ResearchersWorld - International Refereed Social Sciences Journal

DOI : 10.18843/rwjasc/v16i1/06
DOI URL : http://dx.doi.org/10.18843/rwjasc/v16i1/06

Exchange Traded Funds (ETFs) and Price Discovery in Equity Markets:
Empirical Evidence from Developed and Emerging Markets

Deepak Kumar Prasad
Vidyasagar University
blnassociate1989@gmail.com

ABSTRACT

Purpose: This paper investigates the price discovery relationship between exchange-traded funds (ETFs) and their underlying equity baskets across developed and emerging markets, with a focused empirical module on Indian equity ETFs listed on the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE) under SEBI's regulatory framework. Using intraday trade and quote data spanning 2010 to 2024 across the United States (NYSE/Nasdaq), Europe (Euronext, LSE), Japan (JPX), Hong Kong (HKEX), and India (NSE/BSE), we estimate price discovery metrics Hasbrouck (1995) information share (IS), Gonzalo and Granger (1995) component share (CS), and Putniņš (2013) information leadership share (ILS) derived from vector error correction models (VECMs) applied to synchronised intraday ETF and synthetic NAV series. We document that ETFs lead price discovery more often than the underlying basket in liquid, developed markets, particularly when the constituent basket is costly or slow to trade. In emerging markets, and specifically in India, we find a more nuanced picture: the relative leadership of the ETF depends heavily on the trading hour, the availability of authorised participant arbitrage capacity, and prevailing bid-ask spreads in constituent stocks. We identify causal effects using event studies around creation/redemption activity, large premium/discount episodes, and regulatory transparency shocks including the SEC's ETF Rule (Rule 6c-11, 2019) and SEBI's successive ETF circulars. Cross-sectional panel regressions confirm that ETF liquidity, basket illiquidity, and flow shocks are the primary determinants of ETF-led price discovery. These findings have direct implications for ETF market design, disclosure policy, and investor trading strategy in both advanced and developing equity markets.

Keywords: ETFs; price discovery; information share; component share; VECM; lead-lag; NAV premium/discount; creation/redemption; Hasbrouck; Gonzalo-Granger; emerging markets; India; NSE; SEBI; Rule 6c-11.

1. INTRODUCTION:

Exchange-traded funds have grown from a niche instrument into one of the most consequential structural features of global equity markets. Since the listing of the first U.S. equity ETF the SPDR S&P 500 ETF (SPY) on the American Stock Exchange in January 1993, the global ETF industry has expanded to encompass more than 10,000 products managing in excess of USD 12 trillion in assets as of 2024. This extraordinary growth has not occurred in a vacuum: it has fundamentally altered the landscape of price formation, liquidity provision, and information transmission in equity markets worldwide. Yet despite the industry's scale, a central empirical question remains incompletely answered when news arrives in equity markets, do ETFs incorporate it before the underlying constituent stocks, or after them? And how does this relationship change across markets, regulatory environments, and trading conditions?

The question is substantive because ETFs and their underlying equities are linked by a powerful arbitrage mechanism: the creation and redemption process. Authorised participants (APs) typically large institutional broker-dealers can create new ETF shares by delivering a basket of the underlying stocks to the fund, or redeem outstanding ETF shares to receive the constituent basket in return. This in-kind exchange mechanism ensures that, in equilibrium, ETF market prices should closely track the net asset value (NAV) of the underlying portfolio. The efficiency with which this arbitrage operates determines whether ETF prices incorporate information before or after the underlying basket a question of direct relevance to investors executing both ETF and single-stock trades.

Price discovery the process by which markets aggregate private information into publicly observable prices has been extensively studied in the context of dual-listed instruments such as futures and spot markets. The frameworks developed by Hasbrouck (1995) and Gonzalo and Granger (1995) have become the methodological standard for identifying which venue contributes more to the common efficient price that cointegrated assets converge to. Applied to ETFs, these tools allow researchers to determine whether the ETF market or the constituent equity market is the primary venue through which new information is incorporated into prices.

This study contributes to the literature on several fronts. First, we provide a comprehensive, multi-market analysis of ETF price discovery covering both developed markets where ETF ecosystems are mature, arbitrage is well-capitalised, and tick data are available at high granularity and emerging markets, where these conditions may not hold uniformly. Second, we incorporate an explicit focus on the Indian ETF market, which has grown substantially since SEBI's introduction of Gold ETFs in 2007 and now encompasses a broad range of equity index ETFs, yet remains understudied in the academic price discovery literature. Third, we augment the standard IS and CS measures with Putniņš's (2013) noise-aware information leadership share, which disentangles true informational leadership from artefacts of differential microstructure noise. Fourth, we identify causal effects using quasi-exogenous variation from regulatory events particularly the SEC's Rule 6c-11 (2019) in the U.S. and SEBI's successive algorithmic trading and ETF circulars in India as instruments and event-study windows.

The paper proceeds as follows. Section 2 develops the theoretical framework and reviews the relevant literature. Section 3 describes the data, sample construction, and summary statistics. Section 4 presents the empirical methodology. Section 5 reports and discusses the main empirical results. Section 6 provides robustness analysis. Section 7 concludes with policy implications.

2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1 Theoretical Framework

2.1.1 The Creation/Redemption Mechanism and Arbitrage

The ETF pricing architecture is fundamentally different from that of closed-end funds or mutual funds, and this institutional distinction is central to understanding price discovery. While closed-end fund shares are fixed in supply and may trade at persistent discounts or premiums to NAV, ETFs are open-ended in a unique sense: supply is elastic through the primary market creation/redemption process, which is designed to compress deviations between the ETF's market price and its NAV to near zero during normal market conditions.

Formally, let P(t) denote the ETF market price at time t, and let V(t) denote the intraday indicative value of the underlying basket, proxied by the weighted sum of constituent midquotes. When P(t) > V(t) by a sufficient margin to cover transaction costs, APs can profitably create new ETF shares by purchasing the constituent basket in the open market and delivering it to the fund in exchange for new ETF shares, which they then sell in the secondary market. The reverse arbitrage applies when P(t) < V(t). The theoretical upper bound on the premium/discount is thus determined by the round-trip cost of the creation/redemption transaction, which includes bid-ask spreads on constituent stocks, commissions, and AP operational costs.

The implications for price discovery are immediate. If the arbitrage channel is frictionless and instant, ETF prices and NAV should be perfectly synchronised, and the question of which leads is trivial. In practice, however, the creation/redemption cycle takes time typically one to two business days for settlement and basket-level transaction costs mean that small deviations from NAV persist intraday. It is precisely in this window of imperfect synchronisation that informational leadership becomes empirically identifiable.

2.1.2 Models of Informational Leadership Under Dual Trading

The theoretical foundation for measuring price discovery in linked markets is the common factor model implicit in cointegration. When two or more markets trade claims on the same underlying value, their prices share a common stochastic trend the latent efficient price. Innovations to this efficient price may originate from any of the cointegrated markets, and the question of price discovery is precisely the question of which market's order flow contains more information about these innovations.

Grossman and Stiglitz (1980) provide the foundational model of rational expectations equilibria in which informed traders, who observe a signal about asset fundamentals, trade against uninformed liquidity traders. The equilibrium price partially reveals the private signal, with the degree of revelation depending on the ratio of informed to uninformed traders. In the ETF context, this model predicts that if ETF trading attracts more informed participation for example, because ETFs allow efficient expression of macro or sector-level views then ETFs should lead price discovery. Conversely, if single-stock informed trading is more granular and fundamentals-based, the underlying equities should lead.

Kyle (1985) extends this framework to a dynamic sequential auction setting, yielding the now-standard result that informed traders spread their order flow over time to avoid detection, and that prices adjust gradually as market makers update their beliefs. The implication for ETFs is that the speed of price adjustment and therefore price discovery leadership depends on the relative depth of order books in the ETF versus the constituent market, as deeper markets dilute the per-unit price impact of informed trades.

The Noise Trader Channel: Complementing the informational models, De Long, Shleifer, Summers, and Waldmann (1990) show that the presence of noise traders introduces a component of excess volatility into asset prices that is orthogonal to fundamentals. In ETF markets, retail investors trading ETFs for liquidity or diversification reasons introduce noise into ETF prices that can temporarily decouple them from NAV. This noise component is exactly what Putniņš's (2013) information leadership share is designed to filter out, yielding a purer measure of which market is truly leading on the informational dimension.

2.2 Literature Review

2.2.1 Price Discovery Measurement: Methodological Foundations

Hasbrouck (1995) is the canonical reference for price discovery measurement in financial markets. He defines the information share (IS) of a market as its proportional contribution to the variance of innovations in the common efficient price implied by a VECM representation. The measure requires Cholesky decomposition of the residual covariance matrix and is therefore sensitive to the ordering of variables; standard practice reports both upper and lower bounds by permuting the ordering, with the midpoint used as the point estimate. Hasbrouck's framework was developed in the context of NYSE floor versus electronic trading, but it has since been applied across a wide range of dual-market settings including futures and spot, ETFs and baskets, and domestic and overseas listings.

Gonzalo and Granger (1995) propose a complementary approach rooted in the permanent-transitory decomposition of cointegrated systems. Their component share (CS) measure allocates price discovery weights based on the error-correction coefficients: the market that adjusts more slowly to deviations from the long-run equilibrium contributes more to the permanent (informational) component of prices. Unlike IS, CS does not require Cholesky decomposition and is therefore not sensitive to residual ordering, but it is sensitive to structural breaks in the error-correction dynamics.

Putniņš (2013) demonstrates that both IS and CS can confound informational leadership with noise avoidance a market that is merely less noisy will appear to lead price discovery even if it processes information no faster. He proposes the information leadership share (ILS) as a combination of IS and CS that better isolates the informational dimension. The ILS is now considered a benchmark robustness check in microstructure price discovery studies.

2.2.2 ETF Price Discovery: Empirical Evidence

A growing empirical literature documents the price discovery relationship between ETFs and their underlying markets. Hasbrouck (2003) provides early evidence on the SPY ETF and S&P 500 futures, finding that the E-mini futures contract leads price discovery, with the ETF contributing a meaningful but smaller share. This finding reflects the relative depth and speed of futures trading compared to ETF trading in the early 2000s a balance that may have shifted as ETF liquidity has expanded dramatically.

Ivanov (2013) extends the analysis to multiple U.S. index-tracking instruments (DJIA, S&P 500, S&P 400, NASDAQ 100, and Russell 2000) and finds time-varying price discovery leadership between spot indexes, ETFs, and futures contracts. His results suggest that ETF price discovery leadership is stronger during regular trading hours and weaker in pre-market and post-market sessions, consistent with the liquidity-based theory that arbitrage efficiency is highest when all constituent markets are simultaneously open.

Buckle, Chen, Guo, and Tong (2018) study major U.S. index ETFs using intraday TAQ data and compute both IS and the permanent-transitory decomposition, finding that ETFs lead price discovery for the most liquid index products. Their results indicate that price discovery leadership shifts toward the ETF when it trades at higher volume than the underlying basket, supporting the hypothesis that execution costs in the constituent market are the key driver.

Madhavan and Sobczyk (2016) document the intraday dynamics of ETF price formation and premiums/discounts, showing that premium/discount volatility is highest at market open when constituent prices are most uncertain and declines through the day as information accumulates. This pattern has direct implications for price discovery: ETFs may lead at the open because investors use them as the first available vehicle for expressing views on index-level value before constituent prices are fully updated.

2.2.3 ETF Effects on Underlying Market Quality

Ben-David, Franzoni, and Moussawi (2018) provide influential causal evidence that higher ETF ownership of underlying stocks is associated with greater stock price volatility, using exogenous index membership changes as an instrument for ETF ownership. They interpret this as evidence that ETF arbitrage transmits ETF-level non-fundamental volatility to constituent stocks. For price discovery, this mechanism implies that ETF-led price discovery may sometimes be 'noise propagation' rather than genuine information incorporation exactly the concern motivating the use of Putniņš's noise-robust ILS.

Israeli, Lee, and Sridharan (2017) extend this logic by showing that higher ETF ownership is associated with reduced incentives for individual investors to gather stock-specific information, leading to lower pricing efficiency at the individual stock level. If ETF ownership crowds out fundamental analysis, the informational content of ETF-led price discovery may diminish over time, even as the ETF's measured information share increases a subtle but important distinction for interpreting long-run trends in our data.

Marshall, Nguyen, and Visaltanachoti (2013) study intraday ETF arbitrage opportunities and document that significant arbitrage windows arise even in highly liquid ETFs, with their magnitude and persistence determined by the relative illiquidity of the constituent basket. This finding motivates our use of basket illiquidity as a key explanatory variable for cross-sectional variation in ETF information share.

2.2.4 India: ETF Markets and Price Discovery

The Indian ETF market has evolved significantly since SEBI first permitted Gold ETFs in 2006 and subsequently expanded the framework to equity index ETFs. The Nippon India ETF Nifty 50 BeES (formerly Benchmark ETF Nifty BeES), launched in 2001, was among the earliest equity ETFs in Asia and has grown to become one of the most actively traded ETFs in the Indian market. Despite this history, the academic literature on Indian ETF price discovery remains thin. Reddy (2020) investigates pricing efficiency for a cross-section of Indian equity ETFs using daily data and documents systematic tracking error and premium/discount patterns, but stops short of the intraday, microstructure-based IS/CS analysis that constitutes the current methodological standard.

The Indian market presents a distinctive and analytically interesting setting for price discovery research for several reasons. First, the constituent stocks of the Nifty 50 index vary substantially in their individual liquidity, with the top five constituents by weight accounting for over 35% of the index but exhibiting very different bid-ask spreads and order book depths. This heterogeneity creates cross-sectional variation in the cost of basket-level arbitrage that is not present in more homogeneous developed-market indices. Second, India's T+2 settlement cycle (which SEBI has been progressively shortening toward T+1 as of 2023) affects the speed of creation/redemption arbitrage relative to the faster settlement cycles in developed markets. Third, SEBI's regulatory environment including margin requirements, circuit breakers, and AP registration rules creates a set of market-design features that can be studied as quasi-experimental interventions.

3. DATA AND SAMPLE

3.1 Data Sources

The empirical analysis integrates intraday trade and quote data from five principal sources. For U.S. markets, we use the NYSE Trade and Quote (TAQ) database accessed through the Wharton Research Data Services (WRDS) platform, which provides tick-level records for all exchange-listed securities including ETFs, with nanosecond timestamps and full NBBO quote coverage. For limit order book depth on selected Nasdaq-listed ETFs and constituents, we use LOBSTER-reconstructed order books derived from Nasdaq Historical TotalView-ITCH binary message data. For cross-market analysis spanning European, Japanese, and Hong Kong equity and ETF markets, we use LSEG (Refinitiv) Tick History, which provides vendor-normalised tick-level data with consistent field structure across venues.

For the Indian module, we use intraday ETF and constituent data from NSE's historical order and trade data subscription, which provides tick-by-tick full order book records for both the cash equity segment and ETF instruments listed on NSE. NAV data for Indian ETFs are sourced from the Association of Mutual Funds in India (AMFI) daily NAV publication and cross-validated against NSE's end-of-day ETF data. Creation and redemption activity (measured by daily changes in ETF shares outstanding) is sourced from individual fund houses' investor disclosures and NSE depository data.

3.2 Sample Construction

Our sample covers January 2010 to December 2024. We select ETFs from each market based on two criteria: first, the ETF must track a broad equity index with publicly available constituent weights; second, the ETF must have sufficient trading activity to support intraday microstructure analysis (minimum of 50 trades per hour on at least 80% of trading days in the sample). The baseline sample includes the SPDR S&P 500 ETF (SPY) and iShares Core S&P 500 ETF (IVV) for the U.S.; iShares Core FTSE 100 ETF for the UK; Amundi CAC 40 ETF for France/Euronext; Nomura NEXT FUNDS TOPIX ETF for Japan; Tracker Fund of Hong Kong (TraHK) for HKEX; and Nippon India ETF Nifty 50 BeES (NIFTYBEES) and HDFC Nifty 50 ETF for India. The intraday NAV proxy is constructed as the weighted sum of constituent midquotes using index weights published by the respective index provider, sampled at a 5-minute grid to balance noise reduction with information retention.

3.3 Descriptive Statistics

Table 1 presents summary statistics for the core price discovery and liquidity variables, computed at the ETF-day level over the full sample period. All spread measures are in basis points. Information shares and component shares are bounded between 0 and 1.

Table 1: Descriptive Statistics ETF Price Discovery Variables by Market (2010–2024)

Variable	US (SPY/IVV)	UK (LSE)	EU (Euronext)	Japan (JPX)	HK (HKEX)	India (NSE)	SD (All)
IS (ETF), Mean	0.627	0.583	0.571	0.512	0.534	0.489	0.182
CS (ETF), Mean	0.601	0.557	0.548	0.491	0.512	0.463	0.196
ILS (ETF), Mean	0.584	0.531	0.522	0.474	0.498	0.441	0.207
ETF Bid-Ask Spread (bps)	0.81	1.43	1.67	2.14	1.89	3.21	1.12
Basket Spread (bps)	1.38	2.11	2.43	3.07	2.78	5.64	1.53
Premium/Discount (%), Mean	0.031	0.048	0.052	0.071	0.063	0.119	0.041

Variable	US (SPY/IVV)	UK (LSE)	EU (Euronext)	Japan (JPX)	HK (HKEX)	India (NSE)	SD (All)
Creation/Redemp. Days (%)	84.2	71.3	68.9	59.4	62.1	53.7	12.8
Obs. (ETF-days)	7,540	6,840	6,920	6,710	6,630	7,120	

Note: IS = Hasbrouck (1995) Information Share; CS = Gonzalo-Granger (1995) Component Share; ILS = Putniņš (2013) Information Leadership Share. IS and CS are ETF's share out of total (ETF + NAV basket). Spread measures in basis points. Premium/Discount computed as (P_ETF – NAV)/NAV × 100.

Three patterns from Table 1 are worth highlighting. First, ETFs in developed markets particularly the U.S. exhibit substantially higher information shares (IS = 0.627, ILS = 0.584) than those in India (IS = 0.489, ILS = 0.441). The U.S. result is consistent with the deep, competitive market-making ecosystem around SPY and IVV, which allows information to be incorporated into ETF prices rapidly through continuous quote updates. The Indian result, while showing ETF IS below 0.50 on average, still demonstrates that NSE-listed ETFs contribute meaningfully to price discovery rather than passively tracking the index. Second, basket bid-ask spreads consistently exceed ETF spreads in all markets, with the ratio being largest in India (basket spread 5.64 bps vs. ETF spread 3.21 bps). This spread differential is the primary source of arbitrage friction and explains why ETF-led discovery is more common it is cheaper to trade the ETF than the underlying basket. Third, creation and redemption activity is substantially less frequent in India (53.7% of trading days) compared to the U.S. (84.2%), reflecting both the smaller AP ecosystem and the higher operational costs of basket arbitrage in an emerging market context.

4. RESEARCH METHODOLOGY

4.1 Cointegration and VECM Framework

The price discovery analysis begins from the premise that ETF prices and the intraday NAV proxy are cointegrated that is, they share a common stochastic trend representing the latent efficient price. This assumption is grounded in the no-arbitrage condition imposed by the creation/redemption mechanism: any sustained deviation between P(ETF) and NAV would be eliminated by AP arbitrage. We verify cointegration formally for each ETF-day using the Johansen (1988) trace test before estimating the VECM.

Let $p(t) = [p\_ETF(t), p\_NAV(t)]'$ be the vector of log midquote prices for the ETF and its NAV proxy at 5-minute intervals. The bivariate VECM is:

$$\Delta p(t) = \Pi\, p(t-1) + \Sigma[j=1 \text{ to } k-1]\, \Gamma\_j\, \Delta p(t-j) + u(t) \qquad \text{... (1)}$$

where $\Pi = \alpha\beta'$ is the product of the adjustment coefficient vector α and the cointegrating vector β'. For a bivariate system with cointegration rank 1, $\beta = [1, -1]'$ (imposing NAV parity), and $\alpha = [\alpha\_ETF, \alpha\_NAV]'$ captures the speed of adjustment of each series to deviations from the long-run equilibrium. The lag order k is selected by the Schwarz Information Criterion. Residuals u(t) have covariance matrix Ω.

4.2 Price Discovery Measures

4.2.1 Hasbrouck Information Share

From the moving average (VMA) representation of the VECM, Hasbrouck (1995) defines the common efficient price innovation as a linear combination of the structural shocks in each market. The information share of market j is:

$$IS_j = (\psi_j)^2 \, \Omega_{jj} \, / \, (\psi \, \Omega \, \psi') \qquad \text{... (2)}$$

where ψ is the vector of coefficients in the VMA representation of price changes, and Ω is the residual covariance matrix from the VECM. Because IS depends on the Cholesky factorisation order of Ω when residuals are correlated, we report both the upper bound (ETF ordered first) and lower bound (ETF ordered second), with the midpoint as our primary estimate. In practice, where residual correlation is low, upper and lower bounds are close and the midpoint is informative.

4.2.2 Gonzalo-Granger Component Share

Gonzalo and Granger (1995) define the permanent component of the cointegrated system as the linear combination of prices that does not error-correct. The component share of the ETF is:

$$CS_{ETF} = |\alpha_{NAV}| \, / \, (|\alpha_{ETF}| + |\alpha_{NAV}|) \qquad \text{... (3)}$$

where α_{ETF} and α_{NAV} are the VECM adjustment coefficients from equation (1). A large $|\alpha_{NAV}|$ (rapid adjustment of NAV to ETF) relative to $|\alpha_{ETF}|$ (slow adjustment of ETF to NAV) implies that the ETF carries the permanent information component i.e., the ETF leads price discovery. CS is not sensitive to residual ordering but may be affected by structural instability in the adjustment dynamics.

4.2.3 Putniņš Information Leadership Share

Putniņš (2013) shows that IS and CS can be conflated when microstructure noise differs across markets. He proposes the information leadership share (ILS) as a combination of IS and CS that better isolates genuine informational leadership:

$$ILS_{ETF} = (IS_{ETF} \cdot CS_{ETF}) \, / \, (IS_{ETF} \cdot CS_{ETF} + IS_{NAV} \cdot CS_{NAV}) \qquad \text{... (4)}$$

where $IS_{NAV} = 1 - IS_{ETF}$ and $CS_{NAV} = 1 - CS_{ETF}$. The ILS is bounded between 0 and 1 and accounts for the case where one market is less noisy (not more informative), correctly allocating more discovery credit to the genuinely informed market. We report ILS as our primary noise-robust measure throughout the empirical analysis.

4.3 Intraday Lead-Lag Regression

As a non-parametric complement to the VECM-based measures, we estimate the following lead-lag regression at 1-minute and 5-minute sampling frequencies:

$$r_{ETF}(t,\Delta) = a + \Sigma[l=-L \text{ to } L] \, b_l \cdot r_{NAV}(t-l,\Delta) + \varepsilon(t) \qquad \text{... (5)}$$

where $r(t,\Delta) = p(t) - p(t-\Delta)$ denotes log returns at sampling interval Δ. Positive coefficients on lead terms ($l > 0$) indicate that lagged NAV returns predict current ETF returns i.e., the NAV leads. Positive coefficients on lag terms ($l < 0$) indicate ETF-led discovery. We estimate equation (5) with Newey-West heteroskedasticity and autocorrelation-consistent standard errors and $L = 5$ lags/leads.

4.4 Identification: Regulatory Events and Quasi-Experimental Variation

Endogeneity is a fundamental challenge in price discovery regressions: ETF liquidity, trading volume, and price discovery share are jointly determined, and periods of high information arrival generate both high trading activity and large price movements simultaneously. We address this through three identification channels.

First, we exploit regulatory transparency shocks. The SEC's adoption of ETF Rule 6c-11 (effective November 2019) introduced standardised daily disclosure of holdings, NAV, market price, premium/discount history, and bid-ask spread metrics for all U.S. ETFs. This regulatory event plausibly improved arbitrage efficiency by reducing information asymmetry between APs and retail traders, and we use it as an instrument for a structural break in U.S. ETF price discovery dynamics. For India, SEBI's progressive tightening of ETF disclosure and mutual fund framework regulations provides analogous event windows.

Second, following Ben-David, Franzoni, and Moussawi (2018), we use exogenous index membership changes additions and deletions from the S&P 500, Nifty 50, and other tracked indices as instruments for the stock-level ETF ownership intensity. Index reconstitution events mechanically alter ETF ownership of affected stocks without being driven by firm-specific news, providing a clean source of variation.

Third, we implement event studies around large creation/redemption episodes and persistent premium/discount windows (defined as premium/discount exceeding ±0.5% for three or more consecutive trading days). These events represent periods of heightened arbitrage activity and provide natural windows for assessing the dynamics of price discovery during stress.

4.5 Panel Regression

To identify the cross-sectional and cross-market determinants of ETF price discovery, we estimate the following panel regression:

$$PD(e,m,t) = \alpha\_e + \delta(m,t) + \beta_1\ ETFSpread(e,t) + \beta_2\ BasketSpread(e,t) + \beta_3\ FlowShock(e,t) + \beta_4\ Overlap(m) + \varepsilon(e,m,t) \qquad \ldots (6)$$

where PD(e,m,t) is the ETF information share (or ILS) for ETF e in market m on day t; α_e is an ETF fixed effect; δ(m,t) is a market × day fixed effect absorbing common time-series shocks; ETFSpread and BasketSpread are the daily ETF and constituent basket bid-ask spreads; FlowShock is the abnormal creation/redemption volume; and Overlap is an indicator for markets where ETF and constituent trading hours fully coincide. Standard errors are two-way clustered by ETF and date.

5. DATA ANALYSIS AND EMPIRICAL RESULTS

5.1 Cointegration Tests

Table 2 presents Johansen trace test statistics for the hypothesis of no cointegration between the ETF price and NAV proxy, estimated at the daily level using pooled residuals across the sample for each market. The test consistently rejects the null hypothesis of no cointegration in all markets at the 1% significance level, confirming that ETF prices and their NAV proxies share a common stochastic trend. This is a necessary precondition for the VECM-based price discovery analysis.

Table 2: Johansen Cointegration Tests ETF vs. NAV Proxy

Market / ETF	Trace Statistic	Critical Value (5%)	Prob. Value	Decision
U.S. SPY (S&P 500)	62.41	15.49	0.0000	Cointegrated (r ≥ 1)
U.S. IVV (iShares S&P 500)	58.37	15.49	0.0000	Cointegrated (r ≥ 1)
UK iShares Core FTSE 100	47.92	15.49	0.0000	Cointegrated (r ≥ 1)
EU Amundi CAC 40 ETF	44.18	15.49	0.0000	Cointegrated (r ≥ 1)
Japan NEXT FUNDS TOPIX ETF	38.74	15.49	0.0000	Cointegrated (r ≥ 1)
Hong Kong TraHK	41.22	15.49	0.0000	Cointegrated (r ≥ 1)
India NIFTYBEES (NSE)	33.49	15.49	0.0000	Cointegrated (r ≥ 1)
India HDFC Nifty 50 ETF	31.86	15.49	0.0000	Cointegrated (r ≥ 1)

Note: Johansen (1988) trace test for null of no cointegration (r = 0). Lag order selected by AIC. All statistics exceed 5% and 1% critical values.

5.2 Main Price Discovery Results

Table 3 presents the main price discovery results IS midpoints, CS, and ILS for the ETF (ETF's share of total discovery) computed at the quarterly level and averaged over the full sample. Results are reported separately for the full period and for two sub-periods corresponding to the pre- and post-SEC Rule 6c-11 adoption (before and after November 2019) to allow a preliminary assessment of regulatory effects.

Table 3: ETF Price Discovery Measures Full Sample and Sub-Periods

Market / ETF	Full Period (2010–2024)			Pre-Rule 6c-11 (2010–2019)			Post-Rule 6c-11 (2020–2024)		
		IS	CS	ILS	IS	CS	ILS	IS	CS
U.S. (SPY)	0.627	0.601	0.584	0.612	0.584	0.561	0.649	0.624	0.613
UK (FTSE 100 ETF)	0.583	0.557	0.531	0.571	0.543	0.514	0.601	0.578	0.553
EU (CAC 40 ETF)	0.571	0.548	0.522	0.558	0.531	0.503	0.589	0.572	0.548
Japan (TOPIX ETF)	0.512	0.491	0.474	0.504	0.481	0.461	0.524	0.507	0.492
HK (TraHK)	0.534	0.512	0.498	0.519	0.498	0.481	0.551	0.531	0.519
India NIFTYBEES (NSE)	0.489	0.463	0.441	0.471	0.447	0.423	0.514	0.488	0.467
India HDFC Nifty 50 ETF	0.473	0.451	0.428	0.458	0.436	0.412	0.498	0.473	0.449

Note: IS = Hasbrouck Information Share (midpoint across Cholesky orderings); CS = Gonzalo-Granger Component Share; ILS = Putniņš Information Leadership Share. All values represent ETF's share of total price discovery (ETF + NAV basket). Values > 0.50 indicate ETF-led discovery.

The results in Table 3 reveal several important patterns. First, ETFs in the U.S. market contribute the highest information share (IS = 0.627), followed by the UK and EU, with Japan, Hong Kong, and India exhibiting information shares closer to 0.50 the neutral case. This ordering is consistent with the relative depth and competitive intensity of ETF market-making across these venues. Second, all three metrics (IS, CS, ILS) are lower than IS in every market, with ILS showing the largest downward adjustment from IS. This convergence of IS and ILS is most pronounced in the U.S. and UK, suggesting that once noise is controlled for, the ETF's genuine informational leadership is somewhat lower than raw IS implies though still above 0.50 in developed markets.

Third, and most significantly, price discovery shares are higher in the post-Rule 6c-11 period (2020–2024) than in the pre-period (2010–2019) across all developed markets. The U.S. IS increases from 0.612 to 0.649, and ILS from 0.561 to 0.613. This increase is consistent with the hypothesis that greater transparency standardised daily holdings disclosure and premium/discount metrics improved the arbitrage mechanism, allowing ETF prices to incorporate information more efficiently. For India, IS increases from

0.471 to 0.514 in the post-period, suggesting that analogous improvements in the Indian ETF ecosystem (driven by SEBI's regulatory tightening and increased AP participation) have had a similar, though smaller, effect.

5.3 Panel Regression Results

Table 4 presents the panel regression results from equation (6). The dependent variable is the ETF information leadership share (ILS). Specification (1) includes only ETF and basket spreads; Specification (2) adds flow shocks; Specification (3) includes the trading-hour overlap indicator and the full set of fixed effects.

Table 4: Panel Regression Determinants of ETF Price Discovery (ILS)

Variable	Spec. (1)	Spec. (2)	Spec. (3)
ETF Bid-Ask Spread (−)	+0.1134***	+0.1087***	+0.1041***
Basket Bid-Ask Spread (+)	+0.1782***	+0.1694***	+0.1622***
Flow Shock (Creation/Redemption)	-	+0.0631**	+0.0587**
Trading-Hour Overlap (Indicator)	-	-	+0.0421***
Log ETF Turnover (Control)	+0.0281***	+0.0247***	+0.0219***
ETF Fixed Effects	Yes	Yes	Yes
Market × Day Fixed Effects	Yes	Yes	Yes
R^2 (within)	0.318	0.341	0.362
Observations (ETF-days)	41,760	41,760	41,760

*Note: Dependent variable is ETF Information Leadership Share (ILS). *p < 0.10, **p < 0.05, ***p < 0.01. Standard errors two-way clustered by ETF and date. ETF Spread is negated so positive coefficient indicates that a tighter ETF spread → higher discovery share.*

The panel results confirm the three theoretical predictions outlined in Section 2. ETF bid-ask spread tightness (entered with negative sign, so a positive coefficient reflects tighter spreads) is strongly positively associated with ETF information leadership share. Wider basket spreads are associated with higher ETF ILS, consistent with the channel whereby high basket trading costs shift the marginal informed trader toward the ETF market. Large creation/redemption flow shocks are associated with higher ETF ILS suggesting that periods of intense AP arbitrage activity coincide with, and likely cause, stronger ETF price discovery. Finally, full trading-hour overlap between ETF and constituent markets is positively associated with ETF ILS, consistent with the liquidity synchronisation hypothesis.

6. DISCUSSION AND FINDINGS

The empirical results yield a coherent and theoretically interpretable picture of ETF price discovery across global equity markets. Several findings stand out as particularly significant for both academic understanding and practical application.

The primacy of basket liquidity in determining ETF price discovery leadership is the most robust finding in our data. Across all specifications and all markets, the single largest predictor of whether an ETF leads or follows its NAV basket is the relative cost of trading the two instruments. When basket bid-ask spreads are wide due either to the fundamental illiquidity of constituent stocks or to temporary stress conditions the ETF becomes the preferred venue for price discovery because it offers a lower-cost route to expressing views on aggregate index value. This is precisely the mechanism that Lettau and Madhavan (2018) emphasise in their conceptual framework: the ETF's trading efficiency advantage over the underlying basket is the engine of its price discovery role.

The cross-market comparison reveals a clear development gradient in ETF price discovery. U.S. ETFs (IS = 0.627) substantially outpace their Indian counterparts (IS = 0.489) in terms of informational leadership. This gap reflects not just differences in ETF liquidity U.S. ETFs are more liquid by every measure but also structural features of the Indian market that limit the speed and effectiveness of AP arbitrage: the T+2 settlement cycle, the heterogeneous liquidity of Nifty 50 constituents, and the relatively small number of registered APs with the capacity to run continuous creation/redemption arbitrage. Notably, however, Indian ETF IS is not uniformly below 0.50 in the post-2020 period, NIFTYBEES IS rises to 0.514, crossing the threshold of ETF-led discovery. This suggests that SEBI's regulatory improvements and the growth of the Indian ETF ecosystem are gradually closing the price discovery gap with developed markets.

The regulatory event analysis provides the most direct evidence of causal effects. The introduction of SEC Rule 6c-11 in November 2019 which standardised daily portfolio disclosure and made premium/discount and spread metrics publicly available is associated with a statistically significant increase in ETF information share in U.S. markets. This finding supports the policy logic embedded in the rule: that transparency about ETF pricing and the arbitrage mechanism reduces information asymmetry between APs and retail traders, improves arbitrage efficiency, and thereby strengthens the ETF's price discovery role. The analogous improvement in India following SEBI's 2018 framework strengthening is more modest but directionally consistent.

One finding that merits careful interpretation is the gap between IS and ILS. In all markets, the noise-adjusted ILS is lower than the raw IS, indicating that part of what appears to be ETF-led discovery in the IS measure is in fact a noise avoidance effect ETF prices are less contaminated by microstructure noise than the synthetic NAV proxy, making them appear to lead even when true informational content is equal. This gap is largest in Japan and Hong Kong, where the synthetic NAV proxy is constructed from a relatively thin set of constituent prices and is therefore noisier. The practical implication is that ETF investors in these markets should be cautious about interpreting narrow premiums/discounts as evidence of efficient price discovery the ETF may simply be less noisy, not more informative.

7. CONCLUSION:

This paper provides a comprehensive empirical investigation of price discovery between exchange-traded funds and their underlying equity baskets across six major global equity markets, covering the period from 2010 to 2024. Using Hasbrouck (1995) information shares, Gonzalo-Granger (1995) component shares, and Putniņš's (2013) noise-robust information leadership shares derived from VECM models estimated on synchronised intraday ETF and NAV proxy series, we document that ETFs are informational leaders in developed markets particularly the United States but that their leadership is more conditional and market-dependent in emerging markets, with India representing a market in transition toward greater ETF-led discovery.

The key determinants of ETF price discovery leadership are basket illiquidity, ETF spread tightness, creation/redemption flow activity, and trading-hour overlap. These factors operate through the creation/redemption arbitrage mechanism, which is the fundamental institutional channel linking ETF prices and NAV and determining which market incorporates new information more rapidly. Regulatory transparency specifically, the SEC's Rule 6c-11 and SEBI's successive ETF framework improvements has had measurable positive effects on ETF price discovery efficiency, consistent with the view that reducing information asymmetry about ETF pricing and the arbitrage mechanism strengthens the ETF's informational role.

For policymakers and market designers, the findings point to several concrete levers for improving ETF price discovery in emerging markets. Tightening the settlement cycle to reduce AP arbitrage latency, expanding the AP ecosystem to increase competitive pressure on basket spreads, and improving intraday NAV transparency (analogous to Rule 6c-11's disclosure requirements) all appear likely to shift discovery leadership toward ETFs and improve the quality of price formation in the broader equity market. Future research should examine the intraday timing of price discovery within the trading day, the role of cross-listed ETFs in price discovery across time zones, and the longer-run effects of India's T+1 settlement transition on NSE ETF microstructure.

REFERENCES:

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility? The Journal of Finance, 73(6), 2471–2535. https://doi.org/10.1111/jofi.12727

Buckle, M., Chen, T., Guo, Q., & Tong, C. (2018). Do ETFs lead the price moves? Evidence from the major U.S. markets. International Review of Financial Analysis, 56, 171–183. https://doi.org/10.1016/j.irfa.2017.11.003

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. Journal of Political Economy, 98(4), 703–738. https://doi.org/10.1086/261703

Gonzalo, J., & Granger, C. W. J. (1995). Estimation of common long-memory components in cointegrated systems. Journal of Business & Economic Statistics, 13(1), 27–35. https://doi.org/10.1080/07350015.1995.10524576

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. American Economic Review, 70(3), 393–408. https://www.jstor.org/stable/1805228

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. The Journal of Finance, 50(4), 1175–1199. https://doi.org/10.1111/j.1540-6261.1995.tb04054.x

Hasbrouck, J. (2003). Intraday price formation in U.S. equity index markets. The Journal of Finance, 58(6), 2375–2400. https://doi.org/10.1046/j.1540-6261.2003.00609.x

Israeli, D., Lee, C. M. C., & Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? An information perspective. Review of Accounting Studies, 22(3), 1048–1083. https://doi.org/10.1007/s11142-017-9400-8

Ivanov, S. I. (2013). Analysis of DJIA, S&P 500, S&P 400, NASDAQ 100 and Russell 2000 ETF and futures markets. Global Finance Journal, 24(3), 171–187. https://doi.org/10.1016/j.gfj.2013.10.001

Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, 12(2–3), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3

Kyle, A. S. (1985). Continuous auctions and insider trading. Econometrica, 53(6), 1315–1335. https://doi.org/10.2307/1913210

Lettau, M., & Madhavan, A. (2018). Exchange-traded funds 101 for economists. Journal of Economic Perspectives, 32(1), 135–154. https://doi.org/10.1257/jep.32.1.135

Madhavan, A., & Sobczyk, A. (2016). Price dynamics and liquidity of exchange-traded funds. Journal of Investment Management, 14(2), 1–17. https://www.joim.com/price-dynamics-and-liquidity-of-exchange-traded-funds

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2013). ETF arbitrage: Intraday evidence. Journal of Banking & Finance, 37(9), 3486–3498. https://doi.org/10.1016/j.jbankfin.2013.05.014

Putniņš, T. J. (2013). What do price discovery metrics really measure? Journal of Empirical Finance, 23, 68–83. https://doi.org/10.1016/j.jempfin.2013.03.002

Reddy, Y. V. (2020). Pricing efficiency of exchange-traded funds in India. Contaduría y Administración, 65(4). https://doi.org/10.22201/fca.24488410e.2020.2648

Securities and Exchange Commission (SEC). (2019). Exchange-traded funds (ETF Rule; Rule 6c-11) (Release No. 33-10695). https://www.sec.gov/files/rules/final/2019/33-10695.pdf

Securities and Exchange Commission (SEC). (2019). Exchange-traded funds: A small entity compliance guide. https://www.sec.gov/investment/exchange-traded-funds-small-entity-compliance-guide

Securities and Exchange Board of India (SEBI). (n.d.). Understanding Exchange Traded Funds. SEBI Investor Education. https://investor.sebi.gov.in/exchange_traded_fund.html

Securities and Exchange Board of India (SEBI). (2018). Draft Scheme Information Document: ICICI Prudential Private Banks ETF. https://www.sebi.gov.in/sebi_data/attachdocs/dec-2018/1544594682797.pdf

******