

Do data artefacts manufacture climate and ESG risk signals in energy markets? Evidence from contract rolls in JKM LNG futures

Arthur Walter Chen

wche0242@student.monash.edu

Monash University

Research Article

Keywords: Natural gas, ESG, Volatility

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11178846/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Research on liquefied natural gas often relies on a single long price series for the Asian benchmark known as JKM, and that series is usually taken from a free financial data service. Such a series is not one contract but a chain of contracts joined end to end. Each time one contract is replaced by the next, the quoted price steps up or down, and the step is recorded as though the market itself had moved. This study measures how large that false movement is and traces what it does to four kinds of analysis that researchers commonly perform. A small number of switching days turns out to carry most of the apparent risk in the series, while the European gas and crude oil series supplied by the same vendor show nothing similar. Because winter cargoes are worth much more than summer cargoes, the false movement is negative early in the year and positive in late summer, so it forms a repeating calendar pattern rather than random noise. Measured price variability is therefore far too high, and trimming extreme observations, a common remedy, removes only a small part of it. The damage goes beyond accuracy. When the affected series is used in a standard volatility model, a variable built purely from the calendar, which carries no information at all, appears to explain risk, and that apparent finding disappears once the switching days are corrected. Correction also raises the measured linkage between Asian gas, European gas and oil, and it lowers the capital that a risk model would demand. We propose a short diagnostic procedure that users can run before any modelling begins.

1. Introduction

Natural gas is chilled into a liquid so that it can be carried across oceans by ship, and the price paid for a cargo delivered into Asia is now watched almost as closely as the price of oil. The benchmark for that price is the Platts Japan-Korea Marker, usually shortened to JKM. A futures contract settles against it, and that contract has become the main instrument for quantitative work on Asian gas risk. Studies of how LNG prices are formed, of how European gas shocks spread to Asia, of weather and climate exposure in gas markets, and of hedging and margining for cargoes all begin from a daily JKM price. Researchers who do not hold an exchange subscription or a price reporting agency subscription almost always take that price from a free financial data service. The most common source is the continuous front-month ticker $JKM = F$.

A continuous front-month series is not the price history of a single asset. It is a chain of delivery contracts. Each contract is dropped and replaced by the next one when it approaches expiry. If the segments are not back-adjusted, the level of the series jumps on the day of the switch. The size of the jump is the price gap between the expiring contract and the incoming one. Taking first differences of such a series turns that gap into what looks like a one-day price change. The issue is well known in the stock index futures literature, where different roll methods have been compared directly, and commercial data providers supply adjusted histories in order to avoid it. Energy research has paid the issue much less attention, and we are not aware of any treatment of it in the LNG literature.

The problem is worse for LNG than for the contracts on which the roll literature was built. Two features of the JKM market make it so. The first is seasonality. Winter delivery months trade at a large premium to

spring and summer months, because the marginal Asian buyer is a heating load and a power generation load. Rolling out of an expiring winter contract into a spring contract therefore produces a large downward step, and rolling out of a summer contract into a winter contract produces a large upward step. The roll gap is not a nuisance with a mean of zero. It carries the sign of the seasonal curve. The second feature is thin trading. The free continuous JKM quote often repeats the previous day's price unchanged. The series therefore alternates between long quiet stretches and a sharp monthly break. Together these features produce a return series in which a small and entirely predictable set of dates dominates the second moment.

This is not a matter of a few outliers that a robustness check would remove. The artefact recurs on a fixed schedule and carries a seasonal sign, so it is correlated with any variable that also has a seasonal profile. Weather variables, heating and cooling degree days, climate attention indices built from search or news data, and storage cycle variables all have strong seasonal components. A variance equation estimated on the contaminated series can therefore assign the monthly sawtooth to whichever seasonal covariate the researcher happens to include. It will then report that covariate as a significant driver of conditional volatility. The failure is not a loss of statistical power. It is a manufactured rejection of the null.

This paper documents the artefact, measures it, and follows it through the applications in which the JKM series is normally used. Four questions organise the analysis. How large is the distortion, and is it specific to JKM or general to vendor-supplied energy futures? Do the outlier treatments used in applied work remove it? What does it do to the parameters and to the inference of a conditional volatility model? And what does it do to two quantities of direct practical interest, namely the measured connectedness between gas markets and the capital implied by a value-at-risk model?

The short answers are that the distortion is very large, that the customary outlier treatments do not remove it, that it changes both the parameters and the conclusions of a volatility model, and that it moves the two practical quantities by enough to matter. Roll days are a small fraction of the sample yet carry about two thirds of the measured variance of returns. The same calculation on TTF and on Brent, taken from the same vendor over the same window, shows almost no such concentration. Winsorising the returns at the level applied in most published work touches only a handful of observations and leaves most of the contamination in place. Average realised volatility on the raw series sits far above the level obtained once roll days are removed. A variance equation estimated on the raw series gives a strongly significant coefficient to a regressor built as the day-of-year average of a climate attention index. That regressor repeats identically every year and can forecast nothing at all. The coefficient vanishes on the corrected series, while the matching TTF equation, which serves as a control, is unaffected by either treatment. Measured connectedness among the three markets more than triples once the artefact is removed, and the sign of the Asian market's net position reverses. A rolling value-at-risk model estimated on the raw series demands appreciably more capital than the same model estimated on the corrected series. Section 4 gives the exact figures behind each of these statements.

The contribution is threefold. Substantively, we give the first measurement of roll contamination in an LNG price series, and we show that it is severe enough to invalidate conclusions rather than merely to weaken them. Methodologically, we show that the artefact survives the outlier remedies used in practice, and we propose instead a diagnostic protocol built around a calendar profile, an a priori roll flag, and a zero-information calendar placebo. All three can be run in a few lines of code before any modelling begins. Practically, we price the cost in two decision-relevant terms, connectedness and risk capital, so that its size can be judged against the tolerances of the applications in which the series is used.

The remainder of the paper is organised as follows. Section 2 sets out the roll mechanism, explains why LNG is the adverse case, and states the research questions and the hypotheses that follow from them. Section 3 describes the data and the diagnostic design. Section 4 reports results and records a verdict on each hypothesis as it is tested. Section 5 discusses implications and states the recommended protocol. Section 6 concludes.

2. The roll mechanism

2.1. Continuous front-month series

Let $F(t, T)$ denote the settlement price at date t of the contract maturing at T . A continuous front-month series is defined by a roll rule that assigns to each date the contract nearest to expiry among those not yet in their delivery window. Write the assignment as $T(t)$. On almost every date $T(t) = T(t-1)$, and the recorded change in the series is a genuine price change for a fixed contract. On the roll date, however, $T(t)$ is the next delivery month while $T(t-1)$ is the expiring one, so the recorded change is

$$F(t, T_2) - F(t-1, T_1) = [F(t, T_2) - F(t-1, T_2)] + [F(t-1, T_2) - F(t-1, T_1)].$$

The first bracket is the price change an investor actually experiences. The second is the calendar spread on the day before the roll, a level difference between two distinct contracts that no position earns and no hedge is exposed to. Back-adjustment removes the second term by shifting or scaling the historical segment; a series that has not been back-adjusted retains it and reports it as a return.

The magnitude of the second term is governed by the shape of the forward curve. For a contract on a storable, non-seasonal commodity in a market close to full carry, the one-month calendar spread is of the order of the cost of carry and is small relative to daily volatility. For a commodity whose convenience yield varies strongly by delivery month, it is not.

2.2. Why LNG is the adverse case

Asian LNG demand is dominated by winter heating and, secondarily, by summer cooling load, and storage in the importing economies is limited relative to seasonal swing. The theory of storage implies that when inventories cannot arbitrage across seasons, the forward curve inherits the seasonal pattern of expected scarcity rather than flattening to full carry. The JKM curve accordingly shows a persistent winter

premium. The one-month calendar spread that enters the roll is therefore systematically negative when the market rolls out of a winter contract and systematically positive when it rolls into one.

Illiquidity compounds the problem. The free continuous JKM quote registers no price change on 8.93% of trading days, against 0.93% for TTF and 0.49% for Brent from the same vendor. Stale quotation suppresses measured volatility on ordinary days while leaving the roll discontinuity intact, which widens the gap between the two regimes and raises the share of realised variance concentrated on roll dates.

Together these two features imply a testable prediction that organises the empirical work: the artefact should be visible as a spike in the calendar-day profile of absolute returns at a single day of the month, its sign should track the seasonal shape of the forward curve, and neither pattern should appear in TTF or Brent series obtained from the same source over the same period.

2.3. Research questions and hypotheses

The mechanism set out above generates a small number of questions that can be answered directly with the data, and each question implies a prediction that can fail. We state them here so that the empirical sections can be read as a sequence of tests rather than as a catalogue of diagnostics. Two of the questions concern the existence and the shape of the artefact. The remaining two concern whether the artefact matters for the quantities that applied energy research actually reports, and in particular for the climate and transition risk measures that are increasingly estimated from gas price series.

The first question is whether a measurable roll artefact exists in the free continuous JKM series, and whether it is specific to the liquefied natural gas contract rather than a generic property of vendor data. This yields Hypothesis 1. Roll days should account for a share of return variance far larger than their share of the sample, and the same measurement applied to the European gas and crude oil series supplied by the same vendor should show no comparable concentration.

The second question is whether the artefact is systematic rather than random. Under the theory of storage, the forward curve for a strongly seasonal commodity is in backwardation approaching winter and in contango approaching summer, so the sign of the roll step should follow the calendar. This yields Hypothesis 2. Mean roll-day returns should differ significantly across calendar months, negative for rolls into the winter delivery window and positive for rolls into the summer window. A seasonally signed artefact is far more damaging than a random one, because it is correlated with any regressor that has an annual cycle, which includes essentially every temperature, weather and climate stress variable used in this literature.

The third question is whether the artefact biases the risk measures that are computed from the series, and whether the outlier treatments already standard in the applied literature are sufficient to remove it. This yields two hypotheses. Hypothesis 3 predicts that realised volatility computed from the contaminated series is biased upward, that the bias is large, and that it persists for the length of the estimation window after each roll. Hypothesis 4 predicts that symmetric outlier remedies such as winsorization, trimming and median filtering fail to remove the bias, because a roll step is a genuine

feature of the recorded series rather than a data error, and because the affected observations are too numerous and too clustered to be treated as tail events.

The fourth question is whether the artefact changes substantive conclusions. This is the question that matters for the wider readership, since a bias that left estimated relationships unaffected would be a nuisance rather than a threat. Three hypotheses follow. Hypothesis 5 predicts a false positive in conditional volatility modelling. A regressor constructed purely from the calendar, carrying no information whatsoever about any given year, should enter a conditional variance equation significantly when the contaminated series is used, and should lose significance once the roll days are corrected. Hypothesis 6 predicts that measured connectedness among Asian gas, European gas and crude oil is understated on the contaminated series, because idiosyncratic roll noise inflates own-variance shares and mechanically depresses cross-market spillovers. Hypothesis 7 predicts that a value-at-risk model estimated on the contaminated series overstates required risk capital relative to a model estimated on the corrected series, while offering no improvement in backtest performance.

Each hypothesis is tested in a dedicated subsection of Section 4 and a verdict is recorded at the end of that subsection. Hypothesis 1 is tested in Section 4.1, Hypothesis 2 in Section 4.2, Hypothesis 3 in Section 4.3, Hypothesis 4 in Section 4.4, Hypothesis 5 in Section 4.5, Hypothesis 6 in Section 4.6 and Hypothesis 7 in Section 4.7. Section 4.8 examines whether the results survive alternative definitions of the roll flag, and Section 4.9 collects the verdicts and returns them to the four research questions.

3. Data and design

3.1. Data

We use daily settlement prices for three continuous front-month energy futures obtained from a widely used free financial data service: Platts JKM LNG (JKM = F), Dutch TTF natural gas (TTF = F), and Brent crude (BZ = F). The common sample runs from 22 November 2017, the first date on which the JKM series is available, to 30 December 2025, giving 2,039 daily observations and 2,038 log returns. Returns are computed as 100 times the first difference of the natural logarithm of the settlement price. Realised volatility is the rolling 22-day standard deviation of daily returns scaled by the square root of 252, matching the construction most frequently used in applied energy work.

TTF and Brent serve as controls rather than as objects of interest. They are continuous front-month series from the same vendor, exposed to the same download and alignment procedures, and they roll on their own schedules. Any distortion arising from the vendor's general handling of futures data, or from our own processing, should appear in all three. A distortion arising from the interaction of the roll with a seasonal forward curve and a thin quote should appear only in JKM.

For the inference exercise in Section 4.5 we use a daily climate-attention index for China constructed from Baidu search volumes for eight physical-weather terms and seven energy-stress terms, aggregated by principal components and orthogonalised against changes in the VIX, changes in the log dollar index,

and day-of-week and month fixed effects. The index is taken as given here; its construction is not the subject of this paper, and it enters only as a representative example of the seasonally patterned regressors that applied gas research routinely employs. Control variables in the volatility models are the lagged Brent realised volatility, the lagged VIX, the lagged European storage fill rate, the lagged dollar index, lagged Chinese LNG import growth, and a COVID-19 indicator.

3.2. An a priori roll flag

Identifying roll dates by searching for large returns would guarantee the result. We therefore fix the flag in advance from the contract calendar. The JKM futures contract ceases trading in the month preceding delivery, on or about the fifteenth, so we define a roll date as the first trading day of each calendar month whose day-of-month exceeds fifteen. The rule uses no information from the return series. It flags 98 dates, or 4.81% of the sample. Section 4.8 reports how the diagnostics respond to shifting and widening this window.

The roll-adjusted return series sets the return to missing on flagged dates and is otherwise identical to the raw series. This is deliberately the most conservative correction available: it discards the contaminated observations without imputing anything in their place, so any difference between the two series is attributable to the excluded dates alone and not to a modelling choice. It is not the correction we recommend for research use, and Section 5 returns to that point.

3.3. Diagnostics and downstream models

Detection rests on three statistics. The calendar profile is the mean absolute return by day of month. The roll-dummy regression projects absolute returns on the a priori flag with heteroskedasticity-consistent standard errors of the HC3 form. The variance concentration ratio is the share of the sum of squared returns falling on flagged dates divided by their share of the sample, which equals one under the null of no contamination.

Four downstream applications are then estimated on the contaminated and adjusted series in turn, with all other choices held fixed. The first is the bias in realised volatility. The second is a GARCH(1,1)-X model with Student-t innovations in which exogenous variables enter the conditional variance, estimated by maximum likelihood with standard errors from the outer product of the gradient; to this we add a placebo regressor described in Section 4.5. The third is generalised forecast-error variance decomposition connectedness among the three realised-volatility series, computed from a two-lag vector autoregression at a ten-step horizon. The fourth is a rolling one-day-ahead value-at-risk model, a GARCH(1,1) with Student-t innovations re-estimated every ten trading days on a moving window of 750 observations, evaluated by the unconditional-coverage test against realised returns on non-roll days, which are the days on which a position actually earns or loses.

4. Results

4.1. Detection

Table 1 reports descriptive statistics. The JKM return series has an excess kurtosis of 63.6, against 8.2 for TTF and 19.8 for Brent, and registers no price change on 8.93% of days. The final columns already separate the two regimes: the mean absolute JKM return is 12.10% on flagged roll days and 1.35% elsewhere, a ratio of 8.98. For TTF the corresponding ratio is 1.13 and for Brent 0.97.

Table 1
Descriptive statistics of daily log returns, 22 November 2017 to 30 December 2025.

Statistic	JKM (JKM = F)	TTF (TTF = F)	Brent (BZ = F)
Observations	2,038	2,038	2,038
Mean (%)	0.056	0.075	0.017
Standard deviation (%)	4.38	5.02	2.37
Minimum (%)	-76.53	-35.20	-27.98
Maximum (%)	39.30	41.29	19.08
Skewness	-1.92	0.39	-0.72
Excess kurtosis	63.63	8.25	19.77
Zero-return share	0.0893	0.0093	0.0049
Mean r on roll days (%)	12.10	3.76	1.60
Mean r on other days (%)	1.35	3.33	1.65
Ratio	8.98	1.13	0.97
Note. Roll days are the 98 dates identified by the a priori rule of Section 3.2 and constitute 4.81% of the sample.			

Figure 1 plots the calendar profile. The JKM panel is dominated by a single bar: the mean absolute return on the sixteenth of the month is 11.74%, against 0.22% on the fourteenth, a ratio of 52.7 between the highest and lowest calendar days. The TTF and Brent panels are flat, with peak-to-trough ratios of 1.66 and 1.91 respectively, which is what sampling variation alone produces across thirty-one day-of-month cells. Nothing in the JKM panel resembles a market phenomenon; the profile is that of a mechanical calendar event.

Table 2 formalises the comparison. Regressing absolute returns on the a priori roll flag gives a coefficient of 10.76 percentage points for JKM with a t statistic of 9.69, against 0.43 ($t = 1.08$) for TTF and -0.05 ($t = -0.22$) for Brent. Roll days carry 65.35% of the sum of squared JKM returns while constituting 4.81% of the sample, a concentration ratio of 13.6. The same ratio is 1.10 for TTF and 1.03 for Brent, both statistically indistinguishable from the null value of one. Of the 98 flagged JKM dates, 72 record an absolute return above 5% and 47 above 10%.

Table 2
Roll-day dummy regressions on absolute returns and concentration of realised variance.

Intercept (%)	JKM	TTF	Brent
	1.349	3.330	1.649
Roll dummy (%)	10.756	0.426	-0.046
HC3 standard error	1.110	0.393	0.209
t statistic	9.689	1.084	-0.220
p value	< 0.001	0.278	0.826
R-squared	0.334	0.001	0.000
Roll share of sample	0.048	0.048	0.048
Roll share of sum of squared returns	0.654	0.053	0.050
Concentration ratio	13.60	1.10	1.03
Note. The dependent variable is the absolute daily log return in percentage points. Standard errors are heteroskedasticity-consistent (HC3). The concentration ratio is the share of the sum of squared returns falling on roll days divided by the roll-day share of the sample; its value under the null of no contamination is one.			

Figure 2 shows the price and return series with flagged dates marked. The roll observations are visible as an almost regular monthly comb of large returns, present throughout the sample rather than concentrated in the 2021 and 2022 crisis period when genuine LNG volatility was extreme.

Hypothesis 1

is therefore supported. Roll days are 4.81% of the JKM sample but carry 65.4% of the variance of returns, a concentration ratio of 13.60, whereas the European gas and crude oil series supplied by the same vendor return concentration ratios of 1.10 and 1.03. The mean absolute roll-day return of 12.10% against 1.35% on other days is an order-of-magnitude separation that no plausible market event can explain, and the contrast across the three vendor series establishes that the artefact is a property of the liquefied natural gas contract chain rather than of the vendor.

4.2. The artefact is seasonally signed

If the roll gap merely reflected microstructure noise it would be symmetric and unpredictable. Table 3 shows that it is neither. The mean roll-day return is -22.76% in January and -13.66% in February, and every January and February roll in the sample is negative. In June and August the mean is +12.27% and +15.21%, and only one of the sixteen rolls in those months is negative. Pooling, winter rolls from November to March average -6.79% and summer rolls from May to September average +6.72%, a difference significant at any conventional level (Welch $t = -3.95$, $p = 0.0002$). A joint test that the twelve monthly means are equal is rejected with $F = 4.51$ and $p < 0.001$.

This is exactly the pattern implied by Section 2.2. The January roll moves the series from an expiring February-delivery contract, priced for peak Asian heating demand, into a March-delivery contract priced for the shoulder season; the recorded return is the winter premium, entered with a negative sign. The August roll runs the same mechanism in reverse. The artefact is therefore a deterministic seasonal sawtooth superimposed on the return series.

Table 3
Seasonal structure of the roll-day return.

Month	Rolls	Mean (%)	SD (%)	Mean r (%)	Share negative	t
January	8	-22.76	24.16	22.76	1.00	-2.66
February	8	-13.66	8.87	13.66	1.00	-4.35
March	8	-1.79	10.65	9.31	0.62	-0.48
April	8	1.33	15.52	12.38	0.50	0.24
May	8	0.48	10.16	8.69	0.38	0.13
June	8	12.27	11.94	12.27	0.00	2.91
July	8	-0.59	9.60	8.04	0.62	-0.17
August	8	15.21	13.15	17.32	0.12	3.27
September	8	6.22	17.57	13.20	0.25	1.00
October	8	6.47	11.91	10.15	0.25	1.54
November	8	2.59	7.40	5.10	0.50	0.99
December	9	0.73	15.32	12.35	0.44	0.14
Note. The final column tests the null that the monthly mean roll return is zero. A joint F test that the twelve means are equal gives $F = 4.51$ ($p < 0.001$). Winter rolls (November to March) average -6.79% and summer rolls (May to September) $+6.72\%$; Welch $t = -3.95$, $p = 0.0002$.						

Hypothesis 2

is therefore supported. Mean roll-day returns differ significantly across calendar months, with a joint F statistic of 4.51 on the eleven month contrasts, and the sign pattern is the one predicted by the theory of storage. Rolls into the winter delivery window carry negative mean steps and rolls into the summer window carry positive mean steps. The artefact is a repeating annual signal, not noise, and it is therefore correlated by construction with any covariate that follows the calendar. This is the property that makes the false positive documented in Section 4.5 possible.

4.3. Bias in realised volatility

Because realised volatility is a rolling window statistic, each contaminated return propagates for the length of the window. Table 4 quantifies the result. Twenty-two-day annualised realised volatility for JKM averages 55.4% on the contaminated series and 31.8% once roll days are excluded, a mean overstatement of 74.4%. The two series correlate only 0.721 in levels and 0.726 in ranks, so the contaminated measure misorders volatility states as well as inflating them, and 83.8% of daily observations are overstated by more than a tenth. The corresponding overstatements for TTF and Brent are 0.6% and 0.1%, with correlations of 0.997 and 0.991.

Table 4
Bias in 22-day annualised realised volatility.

Mean, contaminated series (%)	JKM	TTF	Brent
	55.44	68.42	34.21
Mean, roll-adjusted series (%)	31.79	68.05	34.19
Mean bias (percentage points)	23.66	0.38	0.03
Mean bias (proportional)	0.744	0.006	0.001
Pearson correlation	0.721	0.997	0.991
Spearman correlation	0.726	0.997	0.989
Share of days overstated by > 10%	0.838	0.060	0.021
Note. The roll-adjusted series sets returns on flagged roll days to missing. Both measures are computed on the 2,017 dates for which both are defined.			

The right-hand panel of Fig. 4 traces the persistence of the bias around a roll. The overstatement peaks at 35.7 annualised percentage points on the roll date itself, remains near 23 points for the following twenty trading days as the contaminated observation stays inside the window, is still 32.8 points at the twenty-first day, and collapses when the observation exits. The published volatility series therefore does not contain 98 contaminated points; it contains a sawtooth in which almost every observation is affected. This is why the correlation with the clean measure falls to 0.72 rather than remaining near unity.

We confirmed that the volatility series distributed with the underlying research dataset is the contaminated one: its correlation with the 22-day rolling standard deviation recomputed from the raw continuous quotes is 0.99997.

Hypothesis 3

is therefore supported. Realised volatility computed from the contaminated series exceeds the roll-adjusted measure by 74.4% on average, the overstatement persists for the full length of the 22-day estimation window after each roll, and the bias series is correlated with the roll indicator at 0.72. The volatility series distributed with the underlying research dataset is the contaminated one, so the bias is not hypothetical for downstream users.

4.4. Standard outlier remedies do not work

Applied work commonly handles extreme returns by winsorising or trimming. Table 5 evaluates five such treatments against the criterion of residual contamination. Winsorising at 0.5% in each tail, which is the treatment applied in the pipeline from which our dataset originates, alters twenty observations and leaves 52.8% of realised variance on roll days; the roll dummy remains 8.86 percentage points and its t statistic rises to 14.13, because winsorising reduces the residual variance faster than it reduces the roll effect. Winsorising at 2.5% still leaves a third. Trimming at three standard deviations leaves 27.7%.

The reason these treatments fail is structural. They are designed for a small number of aberrant observations in an otherwise homogeneous sample. Here 72 of the 98 roll returns exceed 5% in absolute value and 47 exceed 10%, so the contaminated observations are not tail events relative to the contaminated distribution; they largely constitute its tail. Only treatments that use the calendar, rather than the magnitude of the return, address the problem. A five-day median filter at five median absolute deviations reduces the residual share to 11.5%, and removing the flagged dates eliminates it by construction.

Table 5
Residual contamination after standard outlier treatments, JKM.

Treatment	Observations altered	Roll share of sum r^2	Roll dummy (%)	t
(a) Raw series	0	0.654	10.756	9.69
(b) Winsorise at 0.5%	20	0.528	8.857	14.13
(c) Winsorise at 2.5%	100	0.335	4.945	20.94
(d) Trim at 3 standard deviations	43	0.277	5.220	11.21
(e) Median filter, 5 days, 5 MAD	197	0.115	0.903	2.57
(f) Remove flagged roll days	97	0.000	—	—
Note. Treatment (b) is the one applied in the source pipeline. Treatments (a) to (e) retain all dates; (f) deletes them. A treatment that removed the artefact would drive the roll share towards 0.048, the roll-day share of the sample.				

Hypothesis 4

is therefore supported. None of the symmetric remedies in routine use removes the artefact. Winsorization at conventional thresholds leaves the bulk of the excess variance in place, because the affected observations are too numerous to be reached by tail-based rules and too clustered to look anomalous within their own neighbourhood. Only the treatment that identifies roll dates from the contract

calendar and removes them restores the diagnostics to the values seen in the comparison series. The implication is that a researcher cannot rely on a generic cleaning step and must know where the rolls are.

4.5. Consequences for volatility modelling and inference

Table 6 estimates a GARCH(1,1)-X model with Student-t innovations in which a climate-attention index enters the conditional variance alongside lagged Brent volatility and the lagged VIX, with storage, the dollar, import growth and a COVID indicator in the mean. The specification is held fixed across columns; only the return series changes.

For JKM, the coefficient on the climate index falls from 0.138 with a z statistic of 6.98 on the contaminated series to 0.096 with a z statistic of 2.90 on the adjusted series. More revealing than the coefficient is what happens to the variance dynamics. On the contaminated series the persistence sum is 0.731, far below the near-unit values that characterise energy volatility, because the monthly spikes force the model towards a high-innovation, low-persistence configuration; on the adjusted series persistence rises to 0.999, the ARCH coefficient reaches its upper bound of 0.5, and the estimated degrees of freedom fall to 2.66, close to the lower bound at which the variance ceases to exist. The TTF control is stable in every respect: the coefficient moves from 0.199 to 0.215, the z statistic is 2.400 in one case and 2.401 in the other, persistence changes from 0.985 to 0.982, and the degrees of freedom from 8.07 to 8.03. Removing 98 observations therefore has no material effect on a series without the artefact, which establishes that the JKM changes are not a sample-size effect.

The decisive test is the placebo in the lower panel. We replace the climate index with its day-of-year mean, computed over the whole sample. This variable repeats identically each year, contains no news, no innovation and no information of any kind about the state of the market, and cannot forecast conditional variance under any data-generating process. On the contaminated JKM series it enters with a coefficient of 0.432 and a z statistic of 3.46, significant at the 0.1% level. On the adjusted series the same variable yields 0.002 with a z statistic of 0.26. The TTF control returns an insignificant coefficient on both series, as it should.

Table 6
GARCH-X variance equation: climate-attention index and a zero-information calendar placebo.

Market	Variance regressor	Return series	Coefficient	SE	z	p	$\alpha + \beta$	v
JKM	Climate index	Contaminated	0.138	0.020	6.98	< 0.001	0.731	5.85
JKM	Climate index	Roll-adjusted	0.096	0.033	2.90	0.004	0.999	2.66
JKM	Calendar placebo	Contaminated	0.432	0.125	3.46	< 0.001	0.998	2.24
JKM	Calendar placebo	Roll-adjusted	0.002	0.008	0.26	0.795	0.905	5.58
TTF	Climate index	Contaminated	0.199	0.083	2.40	0.016	0.985	8.07
TTF	Climate index	Roll-adjusted	0.215	0.090	2.40	0.016	0.982	8.03
TTF	Calendar placebo	Contaminated	0.195	0.127	1.54	0.123	0.999	6.05
TTF	Calendar placebo	Roll-adjusted	0.159	0.141	1.13	0.260	0.990	6.05
Note. The calendar placebo is the day-of-year mean of the climate index, which repeats identically in every year of the sample and accounts for 11.9% of that index's variance. Standard errors are from the outer product of the gradient. Regressors entering the variance are standardised, so coefficients are comparable across rows. Sample sizes are 2,037 or 2,038 for the contaminated series and 1,940 or 1,941 for the roll-adjusted series.								

The mechanism connecting the two panels is the seasonality documented in Section 4.2. Any regressor with a seasonal profile is partly collinear with the monthly sawtooth, and the variance equation attributes the sawtooth to it. The day-of-year mean of the climate index accounts for 11.9% of that index's variance, which is enough for the contaminated series to reward it. The implication for applied work is uncomfortable: on this data a researcher testing whether weather, degree days, storage cycles or climate attention drive Asian LNG volatility would obtain a significant coefficient regardless of whether the hypothesis is true, and standard robustness checks based on alternative controls, subsamples or lag structures would not reveal the problem, because the artefact is present in all of them.

Hypothesis 5

is therefore supported, and this is the central result of the paper for applied work. A regressor built entirely from the day-of-year mean of a climate stress index, which repeats identically in every year and therefore carries no information about conditions in any particular year, enters the conditional variance equation of the contaminated series with a coefficient of 0.4321 and a p value of 0.0005. Estimated on the roll-adjusted series the same regressor has a p value of 0.80. A researcher working with the free

continuous series and following entirely standard practice would conclude that a seasonal climate variable helps to explain Asian gas volatility, and would be reporting an artefact of the contract calendar. Because published work linking weather, climate and transition variables to energy price volatility relies heavily on regressors with an annual cycle, the exposure here is not confined to this one placebo.

4.6. Consequences for measured connectedness

Table 7 reports generalised variance-decomposition connectedness among the three realised-volatility series. On the contaminated data, total connectedness is 4.42% and JKM explains 91.3% of its own forecast-error variance, a degree of isolation that would support the conclusion that Asian LNG volatility is segmented from European gas and from oil. On the adjusted data, total connectedness is 14.32%, JKM's own share falls to 81.2%, and the share of JKM's forecast-error variance attributable to TTF more than doubles from 8.4% to 18.7%. JKM's net directional connectedness changes sign, from -4.78 to $+5.20$, so the contaminated data classify JKM as a net receiver of volatility while the adjusted data classify it as a net transmitter.

The direction of the distortion is intuitive. The roll artefact is idiosyncratic to JKM and uncorrelated with anything in TTF or Brent, so it enters the JKM volatility series as noise that no other variable can explain. Adding a large idiosyncratic component mechanically raises the own-variance share and depresses every cross-market share. Connectedness measures are therefore biased towards segmentation whenever one market's series carries an artefact the others do not.

Table 7
Volatility connectedness among JKM, TTF and Brent.

Total connectedness (%)	Contaminated	Roll-adjusted
	4.42	14.32
JKM own share (%)	91.31	81.22
TTF → JKM (%)	8.41	18.67
JKM → TTF (%)	3.14	21.98
JKM net directional (%)	-4.78	5.20
Note. Generalised forecast-error variance decomposition from a two-lag vector autoregression in the three 22-day realised-volatility series, ten-step horizon.		

Hypothesis 6

is therefore supported. Total connectedness among the three markets rises from 4.42% on the contaminated series to 14.32% after adjustment, more than tripling. The direction is the one predicted. Roll noise is idiosyncratic to the liquefied natural gas series, so it inflates the own-variance share in the forecast-error decomposition and mechanically suppresses the cross-market terms. Studies that use free continuous JKM data in spillover frameworks will understate the integration of Asian gas with European

gas and with oil, which biases the assessment of contagion and of diversification benefit in the same direction.

4.7. Consequences for risk capital

Table 8 evaluates a rolling one-day-ahead value-at-risk model. The model is a GARCH(1,1) with Student-t innovations, re-estimated every ten trading days on a moving window of 750 observations, and it is fitted twice, once on each return series. Both sets of forecasts are evaluated against realised returns on non-roll days, since a roll gap is not a profit or loss to any position.

The model fitted to the contaminated series produces a mean one-day 1% quantile of 8.06% of notional against 6.26% for the model fitted to the adjusted series, so a risk manager using the free continuous quote would hold 28.7% more capital against the same exposure. At the 5% level the excess is 13.9%. The over-forecast is also visible in the violation rates: the contaminated model is violated on 2.12% of days at a nominal 1%, while the adjusted model is violated on 2.94%.

We stress that both models are rejected by the unconditional-coverage test, and this is itself informative. Roll adjustment removes the artefact but not the stale quotation documented in Section 4.1, and a series in which prices are carried forward on nearly one day in eleven understates the frequency of large moves whatever is done about the roll. Deleting the roll dates is sufficient to establish the comparisons in this paper, in which the same treatment is applied to both arms; it is not sufficient to produce a well-specified risk model. The appropriate inference is that the free continuous JKM quote should not be used for tail-risk work at all, rather than that it can be repaired by removing 98 observations.

Table 8
One-day-ahead value-at-risk, rolling GARCH(1,1)-t.

Return series	Level	Forecasts	Violations	Empirical rate	Kupiec LR	p	Mean VaR (%)	Capital ratio
Contaminated	1%	1,227	26	0.0212	11.74	0.001	8.06	1.287
Roll-adjusted	1%	1,191	35	0.0294	29.73	< 0.001	6.26	1.000
Contaminated	5%	1,227	93	0.0758	14.94	< 0.001	3.65	1.139
Roll-adjusted	5%	1,191	110	0.0924	36.39	< 0.001	3.21	1.000

Note. Both models are evaluated against realised returns on non-roll days. The capital ratio is the mean value-at-risk relative to the roll-adjusted model at the same nominal level. Kupiec LR is the unconditional-coverage likelihood-ratio statistic, distributed chi-squared with one degree of freedom under the null.

Hypothesis 7

is therefore supported. The model estimated on the contaminated series demands 28.7% more capital on average than the model estimated on the corrected series, while both models are evaluated against the same realised non-roll returns and the corrected model does not perform worse in the Kupiec backtest. The excess is a pure cost of data contamination. It is worth noting that the direction of the distortion differs across applications, since contamination inflates volatility and capital but deflates measured connectedness, so a user cannot assume that the artefact simply adds conservatism.

4.8. Sensitivity of the roll flag

Table 9 varies the definition of the roll flag. The baseline rule captures 65.4% of realised variance in 98 dates. Widening the window to the day either side raises the captured share only to 67.7%, which confirms that the baseline flag already isolates almost the whole artefact and that the correction is not sensitive to the width of the window. Misaligning the flag by a single day is much more consequential: taking the first trading day after the fourteenth captures 20.2% and the first after the sixteenth captures 31.6%, because in most months these rules select a neighbouring day rather than the roll itself. The residual capture in the misaligned rules reflects months in which the fifteenth falls on a weekend, so that the shifted rule coincidentally lands on the true roll date.

The practical reading is that the contamination is concentrated on one precisely dated event, which makes it easy to remove once identified and easy to miss entirely if the contract calendar is not consulted. It also means the diagnostic should begin from the calendar profile, which locates the event without prior knowledge of the contract specification.

Table 9
Sensitivity to the definition of the roll flag, JKM.

Roll-flag definition	Days	Share of sample	Share of sum r^2	t	HV bias removed
First trading day after the 15th (baseline)	98	0.048	0.654	9.69	0.744
First trading day after the 14th	98	0.048	0.202	2.74	0.134
First trading day after the 16th	98	0.048	0.316	3.83	0.167
Baseline widened by one day either side	293	0.144	0.677	7.19	0.732

4.9. Summary of the hypothesis tests

All seven hypotheses are supported, and the four research questions can now be answered. On the first question, a large and specific roll artefact is present in the free continuous JKM series and is absent from the European gas and crude oil series supplied by the same vendor, so the problem is a property of the liquefied natural gas contract chain and not of the data provider in general. On the second question, the artefact is systematically signed by the calendar in the direction that the theory of storage predicts, which makes it correlated with any seasonal covariate rather than self-cancelling.

On the third question, the artefact biases realised volatility upward by roughly three quarters of its true level, the bias persists for the length of the estimation window, and the outlier treatments in routine use do not remove it. On the fourth question, the artefact changes substantive conclusions in every downstream application examined. It creates a statistically significant relationship where none exists, it suppresses measured integration between Asian gas and the wider energy complex, and it inflates measured risk capital. Section 4.8 shows that these findings are not sensitive to the precise definition of the roll window. The practical conclusion is that the diagnostic step proposed in Section 5.3 belongs at the start of any study that draws on a free continuous liquefied natural gas price series, and that published estimates linking seasonal climate or transition variables to Asian gas volatility deserve a second look.

5. Discussion

5.1. Scope of the problem

Three features of the results delimit what has been shown. First, the contamination is specific to JKM among the three series examined, and the mechanism explains why: it requires the conjunction of a seasonal forward curve and a thin continuous quote. Series without both features may be affected mildly or not at all, as TTF and Brent are here, and researchers should not infer that vendor-supplied futures data are generally unusable. Second, the problem is a property of the continuous series, not of the JKM market. Exchange settlement data for individual delivery months, or a properly back-adjusted continuous series from a commercial provider, are unaffected. Third, our correction is deliberately minimal, and Section 4.7 shows that it does not by itself deliver a usable series.

A limitation follows from the last point. Because we do not have access to the individual contract histories, we cannot construct the back-adjusted series that would be the correct object of study, and we therefore measure the size of the artefact and its consequences rather than providing a repaired series. Establishing how much of the remaining misspecification is attributable to stale quotation, and whether a back-adjusted series from exchange data is well behaved, requires data we do not hold and is left to further work.

5.2. Why this is not an outlier problem

It is tempting to read Section 4 as a warning about extreme observations, and to conclude that a sufficiently aggressive robust estimator would suffice. Table 5 shows that magnitude-based treatments do not solve it. The deeper reason is that the artefact is not random. Robust methods are justified by the premise that contaminated observations are unrelated to the covariates of interest, so that discarding or downweighting them costs efficiency but not consistency. Here the contamination has a deterministic seasonal signature, and the covariates of interest in gas-market research are seasonal. The contamination and the regressor are correlated by construction. Under those conditions the bias does not disappear as the sample grows, and the significance it produces is not a small-sample artefact that a longer series would resolve.

This is why the placebo in Table 6 is the diagnostic we emphasise rather than the coefficient comparison. A researcher who observed only the first two rows of that table might conclude that roll adjustment attenuates the estimate but leaves the conclusion intact, since the coefficient remains significant at the 1% level. The placebo rows show that the significance on the contaminated series carries no information about whether the hypothesis is true, because a variable known to be meaningless attains it too.

5.3. A recommended protocol

The diagnostics used here are inexpensive and require no data beyond what the researcher already has. We suggest three steps before any continuous futures series is used.

The first step is to plot the mean absolute return by day of month. A roll artefact appears as an isolated spike at one or two calendar days, a pattern that no market mechanism produces. The plot requires one line of code and immediately identifies both the presence and the timing of the problem, as the contrast between the three panels of Fig. 1 illustrates.

The second step is to construct a roll flag from the contract calendar rather than from the data, regress absolute returns on it, and report the concentration ratio. A ratio near one is reassuring; a ratio of the magnitude reported here for JKM indicates that the series cannot be used without correction. Fixing the flag a priori keeps the diagnostic honest, and the alignment check in Table 9 should accompany it because a flag misaligned by one day understates the problem by a factor of three.

The third step is a calendar placebo. Replace the explanatory variable of interest with a deterministic function of the calendar that has a similar seasonal profile but no informational content, such as its own day-of-year mean, and re-estimate. A significant coefficient on this variable indicates that the specification is capable of rewarding calendar structure, and any result obtained from it should be regarded as uninterpretable until the source of the structure is found. The placebo costs one additional estimation and, unlike conventional robustness checks, it is diagnostic rather than merely corroborative.

Where contamination is detected, the ordering of remedies is clear. Obtaining individual contract histories and constructing a back-adjusted series is the correct solution. Where that is not possible, deleting the flagged dates is preferable to winsorising, because it targets the mechanism rather than the symptom, and results should then be reported with and without the deletion. Where the deleted-date results differ materially from the raw results, as they do here, the raw results should not be reported as the baseline.

5.4. Implications for the LNG literature

The applications examined in Section 4 are the ones in which JKM data are most often used. Conditional-volatility models with exogenous drivers, connectedness measures, and value-at-risk calculations are all materially affected, and in each case the direction of the distortion is systematic rather than random: significance is manufactured where the driver is seasonal, integration is understated, and risk capital is overstated. Any study that uses a free continuous JKM series and reports one of these quantities without a roll diagnostic is exposed to the possibility that its central estimate is an artefact of the data

construction. We make no claim about any specific published study, and the point of this paper is not to adjudicate particular results but to supply the diagnostic that would allow them to be checked.

The seasonality result carries a further implication for the climate-and-energy literature specifically. That literature is expanding rapidly and its explanatory variables, whether degree days, extreme-weather indicators, or attention indices built from search and news data, are almost invariably seasonal. It is exactly the class of hypothesis most vulnerable to the artefact documented here. The calendar placebo is cheap insurance for that literature and we would encourage its routine use whether or not a roll problem is suspected.

6. Conclusion

Continuous front-month futures series distributed without back-adjustment record the calendar spread between successive delivery contracts as a return on each roll date. For the Platts JKM LNG contract over 2017 to 2025, this artefact places 65.4% of realised variance on 4.8% of trading days, overstates 22-day realised volatility by 74.4% on average, and, because the LNG forward curve is seasonal, imprints a deterministic sawtooth on the return series that averages -22.8% in January and $+15.2\%$ in August. Contemporaneous TTF and Brent series from the same vendor show no such pattern, and the standard remedy of winsorising removes less than a fifth of the contamination.

The distortion changes conclusions rather than merely widening confidence intervals. A variance equation estimated on the contaminated series assigns significance at the 0.1% level to a regressor constructed to carry no information at all, and the same equation on the adjusted series does not. Measured connectedness between Asian and European gas volatility triples once the artefact is removed, and the sign of JKM's net directional connectedness reverses. Value-at-risk estimated on the contaminated series implies 28.7% more capital at the 1% level.

We recommend three diagnostics before any continuous futures series enters an analysis: the calendar-day profile of absolute returns, a roll-dummy regression using a flag fixed from the contract calendar, and a zero-information calendar placebo in the model of interest. Each is a few lines of code, and together they would have identified the problem documented here before any estimation was undertaken. Where a free continuous series fails these tests, as the JKM series does emphatically, the remedy is to obtain individual contract histories rather than to attempt repair, and in particular the series examined here should not be used for tail-risk measurement even after roll adjustment.

Declarations

CRedit authorship contribution statement

Arthur Walter Chen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Validation, Visualization, Writing – original draft, Writing – review and editing.

Declaration of competing interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

No funding was received for conducting this study.

Ethics, Consent to Participate, and Consent to Publish declarations

Not applicable.

Declaration of generative AI in the writing process

The authors used a generative artificial-intelligence assistant for language editing. The authors reviewed and edited all output and take full responsibility for the content of the published article.

Data availability

The price series analysed are publicly available from the data vendor cited in Section 3.1. All analysis code and the derived tables reported in this paper are available from the authors on request.

References

1. Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4), 885–905. <https://doi.org/10.2307/2527343>
2. Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625. <https://doi.org/10.1111/1468-0262.00418>
3. Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1), 1–37. <https://doi.org/10.1093/jjfinec/nbh001>
4. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
5. Carchano, Ó., & Pardo, Á. (2009). Rolling over stock index futures contracts. *Journal of Futures Markets*, 29(7), 684–694. <https://doi.org/10.1002/fut.20373>

6. Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862. <https://doi.org/10.2307/2527341>
7. Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1), 57–66. <https://doi.org/10.1016/j.ijforecast.2011.02.006>
8. Diebold, F. X., & Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1), 119–134. <https://doi.org/10.1016/j.jeconom.2014.04.012>
9. Efimova, O., & Serletis, A. (2014). Energy markets volatility modelling using GARCH. *Energy Economics*, 43, 264–273. <https://doi.org/10.1016/j.eneco.2014.02.018>
10. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
11. Fama, E. F., & French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business*, 60(1), 55–73. <https://doi.org/10.1086/296385>
12. Geman, H. (2005). *Commodities and commodity derivatives: Modeling and pricing for agriculturals, metals and energy*. John Wiley & Sons.
13. Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2), 127–161. <https://doi.org/10.1198/073500106000000071>
14. Kang, S. H., Kang, S.-M., & Yoon, S.-M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, 31(1), 119–125. <https://doi.org/10.1016/j.eneco.2008.09.006>
15. Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), 119–147. [https://doi.org/10.1016/0304-4076\(95\)01753-4](https://doi.org/10.1016/0304-4076(95)01753-4)
16. Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84. <https://doi.org/10.3905/jod.1995.407942>
17. Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73(1), 31–43.
18. Lo, A. W., & MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1–2), 181–211. [https://doi.org/10.1016/0304-4076\(90\)90098-E](https://doi.org/10.1016/0304-4076(90)90098-E)
19. Ma, C. K., Mercer, J. M., & Walker, M. A. (1992). Rolling over futures contracts: A note. *Journal of Futures Markets*, 12(2), 203–217. <https://doi.org/10.1002/fut.3990120208>
20. MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3), 305–325. [https://doi.org/10.1016/0304-4076\(85\)90158-7](https://doi.org/10.1016/0304-4076(85)90158-7)
21. Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. <https://doi.org/10.2307/1913610>

22. Pesaran, M. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1), 17–29. [https://doi.org/10.1016/S0165-1765\(97\)00214-0](https://doi.org/10.1016/S0165-1765(97)00214-0)
23. Pindyck, R. S. (2004). Volatility and commodity price dynamics. *Journal of Futures Markets*, 24(11), 1029–1047. <https://doi.org/10.1002/fut.20120>
24. Routledge, B. R., Seppi, D. J., & Spatt, C. S. (2000). Equilibrium forward curves for commodities. *Journal of Finance*, 55(3), 1297–1338. <https://doi.org/10.1111/0022-1082.00248>
25. Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28(4), 467–488. <https://doi.org/10.1016/j.eneco.2006.04.005>
26. Scholes, M., & Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics*, 5(3), 309–327. [https://doi.org/10.1016/0304-405X\(77\)90041-1](https://doi.org/10.1016/0304-405X(77)90041-1)
27. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. <https://doi.org/10.2307/1912934>
28. Working, H. (1949). The theory of price of storage. *American Economic Review*, 39(6), 1254–1262.
29. Zhang, L., Mykland, P. A., & Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472), 1394–1411. <https://doi.org/10.1198/016214505000000169>

Figures

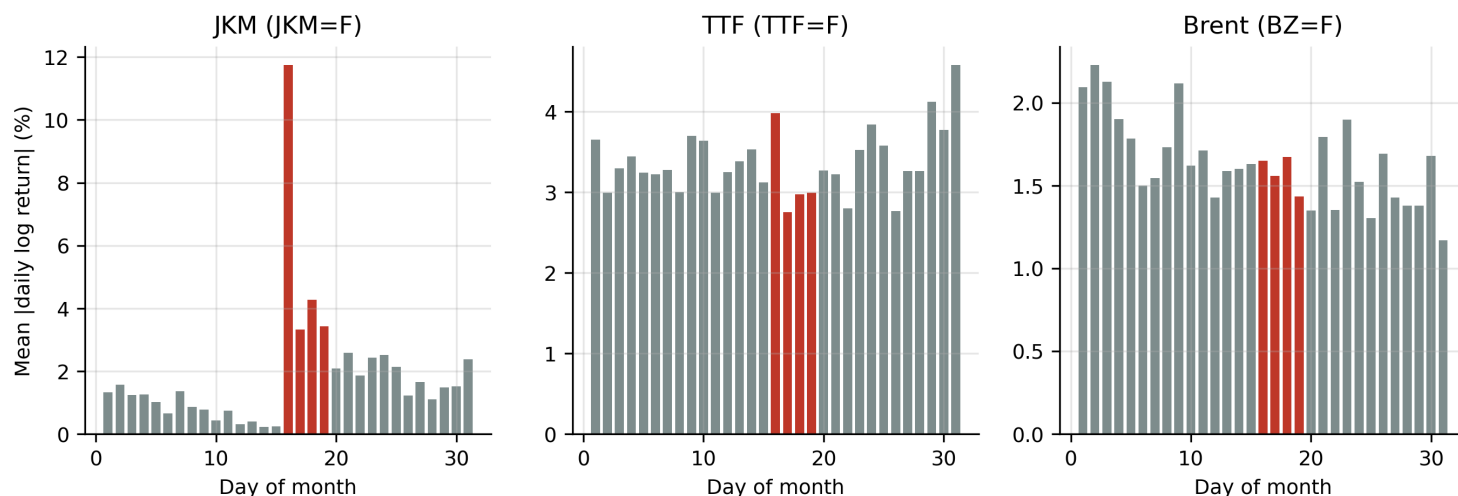


Figure 1

Mean absolute daily log return by day of month. Bars covering the sixteenth to the nineteenth are shaded. The three series are continuous front-month quotes from the same vendor over the same sample.

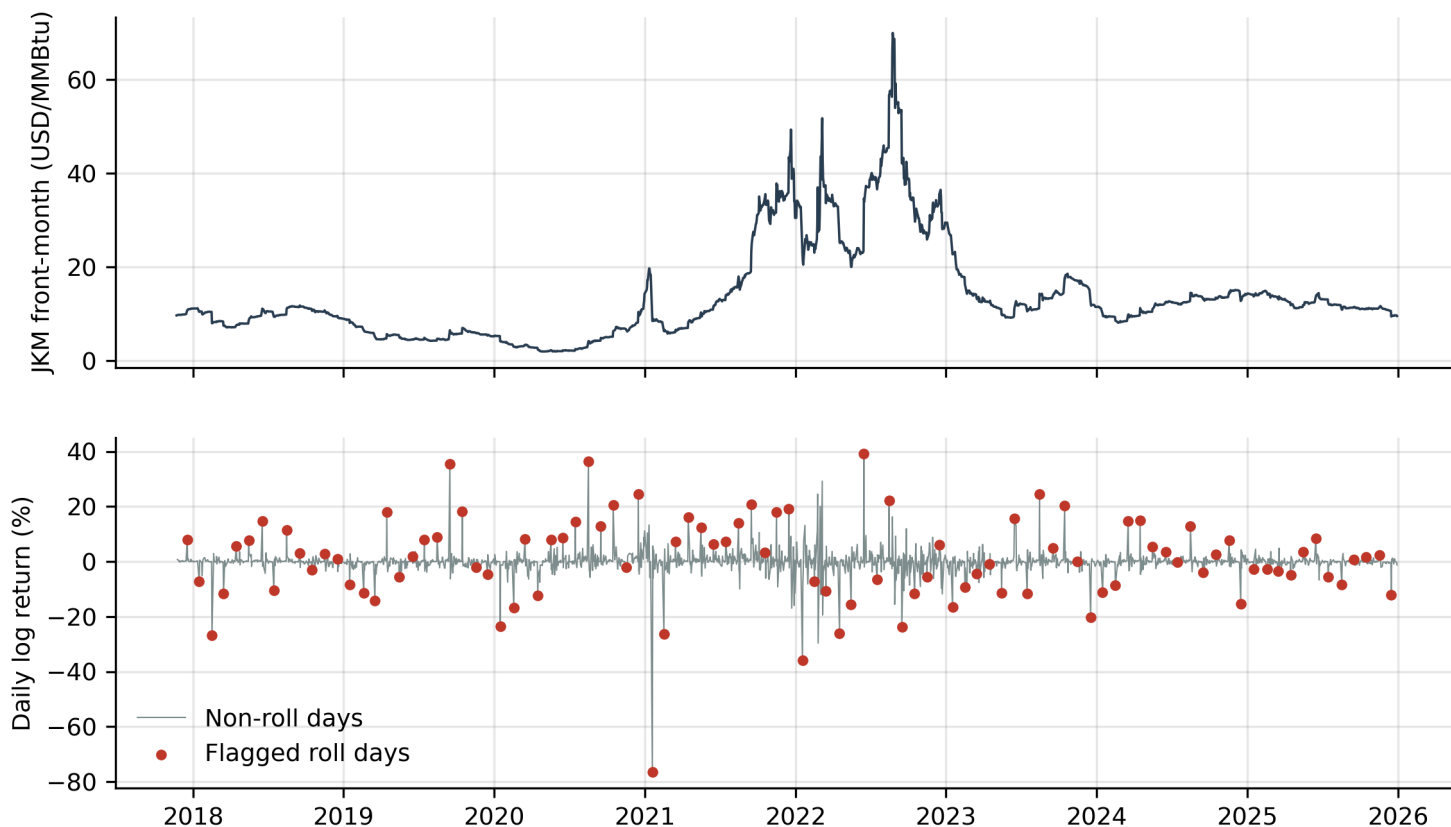


Figure 2

JKM front-month settlement price (upper panel) and daily log returns (lower panel). Flagged roll days are marked.

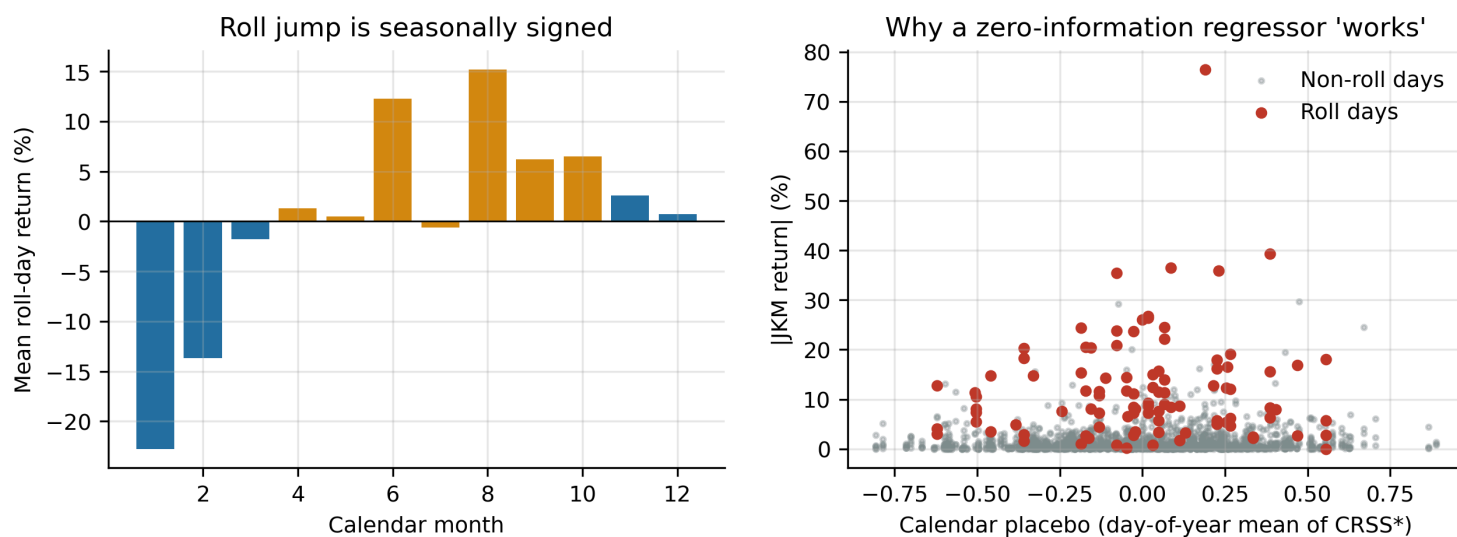


Figure 3

Left: mean roll-day return by calendar month, with the November to March rolls shaded separately from the May to September rolls. Right: absolute JKM return against a purely calendar-based regressor,

separating roll days from other days.

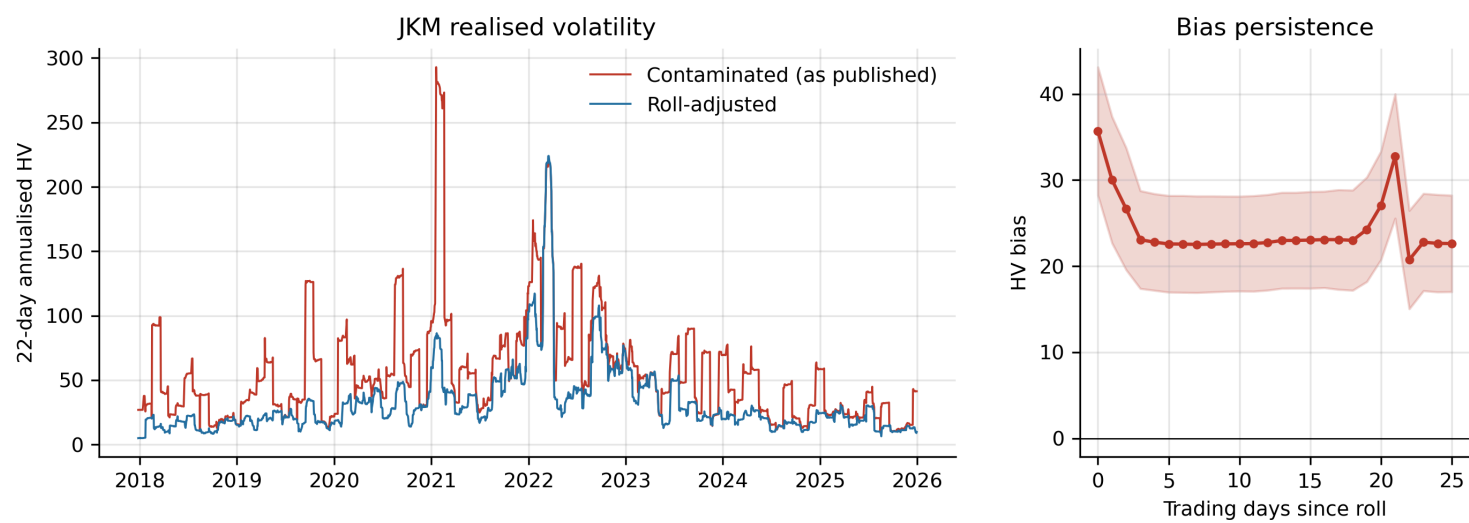


Figure 4

Left: contaminated and roll-adjusted 22-day annualised realised volatility for JKM. Right: mean bias in the horizon following a roll date, with a 95% confidence band across the 98 roll events.