Ahliya Journal of Business Technology and MEAN Economies Vol. 0_ Issue 0_ (2026)
[Published by Palestine Ahliya University] ISSN: 3007-9691

Oil-Price Changes and Economic Growth in MENA: Exposure Heterogeneity and Retrospective Forecast Evaluation

Faeyz Abuamria[1]

[1] Department of International Economics and Development, Graduate School, Palestine Ahliya University (Palestine)

✉ f.abuamria@paluniv.edu.ps

Received:17/09/2026 Accepted:17/10/2026 Published:

Abstract:

This study investigates whether oil-price changes are more strongly associated with real GDP growth in MENA economies with greater predetermined oil exposure, whether perceived corruption control moderates this relationship, and whether oil-related predictors improve growth forecasts. The analysis covers 21 economies, with 285 country-year observations spanning predictors from 2011–2024 and growth outcomes from 2012–2025. Country and year fixed-effects models are complemented by country-clustered standard errors, wild-cluster bootstrap procedures, and Driscoll–Kraay errors. Forecasts use a pseudo-out-of-sample design based on revised data. Results provide no robust evidence that the oil–growth association varies with exposure. The positive one-year interaction coefficient (0.092) is sensitive to alternative inference methods, multiple-testing adjustment, and excluding Libya's post-conflict rebound in 2012. Moderation by perceived corruption control is also not robust. For 2018–2024, ridge regression using lagged growth and country indicators achieves the lowest mean absolute error (4.16 percentage points). Adding oil exposure and governance predictors increases the error to 4.30 points, while no tested oil-related extension improves forecast accuracy. Overall, oil-related predictors provide no additional forecasting value in the models and period examined. The study highlights the sensitivity of inference to cross-country dependence and influential observations, while providing a reproducible forecasting benchmark.

Keywords: *Oil Prices; Economic Growth; MENA; Control of Corruption; Panel Data; Forecast Evaluation.*

Ahliya Journal of Business Technology and MEAN Economies Vol. 0_ Issue 0_ (2026)
[Published by Palestine Ahliya University] ISSN: 3007-9691

تغيرات أسعار النفط والنمو الاقتصادي في منطقة الشرق الأوسط وشمال أفريقيا: تباين درجات التعرض وتقييم التنبؤات بأثر رجعي

فايز أبو عمرية[1]

[1] قسم الاقتصاد الدولي والتنمية، كلية الدراسات العليا، جامعة فلسطين الأهلية (فلسطين)

f.abuamria@paluniv.edu.ps

تاريخ الاستلام:2026/09/17 تاريخ القبول:2026/10/17 تاريخ النشر:

ملخص:

تبحث هذه الدراسة فيما إذا كانت التغيرات في أسعار النفط ترتبط بصورة أقوى بالنمو الحقيقي للناتج المحلي الإجمالي في اقتصادات منطقة الشرق الأوسط وشمال أفريقيا (MENA) ذات الانكشاف النفطي المحدد مسبقًا، وما إذا كان التصور المتعلق بالسيطرة على الفساد يعدّل هذه العلاقة، وما إذا كانت المتغيرات المرتبطة بالنفط تحسّن التنبؤ بالنمو الاقتصادي. تشمل الدراسة 21 اقتصادًا، بواقع 285 مشاهدة من نوع الدولة–السنة، باستخدام متغيرات تنبؤية للفترة 2011–2024 ومعدلات نمو للفترة 2012–2025. وتستخدم نماذج التأثيرات الثابتة للدول والسنوات، إلى جانب الأخطاء المعيارية المجمّعة حسب الدولة، واختبار التمهيد العنقودي البري، والأخطاء المعيارية لدريسكول–كراي. وتعتمد عملية التنبؤ على تقييم شبه خارج العينة باستخدام بيانات منقّحة. لا تُظهر النتائج أدلة متينة على اختلاف العلاقة بين أسعار النفط والنمو الاقتصادي باختلاف مستوى الانكشاف النفطي. فرغم إيجابية معامل التفاعل للسنة الواحدة (0.092)، تتأثر دلالته الإحصائية بأساليب الاستدلال البديلة، وتصحيح تعدد الاختبارات، واستبعاد الانتعاش الاقتصادي في ليبيا بعد الصراع عام 2012. كما أن الدور المعدِّل للتصور المتعلق بالسيطرة على الفساد غير متين. وخلال الفترة 2018–2024، حقق نموذج انحدار ريدج، الذي يستخدم النمو المتأخر ومؤشرات الدول، أدنى متوسط للخطأ المطلق، بلغ 4.16 نقطة مئوية. وأدت إضافة مؤشرات الانكشاف النفطي والحوكمة إلى ارتفاع الخطأ إلى 4.30 نقطة مئوية، دون أن تحسّن أي من الامتدادات النفطية المختبرة دقة التنبؤ. وتشير النتائج إلى محدودية القيمة التنبؤية الإضافية للمتغيرات المرتبطة بالنفط، كما تبرز حساسية الاستدلال الإحصائي للاعتماد المتبادل بين الدول والمشاهدات المؤثرة، مع تقديم معيار تنبؤي قابل لإعادة التطبيق.

الكلمات المفتاحية: *أسعار النفط؛ النمو الاقتصادي؛ منطقة الشرق الأوسط وشمال أفريقيا؛ مكافحة الفساد؛ بيانات اللوحات؛ تقييم التوقعات.*

1. Introduction

Oil dominates macroeconomic discussion of the Middle East and North Africa. The region includes some of the world's largest exporters, several economies that import almost all of their energy, and states in which war and sanctions have repeatedly disrupted oil production. A common view is that an oil-price increase helps exporters and hurts importers, and that the size of the gain depends on how governments manage windfalls. The view has support in multi-country structural models (Cashin et al., 2014; Mohaddes & Raissi, 2019) and in MENA-specific studies (Berument et al., 2010; Abdelsalam, 2023). It also shapes how forecasters adjust growth projections for the region when oil prices move.

The closest regional evidence leaves two questions open. Country vector autoregressions find positive output responses to oil-price increases in most MENA exporters but not in Bahrain or in most importers (Berument et al., 2010), which points to heterogeneity that a simple exporter–importer split may not capture. Abdelsalam (2023) studies 17 MENA economies over 1970–2018 with the West Texas Intermediate price, splits the sample into exporters and importers, and adds a composite institutional-quality index from the PRS Group as a control; the oil-price coefficient falls when the index is included. Entering institutions as a control, however, is not the same as testing whether they change the slope of growth on oil prices across exposure levels. Apergis and Payne (2014) and Sadeghi (2017) link oil to growth through institutions and government size, but for oil abundance and for exporters only. None of these studies combines a continuous predetermined exposure measure with year effects, reports how its estimates change under inference that allows for correlation across countries, or evaluates whether the same variables improve forecasts.

This study evaluates whether a predetermined oil-rent exposure measure is associated with differential growth responses to annual oil-price changes in MENA, and whether oil-related predictors improve retrospective forecast accuracy. We use 21 economies with predictors for 2011–2024 and growth outcomes for 2012–2025, interact the annual Brent change with mean oil rents in 2000–2009 under country and year effects, and test whether perceived control of corruption measured in 2002–2009 modifies that interaction. The forecast comparison is kept separate from the explanatory analysis: it scores simple benchmarks and extended models on identical cases, using only outcomes that would have been published at each forecast origin. The analysis follows a dated plan written after a pilot on the same data (Section 3.4).

The evidence is inconclusive for the explanatory hypotheses. The one-year-ahead interaction is positive under country clustering, but its interval widens to include large negative and positive values once dependence across countries is allowed for, and re-estimating without one observation, Libya's 86.8 percent growth in 2012 after the 2011 conflict, reduces it to about zero. We find no robust evidence of moderation by perceived control of corruption, while the intervals remain too wide to exclude economically relevant moderation. In the primary 2018–2024 forecast window, none of the oil-related extensions lowers mean absolute error relative to the lagged-growth benchmark; one added bundle, the corruption score with a tuned penalty, is slightly more accurate, by less than its sampling uncertainty. Our contribution is therefore a transparent assessment of inference, observation influence and benchmark performance within this design, not a new estimate of the effect of oil prices. Section 2 reviews the literature and states the hypotheses, Section 3 describes the data and design, Section 4 the methods, Section 5 the results, Section 6 the discussion and limitations, and Section 7 concludes.

2. Literature, Mechanisms and Hypotheses

2.1 Oil-price changes and their sources

The modern literature begins with the observation that oil-price increases preceded most post-war US recessions (Hamilton, 1983). Later work stresses that oil prices are endogenous to the world economy and that price changes combine shocks of different origin. Kilian (2009) separates supply shocks, shocks to global demand for industrial commodities and oil-specific demand shocks, and finds that they have different macroeconomic effects. Baumeister and Hamilton (2019) relax the identifying assumptions and give supply disruptions a larger role in historical price movements. Känzig (2021) identifies oil-supply news from high-frequency futures-price reactions to OPEC announcements and shows that news of lower future supply raises oil prices and lowers US economic activity. These are structural results for the United States or the world oil market; they do not imply that an annual Brent change is a structural shock, and we do not treat it as one. For MENA producers the distinction matters in a specific way, because many of the OPEC decisions behind these supply-news shocks change their own production volumes.

2.2 Regional heterogeneity by exposure and production

Global vector autoregressions estimated for a large set of economies find that a supply-driven price increase raises output in major exporters and lowers it in importers, whereas demand-driven increases raise activity more broadly (Cashin et al., 2014). The effects of a supply disruption also depend on whether other producers offset it: disruptions in Iran were largely offset by Saudi Arabia, but disruptions in Saudi Arabia were not (Mohaddes & Pesaran, 2016). The US shale expansion lowered oil prices and reduced output in commodity exporters (Mohaddes & Raissi, 2019). These quarterly structural models estimate country-specific impulse responses; our annual panel estimates a single slope across a continuous exposure measure, so the two kinds of evidence answer related but different questions.

Among MENA economies, Berument et al. (2010) estimate country vector autoregressions in which each economy is too small to move the world price. Oil-price increases have significant positive effects on output in Algeria, Iran, Iraq, Kuwait, Libya, Oman, Qatar, Syria and the United Arab Emirates, but not in Bahrain, Djibouti, Egypt, Israel, Jordan, Morocco or Tunisia. The result differs from ours in three ways that limit direct comparison: shocks are identified within each country's VAR rather than by a common year effect, significance is assessed country by country rather than for a pooled exposure slope, and the response is traced over several periods rather than at a fixed annual horizon. Abdelsalam (2023) is the closest panel study. It covers 17 MENA economies (nine exporters and eight importers, including Turkey) for 1970–2018, uses the annual change in the WTI price and a GARCH measure of its volatility, and estimates pooled, fixed-effects, random-effects, system GMM and panel quantile regressions of real GDP growth. Price changes have positive coefficients for exporters and negative ones for importers. The institutional-quality index enters additively with the other controls (Abdelsalam, 2023, equation 3); its inclusion lowers the oil-price coefficient. That design tests whether institutions are associated with growth and whether controlling for them changes the oil coefficient, not whether institutions change the oil-price response across countries, which is what a triple interaction tests. Apergis and Payne (2014) study ten MENA oil economies over 1990–2013 with time-varying cointegration and find that the negative association between oil reserves and growth turned positive after 2003 and is offset by better institutional quality; their measure is oil abundance and their horizon the long run, not annual price changes.

2.3 Institutions, fiscal channels and volatility

A long literature on the resource curse links resource dependence to slower growth (Sachs & Warner, 2001). It then asks whether institutions condition the effect: the curse appears where institutions favour rent-seeking (Mehlum et al., 2006), and oil dependence harms growth mainly where political institutions are weak (Antonakakis et al., 2017). Surveys emphasise that the evidence is sensitive to measurement and identification (van der Ploeg, 2011; Ross, 2015). Heterogeneous panel estimates that allow for cross-country dependence associate oil abundance with higher, not lower, income (Cavalcanti et al., 2011). These studies concern long-run resource dependence or abundance; they do not imply a particular short-run growth response to an annual price change. Higher oil rents also raise perceived corruption (Arezki & Brückner, 2011), which is one reason not to treat governance scores as exogenous to oil income.

Closer to short-run transmission, oil-revenue volatility lowers growth among major producers, and better fiscal institutions offset part of that effect (El-Anshasy et al., 2017). Commodity booms raise income in the short run but lower it in the long run where governance is poor (Collier & Goderis, 2012). Jarrett et al. (2019) find that better financial institutions mitigate the growth effect of oil-price volatility, using a synthetic-control design around the 2014 price collapse and a panel of 30 oil producers. Across 28 oil exporters, a larger government is associated with a stronger non-oil output response to oil-price shocks and with higher output volatility (Sadeghi, 2017). These results concern financial or fiscal institutions, revenue volatility and non-oil output. Our moderator is a perceptions-based corruption score and our outcome total real GDP growth, so a null result here would not contradict them. Two further studies bear on timing: external shocks, including commodity prices, explain only a modest share of output fluctuations in low-income countries (Raddatz, 2007), and news of giant oil discoveries moves investment and the current account years before GDP (Arezki et al., 2017), so the horizon at which any exposure-dependent response appears is itself uncertain.

2.4 Forecasting and benchmarks

Evaluations of IMF World Economic Outlook forecasts set the benchmark problem (Timmermann, 2007). Short-term growth forecasts for 2004–2017 show little bias, but two- to five-year-ahead forecasts were biased upward and, in up to half of countries, less accurate than a naïve forecast based on recent average growth (Celasun et al., 2021). Simple benchmarks are therefore demanding comparators. Comparisons of forecast accuracy need to account for the dependence of forecast errors (Diebold & Mariano, 1995; Harvey et al., 1997), and evaluation schemes must not use information that was unavailable at the forecast origin (Hyndman & Athanasopoulos, 2021). Both requirements are hard to meet with seven annual test years, and we describe below what our tests can and cannot show.

2.5 Gap and contribution

Table 1 sets the closest studies against this one. The exporter–importer contrast and the role of institutions are not new questions for MENA. What this study adds is a specific design and a standard of reporting: a single continuous exposure measure fixed before the sample, year effects that absorb the common price change, an explicit triple interaction for institutional moderation, inference reported under several dependence assumptions with an influence analysis, and a forecast comparison that respects publication lags. We searched Crossref, OpenAlex and Semantic Scholar on 25 September 2026 and re-checked the closest studies on 4 October 2026 (search log in the Supplementary Material). The search was targeted rather than systematic, so we do not claim that no similar study exists. The evidence matrix in the Supplementary Material records, for each study, its

sample, shock measure, institutional measure and role, outcome, method, finding, the limits of comparison with this study, and which source was read to support the summary.

Table 1: Closest studies compared with the design of this study

Study	Sample and period	Oil measure	Institutions: measure and role	Outcome; horizon	Method; inference	Limit for comparison with this study
Berument et al. (2010)	16 MENA economies; country-specific samples	Oil-price shock; each economy too small to affect the price	None	Output growth; impulse responses	Country VARs; significance by country	No pooled exposure slope or common year effects
Abdelsalam (2023)	17 MENA economies (9 exporters, 8 importers); 1970–2018, annual	WTI change; GARCH volatility	PRS composite index; additive control	Real GDP growth; same year	Pooled, FE, RE, system GMM, panel quantile	Exporter/importer split; institutions not interacted with the oil-price change
Apergis & Payne (2014)	10 MENA oil economies; 1990–2013	Oil reserves (abundance)	Institutional quality; offsets the reserves effect	Growth; long run	Time-varying cointegration	Abundance, not annual price changes; exporters only
Sadeghi (2017)	28 oil exporters; 1990–2016	Oil-price shocks	Government size; conditioning variable	Non-oil output, volatility, spending	Panel estimation	Exporters only; fiscal size, not corruption
Jarrett et al. (2019)	30 oil producers, 1980–2016; 20 oil-rich economies around 2014	Oil-price volatility; 2014 price collapse	Financial institutions; interaction	Growth and output volatility	CS-ARDL; synthetic control	Volatility and financial depth, not price change × exposure
This study	21 MENA economies; predictors 2011–2024, outcomes 2012–2025	Brent annual log change × oil rents 2000–2009; Känzig supply news	WGI control of corruption 2002–2009 (and $t-1$); triple interaction	Real GDP growth; h = 0, 1 (primary), 2	Two-way FE; CR1, wild bootstrap, Driscoll–Kraay; influence analysis; separate forecast evaluation	Association, not a total or causal effect; 14 annual price changes

Note. Entries summarise the design as described in each source; the source read for each study (full text, publisher page or abstract) is recorded in the evidence matrix in the Supplementary Material. FE = fixed effects; RE = random effects; GMM = generalised method of moments; CS-ARDL = cross-sectionally augmented autoregressive distributed lag.

2.6 Mechanisms

Exporters and importers. An oil-price increase transfers income from importers to exporters. For exporters, however, real GDP measures volumes, not prices. A price increase raises real GDP only if producers expand output or if the windfall raises non-oil activity through public spending, credit and confidence. The fiscal channel operates with a lag, so a response may appear in the following year rather than the same year. For importers, the price increase acts as a terms-of-trade loss, partly offset in MENA by remittances, aid and investment from Gulf exporters. Historical oil rents measure exposure among producers, but they do not measure how dependent an importer is on imported energy: a rent of zero covers Jordan, Lebanon and Morocco, which import most of their energy, as

well as economies with other energy sources. We therefore also use a measure based on net energy imports as an alternative.

Supply and demand. Annual Brent price changes combine supply shocks, global demand shocks and oil-specific demand shocks. A price increase driven by global demand arrives with stronger external demand for everyone, and one driven by lower supply does not. For the large MENA producers there is a further complication. The OPEC announcements that identify supply-news shocks often cut their own production, so a supply-news shock that raises prices may lower their real output in the same year even as it raises their income. The sign of the exposure-dependent response to supply news is therefore ambiguous a priori.

Institutions. Stronger governance may smooth windfalls through fiscal rules and savings funds, or it may allow faster use of windfalls in investment. The direction of moderation is therefore not signed in advance. The WGI control-of-corruption score measures *perceived* control of corruption, aggregated from surveys and expert assessments with sampling error; it is not a measure of observed corrupt transactions (Kaufmann et al., 2011; Jiang et al., 2025). Because oil rents can themselves raise perceived corruption (Arezki & Brückner, 2011), our main moderator is the score averaged over 2002–2009, before the estimation period.

Conflict and production. Conflict can change transmission by shutting down production, as in Libya, Syria and Yemen, and production recoveries after conflict can coincide with price movements. We treat this as a possible explanation for influential observations, not as a hypothesis we test. Country-year conflict data from the Uppsala Conflict Data Program (Sundberg & Melander, 2013; Davies et al., 2026) and comparable country oil-production series could not be obtained in the environment used for the analysis, so a conflict or production channel is left for further work.

2.7 Hypotheses

The explanatory hypotheses, fixed in the dated plan before the new specifications were estimated, are:

- **H1 (exposure).** After common year effects are removed, growth is more strongly associated with oil-price changes where predetermined oil exposure is higher. The primary horizon is growth in $t + 1$; growth in t is the second primary test.
- **H2 (supply news and the remaining price change).** We ask two separate questions: whether the exposure interaction with annual supply news differs from zero, and whether the exposure interaction with the part of the Brent change not explained by supply news differs from zero. The two coefficients are in different units, so we do not test their equality. The sign is not predicted.
- **H3 (institutions).** Perceived control of corruption measured in 2002–2009 changes the exposure-dependent association. The direction is not predicted.

The forecasting hypothesis, **FH**, is that adding oil exposure interacted with the Brent change, and the corruption score, to a model with lagged growth and country indicators lowers the mean absolute error of next-year growth forecasts relative to that model and to the country historical mean.

3. Data and Design

3.1 Boundary, candidates and selection

The candidate population is the 22 economies in the IMF's Middle East and North Africa grouping plus Israel, which the World Bank includes in its MENA region, giving 23 candidates. Israel is added because it is geographically in the region and has a complete data record. Its oil exposure is close to zero, so it contributes mainly to the comparison group. Because the IMF grouping excludes Israel, every result is re-estimated on the 22-economy IMF definition.

The eligibility rule uses predictor availability only, not outcomes. A country is eligible if it has at least 10 of 14 predictor-complete years in 2011–2024, at least 3 of 7 in 2018–2024, and at least 5 observed oil-rent years in 2000–2009. A year is predictor-complete when lagged growth, the lagged corruption score, the Brent change and the exposure measure are all observed. The rule admits 21 economies (Table 2). Somalia fails because it has no predictor-complete year; Yemen fails with 9 of 14 and 2 of 7. The West Bank and Gaza and Syria are eligible under this rule, although an earlier seven-predictor design excluded them because it required official exchange-rate and inflation data. The eligible countries provide 285 labelled pairs of predictors in year t and growth in year $t + 1$ for t = 2011–2024, so the outcomes of the primary sample run from 2012 to 2025; 18 pairs have a target year of 2025. Missing values are never set to zero or imputed, and every estimate is complete-case with its sample size reported. All 23 candidates remain in the raw data.

Because the rule uses predictor completeness up to 2024, the forecast comparison is conditional on a country set defined retrospectively: at the first forecast origin it was not known which economies would keep reporting. As a sensitivity check we also apply a rule that uses only predictors observed by that origin (at least 4 of 6 predictor-complete years in 2011–2016). It admits Yemen and drops Djibouti, again 21 economies.

Table 2: Candidate economies, eligibility and estimation samples

Step	Economies	Count	Pairs
Candidates: IMF MENA (22) plus Israel	DZA, BHR, DJI, EGY, IRN, IRQ, ISR, JOR, KWT, LBN, LBY, MRT, MAR, OMN, QAT, SAU, SOM, SDN, SYR, TUN, ARE, PSE, YEM	23	—
Excluded: < 10 of 14 predictor-complete years (2011–2024) or < 3 of 7 (2018–2024)	SOM (0/14, 0/7); YEM (9/14, 2/7)	2	—
Eligible (revised rule; exposure baseline ≥ 5 years available for all)	All others	21	285
Explanatory sample, h = 1	Predictor years 2011–2024; outcome years 2012–2025	21	285 (18 with outcome 2025)
Explanatory sample, h = 0	Predictor and outcome years 2011–2025	21	306 (18 with outcome 2025)
Explanatory sample, h = 2 (secondary)	Predictor years 2011–2023; outcome years 2013–2025	21	264 (18 with outcome 2025)
Common sample across horizons (revision)	Predictor years 2011–2023 with all three outcomes	21	264
IMF-only sensitivity (drop ISR)	20 eligible	20	271
Primary forecast test cases (targets 2018–2024)	Eligible with complete predictors and outcome	21	145
Origin-dated eligibility (forecast sensitivity)	≥ 4 of 6 predictor-complete years 2011–2016: adds YEM, drops DJI	21	139 test cases
Earlier seven-predictor design (not used)	Eligible minus PSE and SYR	19	275

Note. A predictor-complete year has growth in $t - 1$, the WGI corruption score in $t - 1$, the Brent change in t and the exposure measure. Outcomes are not used in selection, but completeness up to 2024 is known only in retrospect. Source: supplied workbook; recomputed by s01_audit.py, s02_explanatory.py and s03_forecast.py.

3.2 Variables, sources and vintages

Table 3 lists the variables. All are rebuilt from the raw sheets of the workbook rather than taken from its prepared panels. Real GDP growth is the World Development Indicators series NY.GDP.MKTP.KD.ZG from a snapshot obtained in July 2026 (World Bank, 2026a). The snapshot

does not flag whether the 2025 values are estimates, and for several economies they are likely to be revised; results without 2025 outcomes are reported in Section 5.2. Oil exposure is the mean of WDI oil rents (NY.GDP.PETR.RT.ZS) over 2000–2009, a period that ends before the estimation sample begins. Exposure ranges from zero in Djibouti, Lebanon and the West Bank and Gaza to 53.8 percent of GDP in Iraq; 11 of the 21 eligible economies exceed 10 percent (Figure 2). The 10 percent threshold is used only for descriptive grouping and for one robustness check; it is not an established economic boundary. The oil price is the World Bank Pink Sheet monthly Brent price in nominal US dollars, averaged over the twelve months of each year (World Bank, 2026b), and enters as 100 times the annual log change, in log points. Brent fell by 63.6 log points in 2015 and 41.5 in 2020, and rose by 51.0 in 2021 and 34.8 in 2022 (Figure 1). Because year effects absorb any common deflator only up to its interaction with exposure, we also report a version deflated by the US consumer price index contained in the Känzig (2021) replication data. Perceived control of corruption is the WGI score on the 0–100 scale introduced in the 2025 methodology revision (Jiang et al., 2025; World Bank, 2025). The WGI provides no observation for 2001 and none for 2025 in this release.

The identified supply measure is the oil supply news shock of Känzig (2021), vintage December 2025, obtained from the author's public repository. Monthly values are summed within each calendar year. A positive value is news of lower future supply that raises oil prices. At annual frequency the supply-news series and the Brent change are almost uncorrelated ($r = 0.06$ over 2001–2025). In a regression of the Brent change on supply news the coefficient is 0.90 (standard error 3.09, $R^2 = 0.004$): the annual supply-news measure explains little contemporaneous annual Brent-price variation in this auxiliary regression. This does not mean that supply news is unimportant for prices, because its effects can be dynamic and annual sums mix announcements of opposite sign. We call the residual of that regression the *unexplained price component*. It is not an identified demand shock. The supply-news series is estimated from a vector autoregression over the full sample, so it uses information after each year; the correlation between the December 2024 and December 2025 vintages for 2000–2024 is 0.98. For forecasting, we use instead the underlying high-frequency oil-supply surprises, which are measured on the announcement day and do not depend on later data.

Table 3: Variables, timing and availability at the forecast origin

Variable	Definition and source	Timing in model	Available at origin (start of target year T)?
Growth g	Real GDP growth, % (WDI NY.GDP.MKTP.KD.ZG; snapshot July 2026)	Outcome $t + 1$ (t, $t + 2$ secondary); predictor $t - 1$	Growth to $T - 2$ published; $T - 1$ usually not. Training uses targets $\leq T - 2$. Revised values used: not real-time
Oil exposure x	Mean oil rents, % of GDP, 2000–2009, ≥ 5 obs. (WDI NY.GDP.PETR.RT.ZS)	Fixed	Yes (subject to later revisions of rents)
Brent change Δp	$100 \times \Delta$ log of the annual mean of 12 monthly Brent prices, nominal USD (World Bank Pink Sheet); real version deflated by US CPI	t	Yes (full-year average of $T - 1$ known)
Control of corruption	WGI score, 0–100 scale, 2025 revision; mean 2002–2009 (primary moderator) or $t - 1$	Fixed or $t - 1$	Score for $T - 2$ normally released during $T - 1$, but the 2025 revision recalculated history: not real-time
Government effectiveness	WGI score, 0–100 scale, mean 2002–2009	Fixed	As above; robustness only

Variable	Definition and source	Timing in model	Available at origin (start of target year *T*)?
Supply news *s*	Känzig (2021) oil supply news shock, monthly, summed by year; vintage 2025M12	*t*	No: estimated with full-sample VAR (explanatory use only)
Supply surprise	Känzig (2021) high-frequency OPEC-announcement surprises, summed by year	*t*	Yes (measured on announcement days); used in M8
Net energy exports	−(net energy imports, % of energy use)/100, mean 2000–2009 (WDI EG.IMP.CONS.ZS)	Fixed	Robustness only; 17 economies

Note. WDI = World Development Indicators; WGI = Worldwide Governance Indicators. Current account, fixed investment, reserves and trade (WDI) are in the workbook but not used, because their coverage differs by country (Supplementary Table S5). Release-date assumptions and file details are in Supplementary Table S2. Source: authors.

The workbook also holds WDI series for the current account, fixed investment, reserves and trade. Their coverage is uneven: the United Arab Emirates, Iran and Syria have almost no current-account or reserves data for 2010–2024, and Jordan, Qatar and the United Arab Emirates lack fixed investment. Using any of them in the main model would change the sample in ways related to country characteristics, so we keep them out of the main analysis; coverage grids are in Supplementary Table S5. Gold prices, the official exchange rate, inflation and unemployment are not in any primary model. Gold has no clear transmission mechanism to MENA growth that is distinct from global conditions absorbed by year effects. The official exchange rate is a poor measure for economies with parallel markets and is missing for the West Bank and Gaza. Inflation and unemployment are plausibly outcomes of the same shocks and would introduce post-treatment bias.

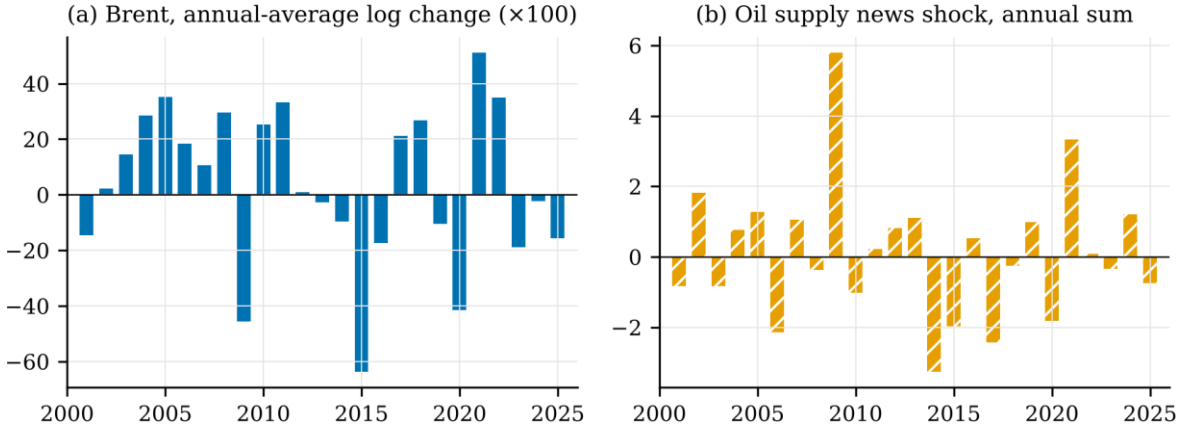

Figure 1: Annual Brent price change and annual oil supply news, 2001–2025

Note. Panel (a): 100 × log change of the annual average of monthly Brent prices, nominal US dollars (World Bank, 2026b, via the supplied workbook). Panel (b): sum of monthly oil supply news shocks, vintage 2025M12 (Känzig, 2021); positive values are news of lower future supply. Source: authors' calculations; code s04_figures.py.

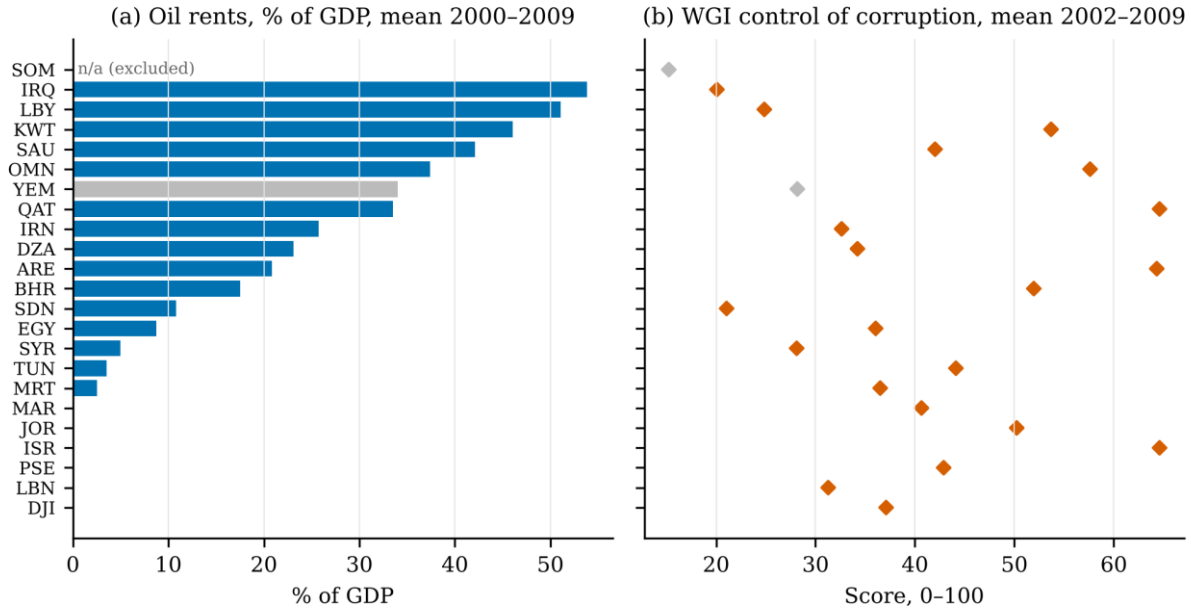

Figure 2: Predetermined oil exposure and perceived control of corruption by country

Note. Panel (a): mean WDI oil rents, 2000–2009 (World Bank, 2026a). Panel (b): mean WGI control-of-corruption score, 2002–2009 (World Bank, 2025). Coloured markers: the 21 eligible economies; grey: excluded (Yemen, Somalia; Somalia has no oil-rent data). Source: authors' calculations.

Table 4: Descriptive statistics, primary estimation sample (h = 1)

Variable	N	Mean	SD	Min	Median	Max
Real GDP growth, t+1 (%)	285	1.97	8.25	−29.46	2.72	86.83
Real GDP growth, t−1 (%)	285	2.24	8.74	−50.34	2.91	86.83
Brent log change, t (log points)	285	−0.01	30.50	−63.62	−2.82	51.02
Oil rents 2000–2009 (% of GDP)	21	18.18	18.91	0.00	10.78	53.82
Brent change × exposure/100	285	0.02	7.94	−34.24	0.00	27.46
Control of corruption, t−1 (0–100)	285	39.90	15.42	13.70	39.24	75.07
Control of corruption 2002–2009 (0–100)	21	41.84	13.84	20.08	40.66	64.60
Supply news shock, annual sum	285	−0.14	1.69	−3.26	0.08	3.33

Note. 21 economies, predictor years 2011–2024, outcome years 2012–2025. Exposure and the 2002–2009 corruption score are country-level (N = 21). Source: authors' calculations.

3.3 Information available at the forecast origin

The forecast origin for target year T is the start of year T. At that date the average Brent price for $T-1$ and the high-frequency supply surprises for $T-1$ are known, and the WGI score for $T-2$ would normally have been released. GDP growth for $T-2$ has normally been published, but growth for $T-1$ usually has not, which is why the predictor is growth two years before the target. The same rule must apply to every use of outcomes, not only to the lag: at origin T, a training pair can be used only if its target year is $T-2$ or earlier. We apply this rule to the outer training sample, the country-mean benchmark, the inner validation that selects the ridge penalty, the volatility-based country exclusion and the calibration of forecast intervals (Section 4.4). Supplementary Table S2 lists the availability of each source at the origin.

These timing assumptions hold for first releases, but the values we use are not first releases. The WDI snapshot contains revised national accounts, and the WGI 2025 revision recalculated the entire history of the indicators. We could not obtain archived WDI or WGI vintages, or historical WEO forecasts: the World Bank and IMF hosts were blocked by the egress policy of the environment

in which the analysis was run. Every forecast comparison is therefore a retrospective pseudo-out-of-sample exercise on revised data, and none of it is a real-time test.

3.4 Analysis plan, pilot history and deviations

The study started from a research workbook that had already been used in a pilot. Its frozen results showed that adding an oil-exposure interaction and a corruption score to a lagged-growth model did not reduce forecast errors over 2018–2024, and a 2025 holdout from an earlier seven-predictor design had also been inspected. We rebuilt every variable from the raw sheets and reproduced the frozen results (Supplementary Table S4). We then wrote a dated analysis plan (25 September 2026, in the replication package) before estimating any new specification. The plan was written after the pilot and after some results had been seen, so it is not a preregistration: it fixes the additional analyses, not the original choice of design. All forecast comparisons are exploratory for the same reason. Nine deviations are recorded (Supplementary Table S1). D1 to D7 were logged during the original analysis; three of them (D2, D5 and D7) are diagnostics added after seeing results. D8 (stepwise forecast comparisons) and D9 (further robustness and inference checks) were added in response to review, as was the correction of the forecast timing rule. Every specification estimated is listed in the specification log.

4. Methods

4.1 Explanatory specification

The primary equation is

$$g_{i,t+h} = \alpha_i + \delta_t + \beta_h \, (\Delta p_t \times x_i/100) + \varepsilon_{i,t+h} \qquad (1)$$

where g is real GDP growth, α_i and δ_t are country and year effects, Δp_t is the Brent log change (log points) and x_i is oil exposure in percent of GDP. The country effects absorb exposure itself and the year effects absorb the common oil-price change and every other common global shock, so omitting these main effects is not a specification error: only their interaction is identified. β_h is therefore a difference in association: the additional growth, in percentage points, per unit of the interaction, comparing economies with different exposure in the same year. It is not the total effect of oil prices on any country's growth, and it is not the response to an identified structural shock. For example, a Brent increase of 10 log points (about 10.5 percent) and exposure of 40 percent of GDP give an interaction value of 4, and the implied growth difference relative to a zero-exposure economy is $4\beta_h$. The primary horizons are $h = 1$ and $h = 0$; $h = 2$ is secondary. Table 2 shows the predictor and outcome years for each horizon; results on one common sample across horizons are reported in Section 5.2. We estimate a version without year effects that includes the Brent main effect, only to show that the interaction does not depend on how common shocks are modelled.

Exposure is predetermined, because it is measured before 2010, but that does not make it exogenous. The interpretation as a conditional association requires that, given country and year effects, the interaction of a world price change with pre-2010 exposure is uncorrelated with other country-specific shocks to growth. The condition fails if exposure is correlated with other channels through which global shocks reach growth, such as conflict, sanctions, production capacity or the size of the state, or if the shocks that move country growth also move the world price; Libya's 2011 conflict, for example, removed supply from the world market in the same year that Brent rose. A causal interpretation would need an external source of variation and exclusion restrictions that this design does not provide, and adding control variables would not supply them.

Dynamics. A version (E1b) adds growth in $t - 1$, the same lag used in forecasting. Relative to the outcome it is a lag of two years at $h = 1$, one year at $h = 0$ and three years at $h = 2$. With $T =$

14, fixed effects with a lagged dependent variable suffer the Nickell (1981) bias of order $1/T$. The bias is largest for the autoregressive coefficient but can also affect other coefficients when they are correlated with the lagged outcome, so E1b is a robustness check, not the primary model. We do not use system GMM: with 21 countries, the instrument count would quickly approach the number of cross-sectional units, weakening specification tests and biasing estimates toward fixed effects (Roodman, 2009).

4.2 Supply news and institutions

For H2 the interaction is split into supply news × exposure and unexplained price component × exposure. For H3 with the predetermined score, we add $\Delta p_t \times c_i$ and $\Delta p_t \times x_i \times c_i/100$, where c is the 2002–2009 corruption score, centred at the cross-country mean and scaled per 10 points. The lower-order terms x_i, c_i and $x_i \times c_i$ are fixed within countries and absorbed by the country effects, and Δp_t is absorbed by the year effects. The alternative uses the time-varying score in $t - 1$, $c_{i,t-1}$. Because that score varies over time, none of its lower-order terms is absorbed, and the full specification is

$$g_{i,t+h} = \alpha_i + \delta_t + \beta_1 \Delta p_t x_i + \beta_2 c_{i,t-1} + \beta_3 x_i c_{i,t-1} + \beta_4 \Delta p_t c_{i,t-1} + \beta_5 \Delta p_t x_i c_{i,t-1} + \varepsilon_{i,t+h} \quad (2)$$

with the same scaling. The original version of this alternative omitted $x_i \times c_{i,t-1}$; it is corrected here, and both versions are in Supplementary Table S10. The coefficient on a triple interaction is not by itself the effect of institutions. We therefore also report the implied exposure slope, $\beta_1 + \beta_5 c$, at percentiles of the corruption score, with standard errors from the full coefficient covariance. Government effectiveness replaces control of corruption in a robustness check. We do not include both, because the two indices are strongly correlated perceptions measures.

4.3 Inference with few countries and common shocks

Each method below rests on different assumptions, and none is decisive with 21 countries and 14 or 15 years. The country-clustered (CR1) covariance allows arbitrary correlation within a country over time and assumes independence across countries; it uses the small-sample factor $G/(G - 1) \times (N - 1)/(N - K)$, and we report confidence intervals and p-values from a t distribution with $G - 1 = 20$ degrees of freedom (normal-approximation values are in the Supplementary Material). Cluster-robust inference is unreliable with this number of clusters (Cameron et al., 2008), so we also report restricted wild-cluster bootstrap p-values with Webb six-point weights and 9,999 replications (Webb, 2023). The bootstrap improves the small-cluster approximation, but it keeps the assumption of independence across countries. That assumption is doubtful: residual cross-sectional dependence remains after year effects are removed, with a Pesaran (2015) CD statistic of 5.4 for the $h = 1$ residuals with year effects and 17.9 without. Driscoll and Kraay (1998) standard errors allow correlation across countries and over two lags of time (Bartlett weights, normal approximation), but their justification is asymptotic in the number of time periods and we have only 14; we report how they change with zero to four lags. A difference between methods therefore signals that conclusion depend on the dependence assumption, and non-significance under one method is not evidence of no association.

We also report an exploratory placebo exercise that permutes exposure values across economies 9,999 times and compares the absolute coefficient with its permutation distribution (Young, 2019). Exposure was not randomly assigned, and its permutation breaks its link with conflict, institutions and economic structure, so the resulting p-value is valid only if exposure values are exchangeable across economies, which the observational design does not guarantee. We use it as a sensitivity check, not as design-based inference. The three primary tests (H1 at $h = 1$ and $h = 0$, H3 at $h = 1$) are adjusted with the Holm procedure using wild-bootstrap p-values; all other results are secondary. For H3 we examine sensitivity to the published uncertainty of the WGI scores by drawing each annual score 1,000 times from a normal distribution with its published standard error and

recomputing the 2002–2009 mean, with errors independent across years and, as a second scenario, perfectly correlated within each country. This is a perturbation exercise. It does not recover true institutional quality or correct measurement-error bias, and it assumes that the 2002–2009 score remains a relevant measure for 2011–2025.

4.4 Forecast evaluation

The target is growth in year T and the origin is the start of T. The window expands. Following the rule in Section 3.3, models for target year T are trained only on pairs whose target year is $T - 2$ or earlier and whose predictor year is 2011 or later (Hyndman & Athanasopoulos, 2021); the first primary origin, 2018, therefore uses 101 training pairs with targets 2012–2016. The test set for T is all eligible country-years with complete predictors and an observed outcome in T. The primary test years are 2018–2024, with 145 common country-year cases scored for every model; 2015–2024 is secondary. The pre-specified models are:

- M0: country historical mean of published growth;
- M1: no change, using growth in $T - 2$;
- M2: pooled autoregression;
- M3: ridge regression on lagged growth with country indicators (Hoerl & Kennard, 1970), penalty 10, as in the pilot; the country indicators are penalised with the other coefficients, so they are shrunk intercepts rather than unpenalised fixed effects;
- M4: M3 plus corruption and the Brent–exposure interaction;
- M5 and M6: the same two specifications estimated by unpenalised OLS with country effects;
- M7: M4 with the ridge penalty chosen at each origin by inner expanding-window validation over {0.1, 1, 10, 100, 1,000};
- M8: M4 plus the oil-supply surprise × exposure interaction.

The inner validation uses only information published at the origin: its test years are the three latest published target years ($T - 4$ to $T - 2$), and each inner training set again keeps targets at least two years earlier than its own test year. An inner fold is used only if its training set has at least 30 pairs; at the 2018 origin two folds qualify, and at the 2015 and 2016 origins none does, so the plan's default penalty of 10 is used there (Supplementary Table S12c). Because M4 adds two predictor groups at once, it cannot show which one changes accuracy. We therefore add M3t, which applies the same penalty selection to M3, and a stepwise comparison of predictor bundles added to M3: the Brent change alone, the Brent–exposure interaction, both, the corruption score, and their combinations, each with the fixed penalty of 10 and with the tuned penalty (deviation D8). The Brent change alone is included because year effects cannot be used in a forecast: the effect for the target year is unknown at the origin.

We report mean absolute error (MAE), root mean squared error (RMSE), median absolute error, bias, errors in 2020–2021 and in other years, and errors by country. MAE is the primary metric because it was fixed in the plan; RMSE and the median are secondary and are reported for every model. For each model we report the difference in MAE from M3 with two uncertainty measures. The first is an exploratory t-test on the seven annual mean differences in absolute error, which allows for correlation across countries within a year but assumes independence across years; it is not a Diebold–Mariano test with a long-run variance estimate, and we do not apply the Harvey et al. (1997) correction. The second is a 95 percent interval from a bootstrap that resamples countries (9,999 draws), which keeps each country's errors together over time but ignores common shocks across countries. With seven test years neither can establish that two models are equally accurate. Forecast

intervals at the 80 percent level are formed in two ways: from quantiles of training residuals, as in the plan, and from the model's own earlier out-of-sample errors whose targets were published by the origin (41 to 167 errors, pooled across countries). We report coverage and width for both. The 2025 target is reported separately and labelled as previously inspected.

5. Results

5.1 Exposure-dependent association (H1)

A simple comparison comes before the regressions. Split the eligible economies into 11 with oil rents of at least 10 percent of GDP in 2000–2009 and 10 with less, and compare median growth in the year after each of the five largest Brent moves (Supplementary Table S11). In three cases the pattern matches the hypothesis. After the 33-log-point rise in 2011, median growth in 2012 was 4.7 percent in the high-exposure group against 2.6 in the low-exposure group; after the 51-point rise in 2021 it was 6.2 against 3.4; and after the 41-point fall in 2020 it was 3.8 against 4.1. In the other two it runs the other way. After the 64-point collapse in 2015, the high-exposure group grew faster in 2016 (3.8 against 1.8). After the 35-point rise in 2022, it grew more slowly in 2023 (1.4 against 3.1). The raw data thus give no consistent picture before any modelling choices are made.

Table 5 reports the explanatory estimates. For growth one year ahead, the coefficient on the Brent–exposure interaction is 0.092 (CR1 standard error 0.043; 95 percent CI 0.003 to 0.181 with $t(20)$; $N = 285$, 21 economies). For a Brent increase of 10 log points, this implies that growth in the following year is about 0.37 percentage points higher in an economy with oil rents of 40 percent of GDP than in one with none. The wild-cluster bootstrap gives $p = .041$. The Driscoll–Kraay standard error is more than three times larger (0.154; $p = .55$), and the corresponding normal-approximation interval runs from −0.210 to 0.394; with zero to four lags the Driscoll–Kraay p-value stays between .53 and .56 (Supplementary Table S7). The exposure-permutation placebo gives $p = .081$. After the Holm adjustment across the three primary tests the p-value is .12. The data are therefore compatible with a positive association under independence across countries, and with a wide range of positive and negative values once correlation across countries is allowed for. Adding lagged growth changes the estimate little (0.097), and so does removing year effects and including the Brent main effect (0.095). For same-year growth the coefficient is smaller and imprecise (0.068; wild $p = .21$; Holm $p = .42$). For growth two years ahead, it is negative and imprecise (−0.064).

Table 5: Exposure-dependent associations with growth: plan specifications

Specification	Coef.	CR1 SE	95% CI (CR1)	CR1 p	Wild p	DK SE	DK p	Holm p	N
E1: Brent × exposure (h = 1)	0.092	0.043	[0.003, 0.181]	.043	.041	0.154	.549	.122	285
E1: Brent × exposure (h = 0)	0.068	0.049	[−0.035, 0.170]	.183	.212	0.081	.401	.424	306
E1: Brent × exposure (h = 2)	−0.064	0.114	[−0.302, 0.175]	.584	.847	0.066	.337	—	264
E1b: plus growth t−1 (h = 1)	0.097	0.049	[−0.005, 0.199]	.061	.038	0.153	.526	—	285
E1b: plus growth t−1 (h = 0)	0.081	0.080	[−0.086, 0.248]	.326	.350	0.077	.295	—	306
E1, no year effects (h = 1)	0.095	0.043	[0.005, 0.185]	.040	.041	0.153	.536	—	285
E2: supply news × exposure (h = 1)	−1.621	1.739	[−5.248, 2.006]	.362	.605	1.850	.381	—	285
E2: unexplained price component × exposure (h = 1)	0.138	0.085	[−0.039, 0.316]	.120	.087	0.192	.470	—	285
E2: supply news × exposure (h = 0)	3.396	1.453	[0.365, 6.426]	.030	.013	2.165	.117	—	306
E2: unexplained price component × exposure (h = 0)	−0.023	0.033	[−0.092, 0.045]	.488	.472	0.118	.844	—	306
E3: Brent × exposure × CC 2002–09 (h = 1)	−0.042	0.039	[−0.124, 0.041]	.305	.354	0.095	.662	.424	285

Specification	Coef.	CR1 SE	95% CI (CR1)	CR1 p	Wild p	DK SE	DK p	Holm p	N
E3: Brent × exposure × CC 2002–09 (h = 0)	−0.049	0.027	[−0.105, 0.007]	.084	.150	0.077	.522	—	306
E3 alt.: Brent × exposure × CC t−1 (h = 1)	−0.002	0.021	[−0.047, 0.043]	.929	.924	0.088	.982	—	285
E3 alt.: Brent × exposure × CC t−1 (h = 0)	−0.060	0.036	[−0.135, 0.015]	.109	.207	0.077	.436	—	306

Note. All models include country and year effects except "no year effects", which includes the Brent main effect. Coefficients are percentage points of growth per unit of the interaction (Brent log change in log points × exposure/100); supply-news coefficients are per unit of the annual sum of monthly shocks; triple interactions are per 10 points of the corruption score. The $t-1$ alternative includes all lower-order terms (equation 2). CR1: country-clustered standard error (small-sample factor $G/(G-1) \times (N-1)/(N-K)$); 95% CI and p-value from $t(20)$. Wild: restricted wild-cluster bootstrap p-value, Webb weights, 9,999 draws (assumes independence across countries). DK: Driscoll–Kraay standard error (2 lags, Bartlett) and normal-approximation p-value (allows correlation across countries; few time periods). Holm: adjusted over the three primary tests (E1 $h = 1$, E1 $h = 0$, E3 $h = 1$) using wild p-values. Exposure-permutation placebo p-values, DK intervals and normal-approximation CR1 values are in Supplementary Table S7. Source: authors' estimates; code s02_explanatory.py.

5.2 Influential observation and robustness

The leave-one-out estimates in Figure 4 show why the one-year-ahead estimate should not be relied on. Dropping any one country leaves it between 0.056 (without Libya) and 0.104. Dropping any one predictor year gives values between −0.049 (without 2011) and 0.145. One observation, Libya with predictor year 2011, dominates the influence diagnostics: its Cook's distance is 0.47, against 0.06 for the next observation, and its standardised influence on the interaction coefficient (DFBETAS) is 1.66. In 2011 Libya's output fell by about half during the civil conflict, and in 2012 it grew by 86.8 percent. The Brent change in 2011 was among the largest positive changes in the sample, and Libya's exposure is among the highest. Re-estimating the model, including the fixed effects, without this observation gives −0.009 (SE 0.071; $N = 284$). Figure 3 shows both partial-regression plots, each computed from its own fixed-effects estimation, so the slope in each panel equals the corresponding coefficient. The rebound is consistent with a recovery of oil production after the conflict, which may confound the association with the contemporaneous price increase; without production and conflict data we cannot separate the two.

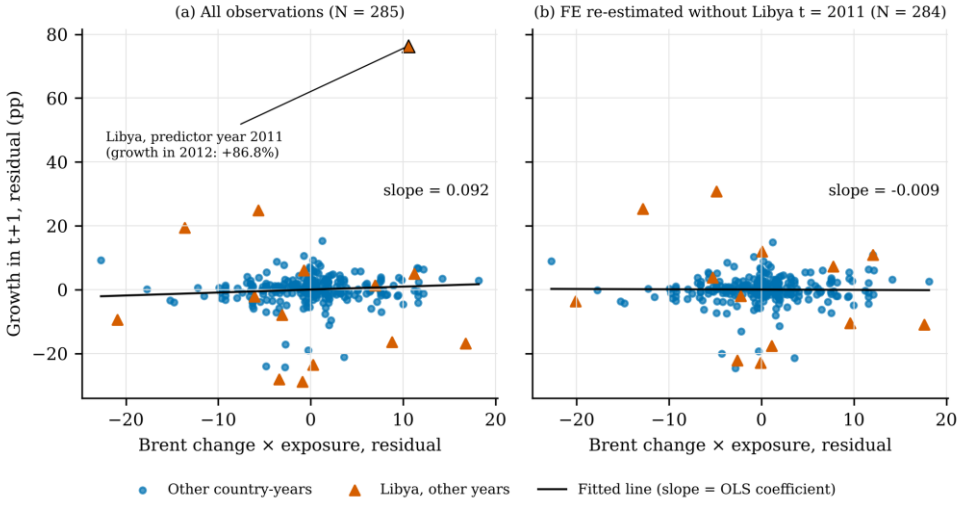

Figure 3: Partial-regression plots for the primary specification, growth in t + 1

Note. Residuals of growth in $t + 1$ and of the Brent–exposure interaction after removing country and year effects. Panel (a): all 285 observations; the highlighted triangle is Libya with predictor year 2011 (growth in 2012: 86.8 percent). Panel

(b): country and year effects re-estimated on the 284 observations without that point. In each panel the fitted line has the slope of the OLS coefficient for that sample (0.092 and −0.009). Source: authors' calculations; code s04_figures.py.

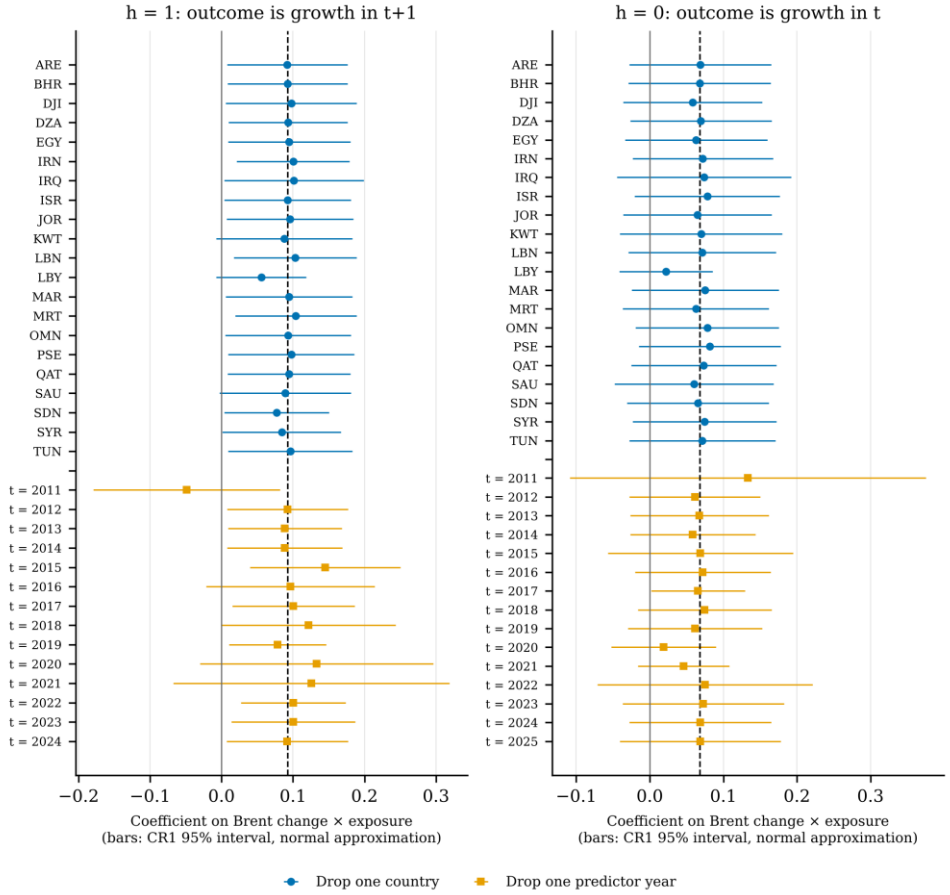

Figure 4: Leave-one-country-out and leave-one-year-out estimates of the Brent–exposure coefficient

Note. Points: coefficient with the named country (circles) or predictor year (squares) removed; bars: 95 percent intervals from CR1 standard errors (normal approximation). Dashed line: full-sample estimate. Source: authors' calculations; code s02_explanatory.py.

Table 6 collects the robustness checks for both primary horizons. Among the pre-specified checks, dropping Israel, as in the 22-economy IMF definition, leaves the one-year-ahead estimate unchanged (0.092), and measuring exposure by oil rents in 2005–2009 gives 0.090. The net-energy-exports measure, which unlike oil rents distinguishes importers from each other, gives a positive coefficient on its own scale (0.011; $p = .029$), but it is available for only 17 economies, and without Libya 2011 it falls to 0.002 (SE 0.006). A binary exporter indicator gives essentially zero (0.001). Excluding the 2020–2021 target years raises the point estimate to 0.113 but widens its interval ($p = .13$), and the pre-COVID sample gives 0.204 with a standard error of 0.166. For same-year growth no pre-specified check gives p below .05 under country clustering; the smallest value, .058, comes from dropping predictor year 2017.

The checks added in revision do not change this picture. Excluding the 18 outcomes for 2025 gives 0.092 ($N = 267$; Driscoll–Kraay $p = .55$), and excluding Libya 2011 as well gives −0.010, so the fragility does not come from the 2025 values. Deflating Brent by the US consumer price index gives 0.094, and −0.014 without Libya 2011. On one common sample of 264 observations for all three horizons (predictor years 2011–2023), the coefficients are 0.076 at $h = 0$, 0.097 at $h = 1$ and −0.064 at $h = 2$ (Supplementary Table S6), so the difference across horizons is not an artefact of different coverage. The disagreement between inference methods also runs in the other direction. At $h = 0$ without Libya 2011 the coefficient is 0.141 with a CR1 p-value of .23 but a Driscoll–Kraay p-

value of .003. We do not read either as decisive; the point is that conclusions at this sample size depend on the assumed dependence structure.

The post-hoc diagnostics point the same way. Excluding outcomes larger than 20 percentage points in absolute value gives 0.035, and excluding Libya, Syria, Sudan and the West Bank and Gaza gives 0.032. A Huber M-estimator gives 0.071 (country-bootstrap SE 0.031). It estimates a different quantity, a typical response that gives less weight to large residuals, so it is a sensitivity analysis for that alternative target rather than a correction of the baseline. Deleting the observation, downweighting it and keeping it answer different questions, and we do not choose among them by their significance.

Table 6: Robustness of the Brent × exposure coefficient

Check	h = 1: coef.	SE	p	N	h = 0: coef.	SE	p	N
Pre-specified checks								
Baseline	0.092	0.043	.043	285	0.068	0.049	.183	306
IMF-only (drop ISR)	0.092	0.045	.055	271	0.078	0.050	.137	291
Drop target years 2020–2021	0.113	0.071	.127	243	−0.015	0.062	.814	264
Sample 2011–2019 (pre-COVID)	0.204	0.166	.234	164	−0.013	0.031	.671	185
Exposure: Oil rents 2005–09	0.090	0.040	.034	285	0.068	0.048	.171	306
Exposure: Net energy exports/use, 2000–09	0.011	0.005	.029	234	0.004	0.005	.425	251
Exposure: binary exporter (rents ≥ 10%)	0.001	0.019	.947	285	0.008	0.016	.630	306
Leave one country out: range	0.056 to 0.104		11/21 < .05		0.022 to 0.081		0/21 < .05	
Leave one year out: range	−0.049 to 0.145		9/14 < .05		0.019 to 0.133		0/15 < .05	
Added in revision (deviation D9)								
Exclude 2025 outcomes	0.092	0.043	.047	267	0.068	0.056	.234	288
Exclude 2025 outcomes and Libya t = 2011	−0.010	0.071	.886	266	0.140	0.119	.255	287
Real Brent (US CPI deflator)	0.094	0.043	.041	285	0.070	0.052	.195	306
Real Brent, drop Libya t = 2011	−0.014	0.077	.860	284	0.148	0.121	.236	305
Post-hoc diagnostics (deviation D2)								
Drop Libya t = 2011 only	−0.009	0.071	.897	284	0.141	0.113	.229	305
Drop Libya t = 2011 and 2020	0.033	0.038	.400	283	0.099	0.076	.212	304
Drop outcomes with |growth| > 20	0.035	0.027	.209	275	0.026	0.037	.491	295
Drop LBY, SYR, SDN and PSE	0.032	0.024	.197	232	0.041	0.031	.211	249
Huber M-estimator (SE: country bootstrap)[a]	0.071	0.031	.023	285	0.041	0.041	.312	306

Note. Country and year effects; CR1 standard errors and $t(G - 1)$ p-values unless stated. Net-energy exports is on a different scale (ratio to energy use) and covers 17 economies. The binary measure multiplies the Brent change by an indicator for rents ≥ 10% of GDP. Real Brent: monthly Brent deflated by the US consumer price index in the Känzig (2021) replication data. Post-hoc diagnostics were added after seeing the results (deviation D2); the checks marked D9 were added in response to review. [a] The Huber M-estimator targets a typical response that downweights large residuals; its SE is the standard deviation of 999 country-cluster pairs-bootstrap estimates and its p-value uses the normal approximation (the default, non-clustered estimator SE is reported in Supplementary Table S9). Source: authors' estimates; code s02_explanatory.py and s02b_posthoc.py.

5.3 Supply news and the unexplained price component (H2)

For growth one year ahead, neither interaction is distinguishable from zero: −1.62 (SE 1.74) for supply news × exposure and 0.138 (SE 0.085; wild $p = .087$) for the unexplained price component × exposure. For same-year growth, the supply-news interaction is positive: 3.40 (SE 1.45; wild $p = .013$). The Driscoll–Kraay p-value is .12, and the result is secondary and was not adjusted for multiplicity. A post-hoc leave-one-out check (36 estimates, each dropping one country or one year) keeps the sign in every case and the estimate between 1.79 and 4.97, with clustered p below .05 (normal approximation) in 34 cases but under Driscoll–Kraay inference in only one. The direction is consistent with price gains from supply news outweighing any output restraint in the more exposed economies in the same year. We regard it as a lead for work with monthly or quarterly data, not as a finding: the supply-news series is estimated with hindsight and is itself measured with error, its annual sum aggregates announcements of opposite sign, and the test is one of several secondary comparisons. Because the two coefficients are in different units, the results do not show that supply-driven and other price changes have different associations with growth.

5.4 Perceived control of corruption (H3)

We find no statistically robust evidence that perceived control of corruption moderates the exposure-dependent association in this sample. The triple interaction with the 2002–2009 score is −0.042 for growth one year ahead (SE 0.039; 95 percent CI −0.124 to 0.041; wild $p = .35$; Holm $p = .42$) and −0.049 for same-year growth (wild $p = .15$). With the time-varying score in $t - 1$ and all its lower-order terms, the coefficients are −0.002 (SE 0.021) and −0.060 (SE 0.036), neither distinguishable from zero; the earlier specification without exposure × score gave 0.002 and −0.060. Government effectiveness gives −0.031 and −0.042 ($p = .30$ and .11). The negative point estimates would mean a smaller exposure-dependent association where perceived corruption control is stronger: the implied one-year-ahead exposure slope is 0.119 (SE 0.061) at the 25th percentile of the 2002–2009 score and 0.039 (SE 0.047) at the 75th (Supplementary Table S10).

These estimates are too imprecise to rule out economically relevant moderation. For a Brent increase of 10 log points, exposure of 40 percent of GDP and a difference of 10 points in the corruption score, the triple interaction implies a growth difference of −0.17 percentage points, with a 95 percent interval from −0.50 to 0.16. The smallest triple interaction that the design would detect with 80 percent power is about 0.12 under country clustering and 0.27 under Driscoll–Kraay errors. When the governance scores are perturbed within their published uncertainty, the triple interaction stays between −0.048 and −0.033 (5th to 95th percentile) with errors independent across years, and between −0.057 and −0.019 with errors fully correlated within countries; its clustered p-value is below .05 in none of the draws. The simulation indicates limited sensitivity to the assumed score perturbations; it does not rule out measurement-error bias or economically relevant moderation.

5.5 Forecast accuracy (FH)

Table 7 reports the primary comparison over 2018–2024 with the corrected timing rule. The ridge model with lagged growth and country indicators (M3) has the lowest MAE, 4.163 percentage points. M3 with the tuned penalty (M3t) is almost identical (4.168), and so is M7, the extended model with the tuned penalty (4.184). With the fixed penalty, adding the Brent–exposure interaction and corruption (M4) raises MAE to 4.303, and adding supply surprises (M8) gives 4.290. M7 has a slightly lower RMSE than M3 (6.915 against 6.926) and M3t the lowest median absolute error (2.004), so the ranking among the top models depends on the loss function; the differences in MAE are small compared with their uncertainty. For M4 against M3 the difference is 0.140, with a country-bootstrap interval of −0.003 to 0.277 and an exploratory t-test p-value of .56 across the seven annual

means; every extended model has a lower MAE than M3 in only three of the seven years. Within the training windows, the tuning chose a penalty of 1,000 for M7 in every origin from 2018 to 2023 and 100 in 2024, which shrinks the added coefficients towards zero. The large penalty is consistent with weak out-of-sample information in the added predictors, but it is not an independent test of that. The no-change forecast (M1) is the least accurate, with an MAE of 6.113, so persistence alone is a weak benchmark in this region; the country mean (4.463) and the pooled autoregression (4.229) are harder to beat. Errors for target years 2020–2021 are about twice those in other years for most models (1.9 to 2.1 times), less so for M6 (1.7) and the no-change forecast (1.1).

Table 7: Retrospective forecasts of growth, target years 2018–2024

Model	MAE	RMSE	MedAE	Bias	ΔMAE vs M3	95% interval (country bootstrap)	t-test p (annual)	Years better than M3	80% cover.
M0 Country mean	4.463	7.456	2.535	1.593	0.300	[−0.185, 0.986]	.181	2/7	0.71
M1 No change (g t−1)	6.113	9.546	3.711	0.790	1.950	[0.999, 3.178]	.054	1/7	0.72
M2 Pooled AR	4.229	6.965	2.110	1.383	0.065	[−0.204, 0.311]	.568	3/7	0.68
M3 Ridge lag + country	4.163	6.926	2.294	1.507	—	—	—	—	0.66
M3t M3 with tuned penalty	4.168	6.936	2.004	1.393	0.005	[−0.296, 0.240]	.942	3/7	0.68
M4 M3 + Brent × exposure + CC	4.303	7.027	2.453	1.900	0.140	[−0.003, 0.277]	.564	3/7	0.64
M5 OLS-FE lag	4.436	7.449	2.413	1.711	0.273	[−0.244, 1.041]	.245	3/7	0.73
M6 OLS-FE extended	5.033	8.035	2.732	2.528	0.870	[0.237, 1.602]	.160	3/7	0.63
M7 M4 with tuned penalty	4.184	6.915	2.122	1.388	0.021	[−0.250, 0.248]	.816	3/7	0.67
M8 M4 + supply surprise × exposure	4.290	7.024	2.539	2.023	0.126	[−0.005, 0.264]	.627	3/7	0.66

Note. 145 identical country-year cases per model; percentage points. Expanding window; at origin T training uses only targets $\leq T - 2$ (published outcomes), for every model, the country mean, penalty tuning and the volatility rule. Bias = mean (forecast − actual). ΔMAE: MAE of the model minus MAE of M3. The interval resamples countries (9,999 draws; ignores common shocks across countries). The *t*-test uses the seven annual mean differences in absolute error (independence across years assumed); it is not a Diebold–Mariano test with a long-run variance. Coverage: intervals from training-residual quantiles. Revised data; exploratory, because these target years were inspected in the pilot. Source: authors' calculations; code s03_forecast.py.

Table 8 separates the predictor groups. With the fixed penalty, adding the Brent change alone leaves MAE unchanged (+0.005), adding the corruption score changes it little (+0.016), and adding the Brent–exposure interaction raises it by 0.133 (country-bootstrap interval 0.032 to 0.233). The deterioration of M4 therefore comes mainly from the exposure interaction, not from the governance score. With the tuned penalty, which shrinks each bundle more heavily, every bundle is within 0.05 of the tuned benchmarks, and the exposure interaction adds 0.020. In the primary window, the corruption score with the tuned penalty is the only specification with a lower MAE than M3 (4.128, a difference of −0.036; −0.041 against the tuned benchmark, interval −0.131 to 0.042). The difference is smaller than its uncertainty, it is one of 11 bundle variants added after review, and we do not treat

it as a forecast gain. At the country level (Figure 6), the interaction raises MAE in 10 of the 11 high-exposure economies, by 0.15 to 0.62 percentage points, the exception being Iraq.

Table 8: Stepwise predictor bundles added to the lagged-growth ridge model, 2018–2024

Predictors (ridge with country indicators)	MAE (α = 10)	RMSE	ΔMAE vs first row [95% interval]	MAE (Tuned α)	RMSE	ΔMAE vs first row [95% interval]
Lagged growth + country indicators (= M3)	4.163	6.926	—	4.168	6.936	—
+ Brent change	4.169	6.922	0.005 [−0.043, 0.049]	4.177	6.878	0.009 [−0.042, 0.067]
+ Brent × exposure	4.296	7.069	0.133 [0.032, 0.233]	4.189	6.908	0.020 [−0.037, 0.089]
+ Brent + Brent × exposure	4.322	7.091	0.159 [0.042, 0.280]	4.208	6.945	0.040 [−0.010, 0.105]
+ corruption score	4.179	6.888	0.016 [−0.086, 0.116]	4.128	6.876	−0.041 [−0.131, 0.042]
+ Brent × exposure + corruption (= M4/M7)	4.303	7.027	0.140 [−0.003, 0.277]	4.184	6.915	0.016 [−0.043, 0.079]
+ Brent + Brent × exposure + corruption	4.320	7.050	0.157 [−0.027, 0.344]	4.184	6.916	0.015 [−0.044, 0.079]

Note. 145 common cases. Tuned α: chosen at each origin by the same inner expanding-window validation for every row, using only published outcomes. Intervals: country-cluster bootstrap, 9,999 draws. Added in response to review (deviation D8). Source: authors' calculations; code s03_forecast.py.

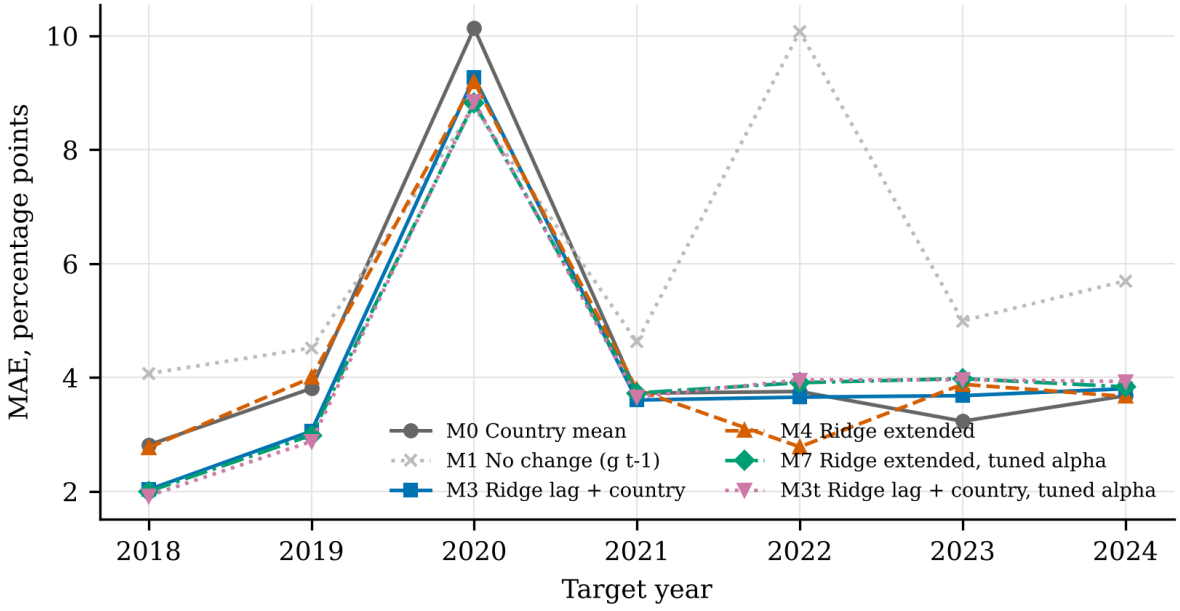

Figure 5: Mean absolute forecast error by target year, 2018–2024

Note. Expanding-window pseudo-out-of-sample forecasts of growth in the target year; training uses outcomes published by the origin (targets up to $T - 2$). Identical country-year cases for every model (about 20 per year). Revised data; exploratory. Source: authors' calculations; code s03_forecast.py.

The 80 percent intervals built from training residuals cover the outcome in 63–73 percent of cases. Intervals calibrated on each model's earlier out-of-sample errors do not fix this: their coverage ranges from 54 to 81 percent, and for the extended models they are more than twice as wide, because early origins with small training samples produced large errors. Coverage is lowest for the five economies with the most volatile growth (39–67 percent for the calibrated intervals, against 57–87 percent for the other 16; Supplementary Table S12). The intervals are therefore descriptive and should not be read as calibrated.

Table 9 shows how the comparison changes across samples. Over 2015–2024 (207 cases) M3 again has the lowest MAE (3.770), and the models with the Brent–exposure interaction deteriorate more: M4 reaches 5.599. Most of the deterioration comes from the 2016 target, whose origin used only 60 training pairs and whose predictor year contains the 63.6-log-point Brent collapse of 2015; the exposure interaction then predicted deep contractions in the exporters that did not occur (M4 MAE 15.8 for 2016, against 2.9 for M3). The IMF-only sample gives the same ranking. Excluding the five economies with the most volatile growth in the data published by the first origin (Libya, Syria, Sudan, Iraq and Iran), the country mean has the lowest MAE (3.283). With the country set defined only from predictors observed by the first origin, M7 and M3t are marginally more accurate than M3 (4.161 and 4.162 against 4.195). For the previously inspected 2025 target (18 cases), all models except the no-change forecast and the pooled autoregression are within about 0.2 percentage points of one another. These sensitivity results confirm that the extended models do not lower MAE in any meaningful way, but also that the ordering among the simpler models is not stable.

Table 9: Mean absolute forecast error across samples

Model	2018–2024 (primary)	2015–2024	IMF-only, 2018–2024	Excl. 5 most volatile	Origin-dated country set	2025 (inspected)
M0 Country mean	4.463	4.433	4.547	**3.283**	4.604	2.139
M1 No change (g t−1)	6.113	5.378	6.179	4.814	6.339	3.982
M2 Pooled AR	4.229	3.812	4.304	3.873	4.219	2.491
M3 Ridge lag + country	**4.163**	**3.770**	**4.240**	3.743	4.195	2.212
M3t M3 with tuned penalty	4.168	3.778	4.242	3.657	4.162	2.054
M4 M3 + Brent × exposure + CC	4.303	5.599	4.395	3.784	4.360	2.236
M5 OLS-FE lag	4.436	5.023	4.530	3.735	4.568	**2.050**
M6 OLS-FE extended	5.033	7.036	5.134	3.775	5.162	2.059
M7 M4 with tuned penalty	4.184	5.391	4.258	3.511	**4.161**	2.102
M8 M4 + supply surprise × exposure	4.290	5.411	4.381	3.870	4.347	2.237
Common cases (n)	145	207	138	112	139	18

Note. Percentage points; lowest value per column in bold. The five most volatile economies, selected from growth published by the 2018 origin (targets ≤ 2016), are LBY, SYR, SDN, IRQ and IRN. The origin-dated country set uses predictors observed by 2016 (adds YEM, drops DJI). The 2025 target was inspected in the pilot and in the supplied workbook, so it is not a holdout. Source: authors' calculations.

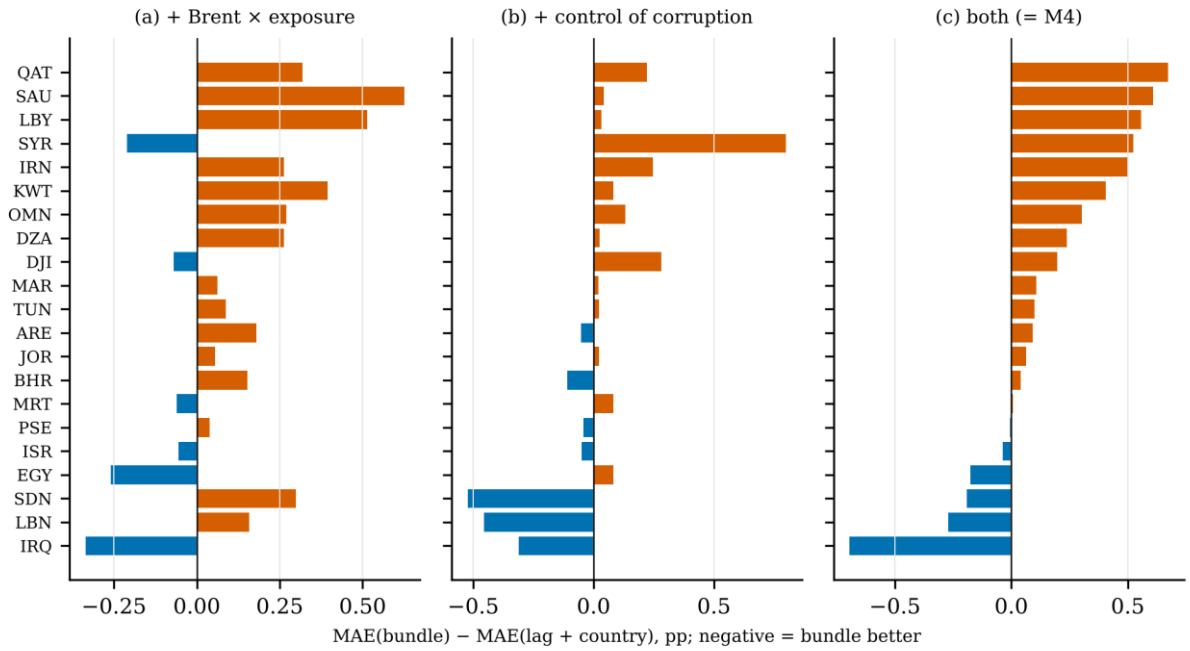

Figure 6: Country-level change in mean absolute error from adding predictor bundles to the lagged-growth ridge model, 2018–2024

Note. Difference between the MAE of the bundle and that of M3 (penalty 10) for each country; negative values: the bundle is more accurate. Panel (c) is M4. Countries are ordered by panel (c). Source: authors' calculations; code s03_forecast.py.

The forecasting hypothesis is not supported: none of the extended models in the plan (M4, M6, M7, M8) achieves a lower mean absolute error than M3 over the primary retrospective evaluation window, and no bundle containing the oil-exposure interaction does so. This statement is specific to the models, the period, the metric and the information rule used here. In line with the plan's decision rule, the article makes no claim of forecast gains.

6. Discussion and Limitations

Table 10 summarizes the evidence for each hypothesis. The central result is that an exposure-dependent association between annual oil-price changes and growth, which many readers would expect, cannot be established robustly with annual data for MENA in 2011–2025. Three features of the setting explain why, and they would apply to similar studies.

Table 10: Summary of evidence for each hypothesis

Hypothesis	Main evidence	Assessment
H1: stronger association with higher exposure (h = 1)	0.092; CR1 95% CI 0.003 to 0.181; DK p = .55; Holm p = .12; −0.009 when re-estimated without Libya t = 2011	No robust evidence; depends on dependence assumptions and one observation
H1: same year (h = 0)	0.068; wild p = .21; Holm p = .42	No evidence
H2: supply news; unexplained price component	h = 0 supply news 3.40 (wild p = .013; DK p = .12); other coefficients imprecise	Not established; lead for higher-frequency work
H3: corruption control moderates the association	−0.042 (95% CI −0.124 to 0.041; Holm p = .42); scenario −0.17 pp (−0.50 to 0.16)	No robust evidence; relevant moderation not ruled out
FH: extended predictors lower MAE	M3 4.163; M4 4.303; M7 4.184; Brent × exposure alone +0.133 (α = 10); corruption with tuned α 4.128 (−0.036; interval includes 0)	Not supported for the planned extensions in this window and model set

Note. Conflict (formerly H4) is treated as a possible explanation for the influential observation, not as a tested hypothesis. Source: Tables 5–9.

First, the identifying variation is thin. With year effects, the information about the interaction comes from 14 annual oil-price changes interacted with one fixed cross-sectional exposure profile, and a few years with large price moves (2011, 2015, 2020, 2021 and 2022) provide most of it. With 21 economies, the uncertainty estimates depend materially on assumptions about cross-country and temporal dependence: under independence across countries the one-year-ahead interval excludes zero, and under Driscoll–Kraay errors it spans values from −0.21 to 0.39. Neither assumption can be verified with this panel, and neither method is reliable at this size. We cannot tell from published descriptions whether the exporter–importer differences reported in earlier MENA panels would survive inference that allows for cross-sectional dependence and influence diagnostics, but our results suggest that such checks should be routine. With the clustered standard error, the smallest interaction detectable with 80 percent power at the 5 percent level is about 0.13; with the Driscoll–Kraay standard error it is about 0.43, which for a 10-log-point Brent rise at 40 percent exposure corresponds to a difference of about 1.7 percentage points in next-year growth. The inconclusive results therefore mean that robust evidence is absent, not that the associations are zero.

Second, in this region oil exposure is correlated with exposure to conflict. The most oil-dependent economies include Libya and Iraq, whose output has been driven for long stretches by war and by the suspension or resumption of production. When such episodes coincide with large price movements, as Libya's 2011 collapse and 2012 rebound coincided with the 2011 price surge, the exposure interaction can pick up conflict and production dynamics; the 2011 Libyan disruption also removed supply from the world market. Country and year effects do not remove this, and dropping the country discards the information that exposure is meant to capture. Conflict is therefore a possible explanation for the influential observation, not a demonstrated one. Testing it requires country-year conflict and production data, which we could not obtain here, and a model of how conflict changes transmission, not the choice of a sample that gives the expected sign.

Third, real GDP measures volumes, and the channel from an oil-price windfall to real activity runs mainly through fiscal policy and its lags. An annual interaction of the price change with exposure is a coarse proxy for that channel. The pattern of point estimates, positive one year ahead and weaker in the same year, is consistent with a lagged fiscal channel (Sadeghi, 2017), but the data cannot distinguish it from zero. The supply-news result for same-year growth, if it holds in higher-frequency data, would fit the view that the nature of the shock matters (Kilian, 2009; Cashin et al., 2014) and that the large producers' income gains from price-raising supply news can outweigh output restraint. We report it as a hypothesis for further work.

The governance result should be read in the same way. It does not show that institutions are irrelevant to how oil windfalls are used. It shows that, in this design, a perceptions-based index measured before 2010 does not robustly predict which economies had larger or smaller exposure-dependent growth associations in 2011–2025. The index may capture general reputation rather than fiscal management; fiscal and financial institutions, such as rules, sovereign funds and financial depth, may matter more than corruption control (El-Anshasy et al., 2017; Jarrett et al., 2019); the 2002–2009 score may be a poor measure of institutions a decade later; and a triple interaction with 21 countries has little power. Our design also differs from Abdelsalam (2023), who includes institutional quality as a control: a smaller oil-price coefficient when institutions are controlled for and a non-significant triple interaction are compatible results. Establishing that institutional moderation is economically negligible would require an equivalence margin fixed in advance and a test designed for it, which this study did not specify.

For forecasting, the result agrees with evidence that simple benchmarks are hard to beat for emerging-economy growth (Celasun et al., 2021). It also shows that the benchmark matters: a model that looks useful against a no-change forecast can be worse than a country mean or a shrunk autoregression. Correcting the training sample so that it uses only published outcomes changed the errors (M3 from 4.108 in the pilot design to 4.163, M4 from 4.313 to 4.303, M6 from 4.805 to 5.033) but did not change the conclusion. The practical implications are limited and follow from this. Scaling a country's expected growth response to an oil-price change by its historical oil rents is not supported by these data, and assessments of oil-price scenarios for MENA economies need country-specific information on production plans, fiscal responses and conflict risk. Growth forecasts for the region should be compared with a country-mean or shrunk autoregressive benchmark, not only with a no-change forecast, and interval forecasts built from in-sample residuals should be presented with their empirical coverage. We draw no implications for fiscal rules or anti-corruption policy from the governance result.

Several further limitations bound these conclusions. The data are revised, not real-time, and the outcomes to 2024 and the 2025 targets had been seen before the plan was written, so the forecast comparison is retrospective and exploratory. Oil rents measure exposure among producers but not the degree of energy import dependence among the others, and the alternative net-energy measure covers only 17 economies. The supply-news series is estimated with hindsight and aggregated to annual frequency. National accounts for Libya, Syria, Sudan, Yemen and the West Bank and Gaza are subject to large revisions and uncertainty, and their growth rates dominate the dispersion of outcomes: of the 285 primary observations, 10 have growth above 20 percent in absolute value, five of them Libyan. Somalia and Yemen are excluded by the coverage rule; their inclusion under a looser rule is not tested. The monthly Pink Sheet prices were taken from the supplied workbook; the World Bank file could not be re-downloaded in the environment used for this revision, so their re-verification is recorded in the Supplementary Material as an external check rather than one we repeated.

A stronger test would state the exposure measure, the interaction, the inference method and the decision rule before the outcomes are seen, and evaluate them on outcomes that did not yet exist when the plan was written. The first such outcome is 2026 GDP growth, which the World Bank will first publish in 2027. The test would use the data vintages available at each forecast origin, and it would include country-year conflict and oil-production measures.

7. Conclusion

This study evaluates exposure-dependent associations between annual oil-price changes and economic growth in MENA, and the retrospective forecasting value of oil-related predictors. The baseline one-year-ahead association is positive under country clustering, but it is sensitive to assumptions about cross-country dependence and to an influential Libyan recovery observation, without which it is close to zero. The data do not provide robust evidence of moderation by perceived control of corruption, but the estimates remain too imprecise to rule out economically relevant effects. In the evaluated forecast window, and with training restricted to outcomes published at each origin, the tested oil-related extensions do not reduce mean absolute error relative to the lagged-growth benchmark. These conclusions are specific to the sample, the exposure measure, the information assumptions and the model set. Stronger causal or real-time claims require a different design and additional data: first-release vintages, outcomes not yet observed, and direct measures of conflict and oil production.

References

Abdelsalam, M. A. M. (2023). Oil price fluctuations and economic growth: The case of MENA countries. *Review of Economics and Political Science, 8*(5), 353–379. https://doi.org/10.1108/REPS-12-2019-0162

Antonakakis, N., Cunado, J., Filis, G., & Perez de Gracia, F. (2017). Oil dependence, quality of political institutions and economic growth: A panel VAR approach. *Resources Policy, 53*, 147–163. https://doi.org/10.1016/j.resourpol.2017.06.005

Apergis, N., & Payne, J. E. (2014). The oil curse, institutional quality, and growth in MENA countries: Evidence from time-varying cointegration. *Energy Economics, 46*, 1–9. https://doi.org/10.1016/j.eneco.2014.08.026

Arezki, R., & Brückner, M. (2011). Oil rents, corruption, and state stability: Evidence from panel data regressions. *European Economic Review, 55*(7), 955–963. https://doi.org/10.1016/j.euroecorev.2011.03.004

Arezki, R., Ramey, V. A., & Sheng, L. (2017). News shocks in open economies: Evidence from giant oil discoveries. *Quarterly Journal of Economics, 132*(1), 103–155. https://doi.org/10.1093/qje/qjw030

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review, 109*(5), 1873–1910. https://doi.org/10.1257/aer.20151569

Berument, M. H., Ceylan, N. B., & Dogan, N. (2010). The impact of oil price shocks on the economic growth of selected MENA countries. *The Energy Journal, 31*(1), 149–176. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol31-No1-7

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics, 90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Cashin, P., Mohaddes, K., Raissi, M., & Raissi, M. (2014). The differential effects of oil demand and supply shocks on the global economy. *Energy Economics, 44*, 113–134. https://doi.org/10.1016/j.eneco.2014.03.014

Cavalcanti, T. V. de V., Mohaddes, K., & Raissi, M. (2011). Growth, development and natural resources: New evidence using a heterogeneous panel analysis. *Quarterly Review of Economics and Finance, 51*(4), 305–318. https://doi.org/10.1016/j.qref.2011.07.007

Celasun, O., Lee, J., Mrkaic, M., & Timmermann, A. (2021). *An evaluation of World Economic Outlook growth forecasts, 2004–17* (IMF Working Paper No. 2021/216). International Monetary Fund. https://doi.org/10.5089/9781513587172.001

Collier, P., & Goderis, B. (2012). Commodity prices and growth: An empirical investigation. *European Economic Review, 56*(6), 1241–1260. https://doi.org/10.1016/j.euroecorev.2012.04.002

Davies, S., Pettersson, T., & Öberg, M. (2026). Organized violence 1989–2025, and violent political protests. *Journal of Peace Research, 63*(4), 705–723. https://doi.org/10.1093/jopres/xjag046

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics, 80*(4), 549–560. https://doi.org/10.1162/003465398557825

El-Anshasy, A., Mohaddes, K., & Nugent, J. B. (2017). *Oil, volatility and institutions: Cross-country evidence from major oil producers* (Globalization and Monetary Policy Institute Working Paper No. 310). Federal Reserve Bank of Dallas. https://doi.org/10.24149/gwp310

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy, 91*(2), 228–248. https://doi.org/10.1086/261140

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

Jarrett, U., Mohaddes, K., & Mohtadi, H. (2019). Oil price volatility, financial institutions and economic growth. *Energy Policy, 126*, 131–144. https://doi.org/10.1016/j.enpol.2018.10.068

Jiang, C. N., Jain, A., Romano, T., & Boyce, T. (2025). *The Worldwide Governance Indicators: Revised methodology for measuring governance using perception data* (WGI Modeling and Methodology Paper). World Bank Group. https://www.worldbank.org/content/dam/sites/govindicators/doc/The%20Worldwide%20Governance%20Indicators%202025%20Methodology%20Revision.pdf

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review, 111*(4), 1092–1125. https://doi.org/10.1257/aer.20190964

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law, 3*(2), 220–246. https://doi.org/10.1017/S1876404511200046

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review, 99*(3), 1053–1069. https://doi.org/10.1257/aer.99.3.1053

Mehlum, H., Moene, K., & Torvik, R. (2006). Institutions and the resource curse. *The Economic Journal, 116*(508), 1–20. https://doi.org/10.1111/j.1468-0297.2006.01045.x

Mohaddes, K., & Pesaran, M. H. (2016). Country-specific oil supply shocks and the global economy: A counterfactual analysis. *Energy Economics, 59*, 382–399. https://doi.org/10.1016/j.eneco.2016.08.007

Mohaddes, K., & Raissi, M. (2019). The US oil supply revolution and the global economy. *Empirical Economics, 57*(5), 1515–1546. https://doi.org/10.1007/s00181-018-1505-9

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*(6), 1417–1426. https://doi.org/10.2307/1911408

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews, 34*(6–10), 1089–1117. https://doi.org/10.1080/07474938.2014.956623

Raddatz, C. (2007). Are external shocks responsible for the instability of output in low-income countries? *Journal of Development Economics, 84*(1), 155–187. https://doi.org/10.1016/j.jdeveco.2006.11.001

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*(1), 135–158. https://doi.org/10.1111/j.1468-0084.2008.00542.x

Ross, M. L. (2015). What have we learned about the resource curse? *Annual Review of Political Science, 18*(1), 239–259. https://doi.org/10.1146/annurev-polisci-052213-040359

Sachs, J. D., & Warner, A. M. (2001). The curse of natural resources. *European Economic Review, 45*(4–6), 827–838. https://doi.org/10.1016/S0014-2921(01)00125-8

Sadeghi, A. (2017). *Oil price shocks and economic growth in oil-exporting countries: Does the size of government matter?* (IMF Working Paper No. 17/287). International Monetary Fund. https://doi.org/10.5089/9781484335390.001

Sundberg, R., & Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research, 50*(4), 523–532. https://doi.org/10.1177/0022343313484347

Timmermann, A. (2007). An evaluation of the World Economic Outlook forecasts. *IMF Staff Papers, 54*(1), 1–33. https://doi.org/10.1057/palgrave.imfsp.9450007

van der Ploeg, F. (2011). Natural resources: Curse or blessing? *Journal of Economic Literature, 49*(2), 366–420. https://doi.org/10.1257/jel.49.2.366

Webb, M. D. (2023). Reworking wild bootstrap-based inference for clustered errors. *Canadian Journal of Economics, 56*(3), 839–858. https://doi.org/10.1111/caje.12661

World Bank. (2025). *Worldwide Governance Indicators, 2025 update* [Data set]. https://www.worldbank.org/en/publication/worldwide-governance-indicators

World Bank. (2026a). *World development indicators* [Data set; snapshot obtained July 2026]. https://databank.worldbank.org/source/world-development-indicators

World Bank. (2026b). *World Bank commodity price data (The Pink Sheet), monthly* [Data set; file updated September 2, 2026]. https://www.worldbank.org/en/research/commodity-markets

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics, 134*(2), 557–598. https://doi.org/10.1093/qje/qjy029