic control; the data and the comparison conditions are not known. A national time indicator is also unable to isolate the effects of an FSDP from the other Vision 2030 changes, from other tax changes or from a

change in global conditions.

The defensible conclusion is that the conditional trajectory of Saudi non-oil growth has changed over the course of the program period in the selected baseline period, whereas the data does not provide evidence of the size of a GDP gain resulting from an FSDP, positive aggregate financial-channel effect or reliable machine-learning advantage. The use of consistent market histories before the reforms, of direct measures of financing provided and of credible differences in reform exposure would aid policy evaluation. The ability to clearly separate observed financial growth from conditional growth patterns and the greater evidence needed for causal attributions marks the current contribution.

7. CONCLUSION AND IMPLICATIONS

This study reviews the Saudi FSDP using three related questions but different approaches based on the observed quarterly data from original statistical producers. The selected interrupted time-series model suggests that a change in the conditional trend of non-oil growth occurred around 2018Q2, with a focus in the trend, not in the level. The evidence is sensitive to the lag order, estimation window, and the type of GDP (non-oil or aggregate). H3 is then supported in the chosen baseline in a qualified manner. The findings are not specific to a particular program and do not quantify a specific economic benefit.

The financial channel results have no bearing on H1 or H2. The point estimates of real claims growth are positive in several models, and while the signs of the point estimates are positive, the intervals grow large and the point estimates become weaker as the specification changes dynamically and seasonally. Market turnover is particularly challenging to analyze due to its younger time span, changing denominators and contradictory stability indications. The exploratory turnover-change equation helps to answer a modified question and assists some diagnostics. It is not able to solve the identification problems of the original market level hypothesis or provide a financing mechanism that the available aggregates do not directly observe.

The machine-learning extension does not assume that a complex model performs better and thus provides the answer to the third research question. The small core RMSE gain of macro ridge over AR(2) is questionable, the historical mean is competitive, and financial additions do not offer consistent improvements. The three forecast numbers reveal the basis for the forecasts, forecast errors and comparisons to features. The practical implication of this is to keep a few basic benchmarks and to sequentially test new predictors and not rely on more complicated forecasting systems. The results are for the tested one-quarter horizon, source vintage and evaluation windows.

Evidence helps to separate financial adoption, financing allocation and realized economic outcomes as a basis for program assessment. Adoption is captured in the documented increase in electronic payments, lending and turnover, which provide other measures of financial activity. None should be used singly as an indicator of other productive output. A standardized publication history, supplemented with direct measures of credit allocation, new equity and debt financing, and differential exposure to reforms by firm or sector, would be more informative. These types of measures would increase the visibility of the conceptual transmission channels and enable more

powerful comparison designs.

These constraints include just one national intervention, reforms and shocks occurring at the same time, financial histories that are short, data that are revised rather than real-time, and a relatively small test set for forecasts. These limits are visible when using robustness checks but cannot be removed from them. Before considering synthetic control or exposure-based designs, future work should take the time to build compatible pre-reform histories and establish credible comparison units. The present contribution is an auditable empirical report of findings to support a conditional program-period trend change, inconclusive aggregate financial associations and no demonstrated reliable machine-learning forecasting gain.

Citations:

[1] Saudi Central Bank, After being endorsed by CEDA: The Financial Sector Development Program holds a forum to introduce the program and its initiatives, Riyadh, Saudi Arabia, May 9, 2018. Original SAMA source. [Accessed: Sep. 15, 2026].

[2] International Monetary Fund, Saudi Arabia: Financial Sector Assessment Program—Financial System Stability Assessment, IMF Country Report No. 2024/281, Washington, DC, USA, International Monetary Fund, 2024. doi: 10.5089/9798400288074.002. [Online]. Available: https://doi.org/10.5089/9798400288074.002

[3] Saudi Central Bank, SAMA: E-Payments Account for 79% of Total Retail Payments in 2024, Apr. 15, 2025. Original SAMA source. [Accessed: Sep. 15, 2026].

[4] S. Alabdulwahab, "The linkage between oil and non-oil GDP in Saudi Arabia," Economies, vol. 9, no. 4, Art. no. 202, 2021. doi: 10.3390/economies9040202.

[5] M. I. Haque, "The growth of private sector and financial development in Saudi Arabia," Economies, vol. 8, no. 2, Art. no. 39, 2020. doi: 10.3390/economies8020039.

[6] J. A. Alzyadat, "Sectoral banking credit facilities and non-oil economic growth in Saudi Arabia: Application of the autoregressive distributed lag (ARDL)," The Journal of Asian Finance, Economics and Business, vol. 8, no. 2, pp. 809–820, 2021. doi: 10.13106/jafeb.2021.vol8.no2.0809.

[7] D. Sethi and D. Acharya, "Financial inclusion and economic growth linkage: Some cross country evidence," Journal of Financial Economic Policy, vol. 10, no. 3, pp. 369–385, 2018. doi: 10.1108/JFEP-11-2016-0073.

[8] D.-W. Kim, J.-S. Yu, and M. K. Hassan, "Financial inclusion and economic growth in OIC countries," Research in International Business and Finance, vol. 43, pp. 1–14, 2018. doi: 10.1016/j.ribaf.2017.07.178.

[9] Z. Abdul Karim, R. Nizam, S. H. Law, and M. K. Hassan, "Does financial inclusiveness affect economic growth? New evidence using a dynamic panel threshold regression," Finance Research Letters, vol. 46, Art. no. 102364, 2022. doi: 10.1016/j.frl.2021.102364.

[10] N. Emara and A. El Said, "Financial inclusion and economic growth: The role of governance in selected MENA countries," International Review of Economics & Finance, vol. 75, pp. 34–54, 2021. doi: 10.1016/j.iref.2021.03.014.

[11] P. K. Ozili, A. Ademiju, and S. Rachid, "Impact of financial inclusion on economic growth: Review of existing literature and directions for future research," International Journal of Social Economics, vol. 50, no. 8, pp. 1105–1122, 2023. doi: 10.1108/IJSE-05-2022-0339.

[12] P. K. Ozili, "Impact of digital finance on financial inclusion and stability," Borsa Istanbul Review, vol. 18, no. 4, pp. 329–340, 2018. doi: 10.1016/j.bir.2017.12.003.

[13] T.-H. Le, A. T. Chuc, and F. Taghizadeh-Hesary, "Financial inclusion and its impact on financial efficiency and sustainability: Empirical evidence from Asia," Borsa Istanbul Review, vol. 19, no. 4, pp. 310–322, 2019. doi: 10.1016/j.bir.2019.07.002.

[14] A. W. A. Boot, P. Hoffmann, L. Laeven, and L. Ratnovski, "Fintech: What's old, what's new?," Journal of Financial Stability, vol. 53, Art. no. 100836, 2021. doi: 10.1016/j.jfs.2020.100836.

[15] A. V. Thakor, "Fintech and banking: What do we know?," Journal of Financial Intermediation, vol. 41, Art. no. 100833, 2020. doi: 10.1016/j.jfi.2019.100833.

[16] S. Hathroubi, "Inclusive finance, growth and socio-economic development in Saudi Arabia: A threshold cointegration approach," Journal of Economic Development, vol. 44, no. 2, pp. 77–111, 2019. doi: 10.35866/caujed.2019.44.2.004.

[17] N. Samargandi, M. Al Mamun, K. Sohag, and M. Alandejani, "Women at work in Saudi Arabia: Impact of ICT diffusion and financial development," Technology in Society, vol. 59, Art. no. 101187, 2019. doi: 10.1016/j.techsoc.2019.101187.

[18] A. M. Alkhaldi, R. M. Elzaki, and I. F. Abdalla, "Determinants of financial development sustainability: Evidence from Saudi Arabia," SAGE Open, vol. 15, no. 3, 2025. doi: 10.1177/21582440251367851.

[19] M. Belloumi and A. Alshehry, "The impacts of domestic and foreign direct investments on economic growth in Saudi Arabia," Economies, vol. 6, no. 1, Art. no. 18, 2018. doi: 10.3390/economies6010018.

[20] S. Athey and G. W. Imbens, "Machine learning methods that economists should know about," Annual Review of Economics, vol. 11, pp. 685–725, 2019. doi: 10.1146/annurev-economics-080217-053433.

[21] P. Goulet Coulombe, M. Leroux, D. Stevanovic, and S. Surprenant, "How is machine learning useful for macroeconomic forecasting?," Journal of Applied Econometrics, vol. 37, no. 5, pp. 920–964, 2022. doi: 10.1002/jae.2910.

[22] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," PLOS ONE, vol. 13, no. 3, Art. no. e0194889, 2018. doi: 10.1371/journal.pone.0194889.

[23] M. C. Medeiros, G. F. R. Vasconcelos, Á. Veiga, and E. Zilberman, "Forecasting inflation in a data-rich environment: The benefits of machine learning methods," Journal of Business & Economic Statistics, vol. 39, no. 1, pp. 98–119, 2021. doi: 10.1080/07350015.2019.1637745.

[24] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," Nature Machine Intelligence, vol. 2, no. 1, pp. 56–67, 2020. doi: 10.1038/s42256-019-0138-9.

[25] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: An empirical study on performance estimation methods," Machine Learning, vol. 109, pp. 1997–2028, 2020. doi: 10.1007/s10994-020-05910-7.

[26] Capital Market Authority, Annual Report 2024, Riyadh, Saudi Arabia, Capital Market Authority, 2025. [Online]. Available: https://cma.gov.sa/en/Market/Reports/Documents/CMA_2024_REPORT_EN.pdf [Accessed: Sep. 15, 2026].

[27] Capital Market Authority, Statistical Bulletin Q1 2026: Equities, Table 11, Riyadh, Saudi Arabia, Capital Market Authority, 2026. Original CMA Equities XLSX. [Accessed: Sep. 15, 2026].

[28] A. Abadie, "Using synthetic controls: Feasibility, data requirements, and methodological aspects," Journal of Economic Literature, vol. 59, no. 2, pp. 391–425, 2021. doi: 10.1257/jel.20191450.

[29] General Authority for Statistics, Statistical Methodology: Comprehensive Revision of Gross Domestic Product (GDP), Riyadh, Saudi Arabia, General Authority for Statistics, 2025. Original GASTAT source. [Accessed: Sep. 15, 2026].

[30] General Authority for Statistics, Gross Domestic Product and National Accounts, Second Quarter 2026, Tables 3.2 and 3.3, Riyadh, Saudi Arabia, General Authority for Statistics, 2026. Original GASTAT Q2 2026 XLSX. [Accessed: Sep. 15, 2026].

[31] General Authority for Statistics, Consumer Price Index July 2026: Data Tables, Table 4.1, All Regions, Riyadh, Saudi Arabia, General Authority for Statistics, 2026. Original GASTAT CPI July 2026 XLSX. [Accessed: Sep. 15, 2026].

[32] KAPSARC Data Portal, Bank Claims on Private Sector, Saudi Central Bank (SAMA), 2026. [Online]. Available: https://datasource.kapsarc.org/explore/dataset/bank-claims-on-private-sector/information/ [Accessed: Sep. 15, 2026].

[33] KAPSARC Data Portal, Monetary Indicators (Interest rates), Saudi Central Bank (SAMA), 2026. [Online]. Available: https://datasource.kapsarc.org/explore/dataset/interest-rates-and-sama-average-bills/ [Accessed: Sep. 15, 2026].

[34] U.S. Energy Information Administration, Europe Brent Spot Price FOB (Dollars per Barrel), series RBRTE, 2026. Original EIA series page. [Accessed: Sep. 15, 2026].