Call the Zookeeper: A Unified Framework for Commodity Risk Premiums*

John Fan[a], Xihang Li[b], Xiao Qiao[b], Tingxi Zhang[c]

[a]Griffith University [b]City University of Hong Kong [c]Curtin University

Presenter: Riven (TINGXI ZHANG)
August 2025 @ JPMCCEM

*This paper was previously titled as "**Call the Zookeeper: Dissecting Commodity Pricing Theories**"

WHAT WE KNOW

- **Theory of Normal Backwardation** links variation in risk premiums with *hedging pressure* (Bessembinder, 1992; De Roon et al., 2000; Basu and Miffre, 2013), movements in *open interest* (Hong and Yogo, 2012), speculators' *risk capacity* (Acharya et al., 2013), *selective hedging* behavior (Fernandez-Perez et al., 2018).

- **Theory of Storage** links risk premiums with *inventory* levels (Deaton and Laroque, 1992; Gorton et al., 2013), or *convenience yield* (Gibson and Schwartz, 1990; Liu and Tang, 2011; Gu et al., 2024), *volatility* (Dewally et al., 2013), *financialization* (Basak and Pavlova, 2016).

- Most empirical studies focus on one specific explanation, limited agreement on which theory dominates.

WHAT WE DID

- Jointly evaluate multiple explanations of commodity futures risk premiums.
- Investigate return predictability in both time-series and cross-sectional dimensions
- First gauge the **degree** of return predictability, then address the **relative importance** of competing explanations.
- Document the **statistical** and **economic** significance of the predictability.

WHY IT IS INTERESTING-(I)

- Discussions of commodity risk premia (Keynes, 1930; Kaldor, 1939) predates efficient markets. Despite economic importance, commodity futures have drawn far fewer studies than traditional asset classes.

- Classical literature emphasizes **time-series influence** of hedgers' positions and inventory dynamics on futures prices (Cootner, 1960; Hirshleifer, 1988; Bessembinder, 1992; De Roon et al., 2000).

- More recent literature underscores the role of inventory proxies in explaining **cross-sectional variation** in commodity returns (Gorton et al., 2013; Bakshi et al., 2019; Gu et al., 2024).

- Studying both allows us to identify factors that simultaneously forecast aggregate returns and explain why certain commodities earn higher average returns.

WHY IT IS INTERESTING-(II)

- Most studies rely on a few predictors (e.g., momentum, basis) and a single method, typically OLS regressions.
- We leveraged the richest dataset and various models
  - 25 commodities across 30+ years
  - 81 predictors covering commodity-specific, macroeconomy and market sentiment
  - SIX linear and FOUR nonlinear models, as well as TWO ensemble models
- Validation across ten distinct methods ensures that findings are robust and generalize across time, sectors, and model classes.

WHAT WE FIND

- Predictability in both time-series (TS) and cross-sectional (CS) dimensions:
  - TS market-timing strategy delivers up to 0.56 annualized Sharpe (vs.0.22 of AVG)
  - CS long-short portfolio yields up to 0.78 annualized Sharpe (vs. 0.40 of traditional factors)
- Theory of Normal Backwardation plays a stronger (not dominating) role in both TS and CS.
- While macro variables play nonnegligible roles in predicting aggregate returns; commodity-specific factors explain more cross-sectional variation, whereas the influence of news-based sentiment is limited.
- Linear ensemble models perform better than nonlinear models, likely due to the underlying data structure.

SO WHAT?

- Zookeeper prevails—Integrating multiple theoretical perspectives yields superior explanatory power compared to any single approach, supporting a unified commodity pricing framework.
- Commodity pricing studies should prioritize evaluating the relative importance of return drivers across theories, variable segments, and return dimensions.
- Parsimony vs. Complexity: Kelly et al. (2024) highlight a "virtue of complexity" in return prediction, yet Buncic (2025) shows a simple ridge model (~15 predictors) outperforms complex ML models in cross-sectional analysis. Cartea et al. (2025) note that noisy predictors lead to diminishing returns from added complexity.

Commodity Data

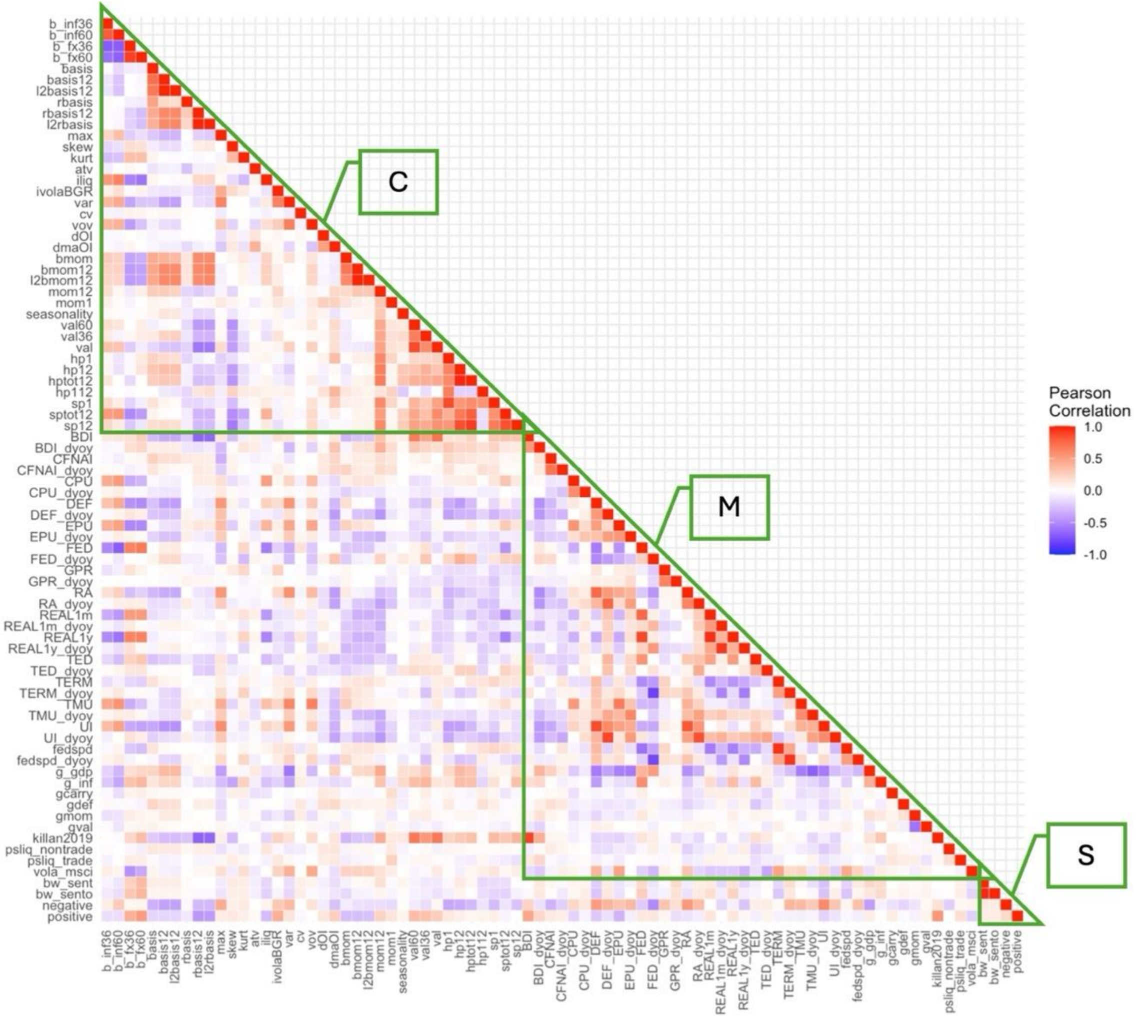
- 25 major commodity markets from all sectors:
  - Softs: coffee, cocoa, sugar, orange juice
  - Grains: wheat, oats, corn, rough rice
  - Oilseeds: soybean, soybean oil, soybean meal
  - Meats: feeder cattle, lean hogs, live cattle
  - Precious metals: gold, copper, silver, platinum, palladium
  - Industrial material: cotton, lumber
  - Energy: heating oil, crude oil, gasoline, natural gas

- Daily settlement prices, volume, open interest, lot size of all maturities in 1979-2021. CFTC Commitment of Traders (CoT)

Predictor Overview

Group	Category	Example Variables / Features
C [37]	HP and Storage	Momentum, Carry, Relative Basis, Seasonality, Hedging/Spec Pressure, Skewness, Open Interest Changes
	Transformations	Smoothed, Lagged (per Kang et al., 2020; Ederington et al., 2021)
	Liquidity & Volatility	Basis-Momentum, Amivest Illiquidity, Variance, Scaled Variance, Abnormal Volume
	Others	FX/Inflation Exposure, Value, Max Return, Kurtosis, Vol-of-Vol
M [40]	Macro & Systemic Risk	Real Rates, Fed Funds Rate, Term/Default Spread, Fed Funds Spread
	Growth & Inflation	Global GDP, G7 Inflation, CFNAI, Kilian Index, Shipping Costs
	Liquidity & Uncertainty	Pastor-Stambaugh Liquidity, TED Spread, Jurado-Ludvigson-Ng TMU, BDI EPU, Caldara-Iacoviello GPR
	Systematic Factors	AQR Style Factors: Value, Momentum, Carry, Defensive
	Transformations	YoY Changes, Global focus > Single-country (e.g., U.S. or China)
S [4]	Investor Sentiment	Baker-Wurgler Sentiment Index, FinBERT WSJ Sentiment (1984–2021, 822k+ articles, monthly aggregated probabilities)

Correlation Heatmap



The Model

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1}$$

$$E_t(r_{i,t+1}) = g_t(c_{i,t}, m_t, s_t)$$

- $c_{i,t}$: 37 commodity-specific variables (Group "C")
- m_t: 40 macroeconomic variables (Group "M")
- s_t: 4 sentiment variables (Group "S")

- For **time-series** analysis, target variable is an equally weighted market portfolio (AVG) of 25 commodities
  - $c_{i,t}$ is averaged at each time t
- For **cross-sectional** analysis, target variable is commodity-specific return
  - m_t and s_t are replaced by regression betas of commodity i on each m and s variables

- The predictive model g_t includes
  - **Linear models**: Ordinary least square (OLS), Stepwise regression, Partial least squares regression (PLS), Ridge, LASSO, Elastic net
  - **Nonlinear models**: Random Forest, Gradient Boosted Regression Trees(GBRT), Extreme Gradient Boosting (XGBoost), Multilayer perception (MLP)

- Hyperparameters tunning and OOS testing
  - **201 Months** used for training (Apr 1988-Dec 2004)
  - **192 Months** used for OOS testing **(**Jan 2005- Dec 2021**)**

Statistical Performance

OOS R2: Time-series predictability

$$R^2_{OOS} = 1 - \frac{\frac{1}{T-t}\sum_{\tau=t}^{T}(r_\tau - \hat{r}_\tau)^2}{\frac{1}{T-t}\sum_{\tau=t}^{T}(r_\tau - \bar{r}_\tau)^2}$$

r_τ : Actual return of market portfolio (AVG)
$\hat{r}_\tau$: Model predicted return of AVG
$\bar{r}_\tau$: Historical returns average of AVG

	C		M	S	Forecast Combination
Panel A: Linear Models					
OLS	-1.0168		-1.5033	-0.0664	-0.2093
Stepwise	-0.1145		-0.1681	0.0015	0.0192
PLS	-0.0126	<	0.0114	-0.0268	0.0372
Ridge	0.0280	<	0.0381	-0.0033	0.0311
LASSO	0.0119		-0.0425	-0.0235	0.0136
Elastic Net	0.0300	<	0.0381	-0.0232	0.0259
L-En	0.0268		0.0204	-0.0107	0.0355
Panel B: Nonlinear Models					
Random Forest	0.0186	<	0.0278	-0.0041	0.0195
GBRT	0.0243	<	0.0317	-0.0017	0.0265
XGBoost	0.0058	<	0.0130	-0.0061	0.0176
MLP	0.0033	<	0.0036	-0.0048	0.0019
N-En	0.0235		0.0221	-0.0002	0.0185

OOS R2: Cross-sectional predictability

$$R_{OOS}^2 = 1 - \frac{\frac{1}{T-t}\frac{1}{N}\sum_{\tau=t}^{T}\sum_i\left(r_{i,\tau} - \hat{r}_{i,\tau}\right)^2}{\frac{1}{T-t}\frac{1}{N}\sum_{\tau=t}^{T}\sum_i\left(r_{i,\tau} - \bar{r}_{i,\tau}\right)^2}$$

$r_{i,\tau}$: Actual return of commodity i

$\hat{r}_{i,\tau}$: Model predicted return of commodity i

$\bar{r}_{i,\tau}$: Historical average return of of commodity i

	C		M	S	Forecast Combination
Panel A: Linear Models					
OLS	-0.0071		-0.0300	0.0055	0.0032
Stepwise	0.0072	>	-0.0003	0.0066	0.0083
PLS	0.0066	>	-0.0030	0.0057	0.0070
Ridge	0.0047		0.0050	0.0073	0.0079
LASSO	0.0028		0.0031	0.0072	0.0076
Elastic Net	0.0045	>	0.0043	0.0072	0.0077
L-En	0.0064		0.0031	0.0073	0.0081
Panel B: Nonlinear Models					
Random Forest	-0.0031		-0.0056	0.0011	0.0028
GBRT	0.0071	>	0.0061	0.0072	0.0068
XGBoost	0.0007	>	-0.0027	0.0061	0.0054
MLP	0.0068	>	0.0064	0.0064	0.0066
N-En	0.0069		0.0042	0.0067	0.0068

Economic Significance

Time-series market timing strategies

$\alpha_t^* = \frac{1}{\gamma}\frac{\mu_t}{\sigma_t^2}$ Campbell & Thompson (2008), the **optimal weight** on risky portfolio for risk-averse investors, where μ_t is proxied by model predicted return of market portfolio

	AVG	Linear Ensemble				Nonlinear Ensemble			
	Constant Weight	C	M	S	Forecast Combination	C	M	S	Forecast Combination
Panel A: Weight {0, 1}									
Average Exposure	0.71	0.78	0.68	0.64	0.73	0.78	0.71	0.63	0.75
Annualized Mean	2.3%	5.7%	3.2%	0.9%	5.1%	4.8%	3.1%	1.8%	3.9%
t-statistics (Newey-West)	0.8	1.8	1.0	0.3	1.6	1.5	0.9	0.5	1.2
Annualized S.D.	10.5%	11.7%	11.5%	12.6%	12.0%	12.0%	12.2%	12.5%	12.3%
Annualized Sharpe	0.22	0.49	0.27	0.07	0.43	0.40	0.25	0.14	0.32
Annualized Sortino	0.30	0.80	0.41	0.09	0.64	0.63	0.36	0.19	0.45
Skewness	-0.358	0.447	0.108	-0.642	0.136	0.337	0.011	-0.538	-0.009
Maximum Drawdown	-33.1%	-20.3%	-36.7%	-40.1%	-30.6%	-29.7%	-36.8%	-38.9%	-29.0%
CER	-0.6%	2.3%	-0.2%	-3.3%	1.5%	1.3%	-0.7%	-2.4%	0.1%
Panel B: Weight {-1, 1}									
Average Exposure	0.49	0.60	0.39	0.37	0.49	0.62	0.50	0.38	0.57
Annualized Mean	1.5%	7.1%	3.6%	-0.2%	7.9%	6.7%	3.2%	0.7%	5.0%
t-statistics (Newey-West)	0.8	1.9	0.8	-0.1	2.0	1.8	0.9	0.2	1.5
Annualized S.D.	7.1%	13.9%	14.6%	14.2%	14.2%	14.0%	13.5%	14.0%	13.3%
Annualized Sharpe	0.22	0.51	0.24	-0.02	0.56	0.48	0.23	0.05	0.37
Annualized Sortino	0.30	0.89	0.39	-0.02	0.97	0.79	0.36	0.06	0.55
Skewness	-0.358	0.517	0.364	-0.440	0.485	0.461	0.091	-0.368	-0.004
Maximum Drawdown	-23.5%	-21.6%	-43.4%	-49.5%	-30.9%	-27.7%	-43.2%	-42.7%	-24.9%
CER	0.2%	2.4%	-1.7%	-5.6%	3.0%	1.9%	-1.4%	-4.6%	0.5%

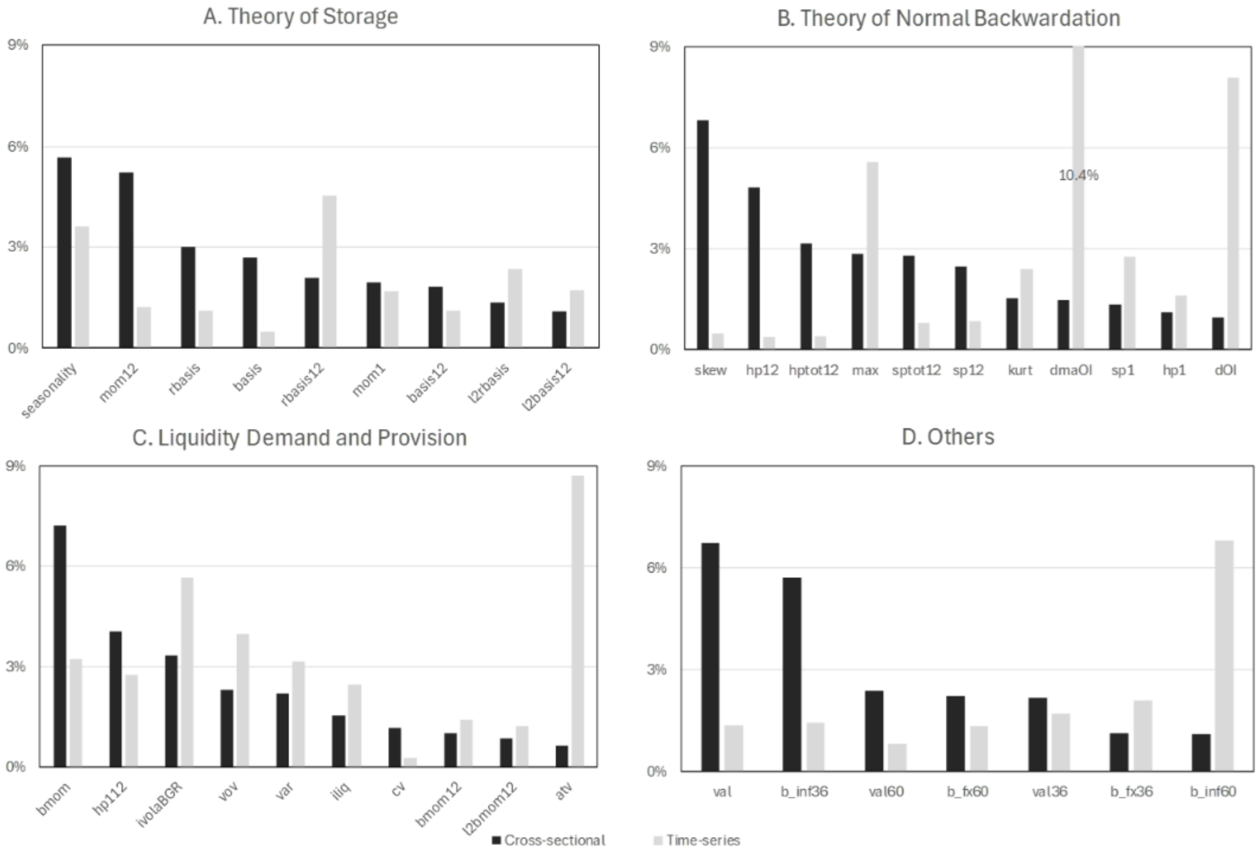
Cross-sectional long-short strategies

The cross section of commodities are sorted on model predicted returns. Long/short positions are entered in high-/low-ranked commodities.

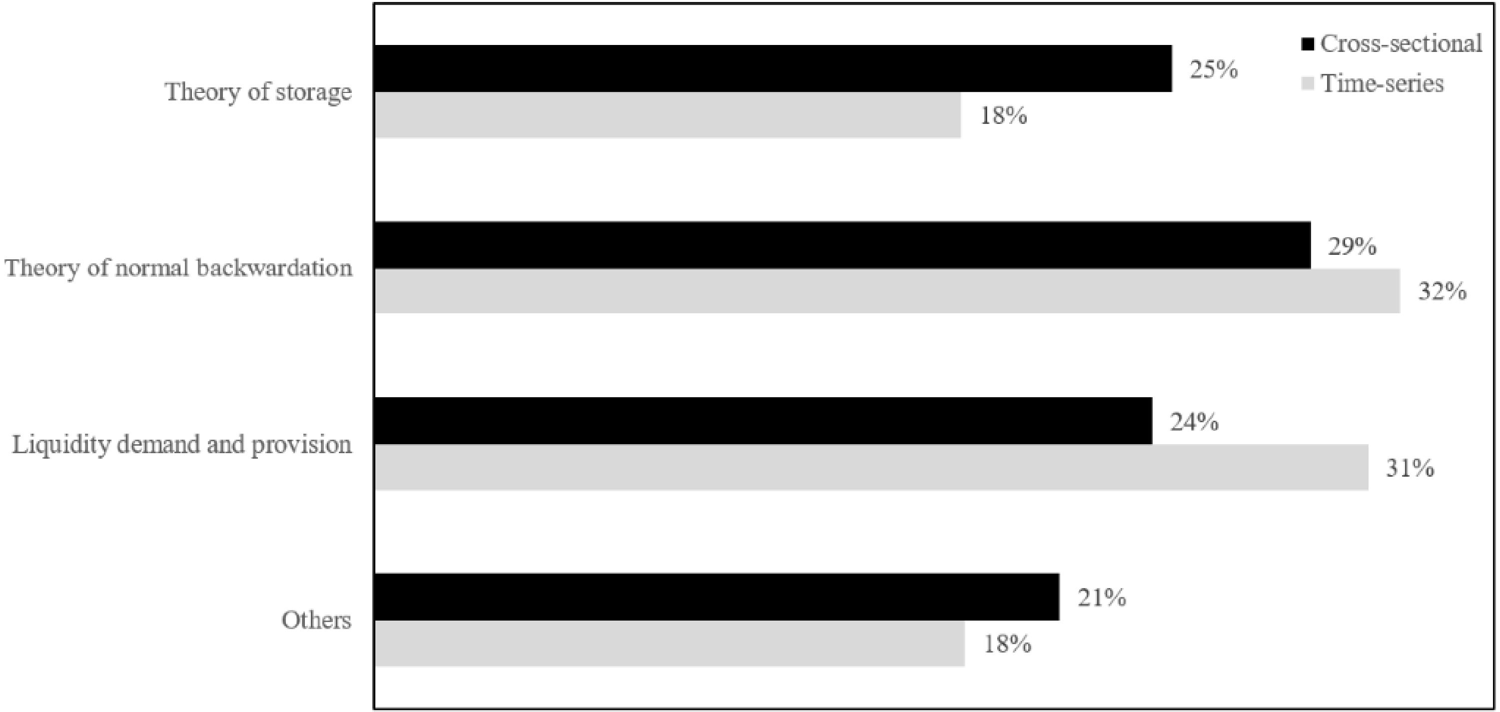
	Linear Ensemble				Nonlinear Ensemble			
	C	M	S	Forecast Combination	C	M	S	Forecast Combination
Annualized Mean	5.8%	-2.1%	0.9%	3.4%	1.7%	0.7%	-0.8%	1.1%
t-statistics (Newey-West)	3.2	-0.6	0.6	1.9	0.9	0.5	-0.2	0.6
Annualized S.D.	9.0%	9.9%	10.6%	9.4%	9.2%	8.8%	9.0%	10.0%
Annualized Sharpe	0.64	-0.21	0.08	0.36	0.18	0.08	-0.09	0.11
Annualized Downside S.D.	5.2%	6.2%	7.4%	6.4%	6.0%	5.1%	6.1%	7.1%
Annualized Sortino	1.10	-0.33	0.11	0.53	0.28	0.14	-0.13	0.15
Skewness	0.180	-0.015	-0.552	0.011	-0.166	0.118	-0.323	-0.279
Excess Kurtosis	0.952	0.076	4.179	1.779	1.754	-0.113	2.499	2.466
99% VaR (Cornish-Fisher)	-5.8%	-6.9%	-10.8%	-7.1%	-7.4%	-5.5%	-8.1%	-8.8%
Maximum Drawdown	-10.6%	-37.6%	-50.9%	-22.6%	-25.2%	-25.5%	-51.6%	-38.6%
Omega	1.49	0.84	0.96	1.24	1.19	1.04	1.04	1.24

Feature Importance

Variable importance by commodity theory



Variable importance by commodity theory

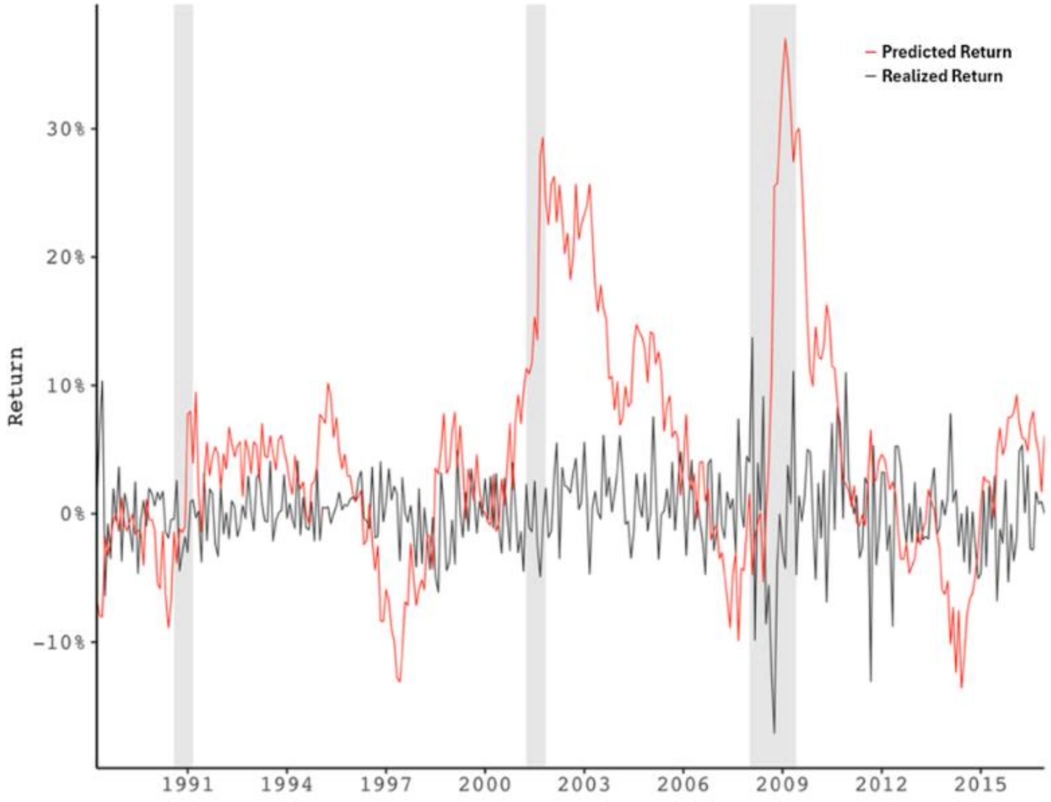


Business Cycle

Commodity returns & business cycles

		Avg. Realized Return	Avg. Forecasted Return				
	N	T=0	T=1	T=6	T=12	T=24	T=60
Expansion	311	3.87%	3.10%	2.41%	2.24%	1.79%	2.15%
Early Expansion	160	7.28%	3.79%	3.19%	2.78%	3.19%	2.06%
Late Expansion	151	0.27%	2.36%	1.58%	1.67%	0.30%	2.24%
Recession	34	-12.23%	-4.87%	-0.21%	2.14%	6.39%	6.55%
Early Recession	17	-5.93%	-4.37%	-1.85%	0.76%	1.33%	5.02%
Late Recession	17	-18.54%	-5.38%	1.44%	3.52%	11.45%	8.09%
Unconditional	345	2.29%	2.31%	2.15%	2.23%	2.24%	2.58%

Commodity returns & business cycles



Conclusion

- We document statistical and economic magnitude of return predictability in both the time-series and cross-section of commodity futures returns.
- All theories are important in explaining commodity futures risk premium.
- Moves the field towards a unified, more parsimonious theory of why commodities earn risk premia in both dimensions
- Business cycle analysis confirms the cyclical/counter-cyclical nature of realized/expected commodity returns.
- Practitioners should focus on a select set of predictors across theories, return dimension, and data category.
- Cautionary on model complexity.

THANK YOU

Q&A

APPENDICES

Spanning Test

	Basis	C	Momentum	C	Basis-momentum	C	Relative-basis	C	Skewness	C	Hedging pressure	C
C	0.319		0.272		0.265		0.105		0.497		0.572	
	3.0		2.1		2.7		0.9		8.6		8.9	
Basis		0.266										
		3.2										
Momentum				0.228								
				2.5								
Basis-momentum						0.263						
						2.4						
Relative-basis								0.101				
								1.0				
Skewness										0.516		
										8.0		
Hedging pressure												0.537
												8.4
Annualized	0.0	**6.0**	0.0	**6.0**	2.4	**4.8**	1.2	**6.0**	**4.8**	2.4	1.2	**3.6**
	-0.0	**3.0**	-0.1	**3.0**	1.2	**2.4**	0.5	**3.1**	**2.2**	1.3	0.9	**1.9**
R2	0.085		0.062		0.070		0.011		0.256		0.307	
N	203		203		203		203		203		203	

Pricing implications-12 pricing models

- Four benchmark pricing models from the literature
  - AVG + Basis; Szymanowska et al. (2014)
  - AVG + Basis + Momentum; Bakshi et al. (2019)
  - AVG + Basis + Basis-Momentum; Boons and Prado (2019)
  - AVG + Basis-Momentum + Relative Basis; Gu et al. (2024)
- Four augmented models with the addition of proposed zookeeper factor
- One single factor model of zookeeper and one two-factor model of zookeeper and AVG
- Kitchen sink
  - AVG + Basis + Momentum + Basis-Momentum + Relative Basis + Skewness + Hedging Pressure
- Kitchen sink + zookeeper
- Testing assets: 148 characteristic-sorted portfolios + 7 sector portfolios

GRS Test

	(1) Szymanowska et al (2014)	(2) Bakshi et al (2019)	(3) Boons & Prado (2019)	(4) Gu et al (2024)	(5) Augmented Szymanowska et al (2014)	(6) Augmented Bakshi et al (2019)	(7) Augmented Boons & Prado (2019)	(8) Augmented Gu et al (2024)	(9) C	(10) C + AVG	(11) Kitchen sink	(12) Augmented Kitchen sink
Avg|alpha|	6.24	6.19	6.15	6.17	5.98	6.00	5.90	5.85	6.38	5.97	5.99	5.97
Frac of alpha >0	52%	52%	51%	52%	50%	50%	47%	46%	68%	50%	47%	48%
Avg|t-stat|	2.74	2.79	2.72	2.69	2.61	2.68	2.61	2.57	1.47	2.57	2.75	2.74
Frac of *t* >1.96	17%	17%	15%	15%	14%	14%	14%	13%	8%	14%	17%	16%
N test assets	155	155	155	155	155	155	155	155	155	155	155	155
GRS	33.71	34.59	35.16	37.92	32.17	33.13	33.96	36.87	32.62	32.35	38.87	38.28
p -value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Fama-MacBeth Test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Szymanowska et al (2014)	Bakshi et al (2019)	Boons & Prado (2019)	Gu et al (2024)	Augmented Szymanowska et al (2014)	Augmented Bakshi et al (2019)	Augmented Boons & Prado (2019)	Augmented Gu et al (2024)	C	C + AVG	Kitchen sink	Augmented Kitchen sink
C					0.008**	0.007**	0.007**	0.012***	0.011***	0.011***		0.008***
					2.8	2.2	2.6	4.3	4.0	4.2		2.6
AVG	0.006	0.004	0.006	0.008*	0.004	0.004	0.005	0.005		0.004	0.004	0.004
	1.2	0.8	1.3	1.5	0.8	0.7	1.1	0.9		0.8	0.7	0.6
Basis	0.012***	0.010***	0.013***		0.011***	0.010***	0.012***				0.011***	0.010***
	4.2	3.2	4.3		3.3	2.9	3.6				2.9	2.8
Momentum		0.013***				0.013***					0.016***	0.015***
		4.2				4.1					4.2	4.3
Basis-momentum			0.004*	0.005**			0.004*	0.005**			0.004	0.004
			1.7	2.69			2.0	2.1			1.3	1.4
Relative-basis				0.013***				0.007*			0.008**	0.008*
				3.2				1.7			1.9	1.7
Skewness											0.005	0.005
											1.3	1.3
Hedging pressure											0.007**	0.007**
											2.3	2.3
Adj-R2	0.145	0.185	0.185	0.154	0.184	0.224	0.221	0.215	0.082	0.143	0.308	0.333
N	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760