# Constructing and Statistical Validation of China's Implied Volatility Index

October 2025

## 1 Computation of VIX

### 1.1 Core Logic

- Set the target maturity $T^* = 30$ days.
- Identify two option maturities:
  - $T_1$: the closest maturity *less than* 30 days
  - $T_2$: the closest maturity *greater than* 30 days
- Compute the variance for each maturity, then linearly interpolate to obtain the 30-day variance.

### 1.2 Variance Computation Formula

$$\sigma^2 = \frac{2e^{rT}}{T} \sum_i \frac{\Delta K_i}{K_i^2} Q(K_i) - \frac{1}{T} \left( \frac{F}{K_0} - 1 \right)^2 \tag{1}$$

where:

- $Q(K_i)$: the mid-price of options at strike $K_i$
- $\Delta K_i$: the interval between adjacent strike prices
- $K_0$: the strike price immediately below the forward price $F$
- $F$: the forward index level derived from put–call parity:

$$F = K_i + e^{rT}(C_i - P_i)$$

- $T$: the time to expiration (in years),

$$T = \frac{D}{365}$$

### 1.3 Interpolation to 30-Day Variance

Let $\sigma_1^2, \sigma_2^2$ be the variances for maturities $T_1, T_2$. The 30-day variance is linearly interpolated as:

$$\sigma_{30}^2 = \frac{T_1 \sigma_1^2 (T_2 - 30) + T_2 \sigma_2^2 (30 - T_1)}{T_2 - T_1} \tag{2}$$

### 1.4 Final CNVIX Formula

The annualized CNVIX index is then given by:

$$\text{CNVIX} = 100 \times \sqrt{\sigma_{30}^2 \times \frac{365}{30}} \tag{3}$$

## 1.5 Visualization of CNVIX Time Series

To visualize the daily implied volatility level of the Chinese stock market, we plot the constructed CNVIX index based on option market data.

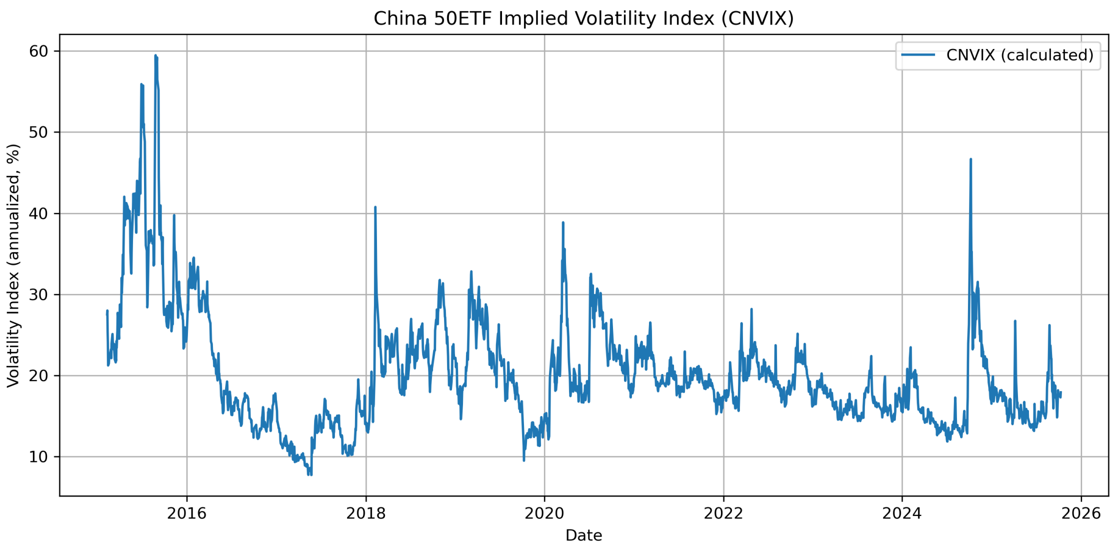


Figure 1: Time Series of CNVIX (China 50ETF Implied Volatility Index)

As shown in Figure 1, the CNVIX exhibits clear time-varying behavior.

# 2 Comparison Between Implied and Realized Volatility

To assess the information content of the CNVIX, we compare the 30-day implied volatility (CNVIX) with the corresponding 30-day *future realized volatility* of the underlying 50ETF. The realized volatility is computed using high-frequency or daily return data and then shifted forward by 30 calendar days to align with the CNVIX's forecasting horizon.

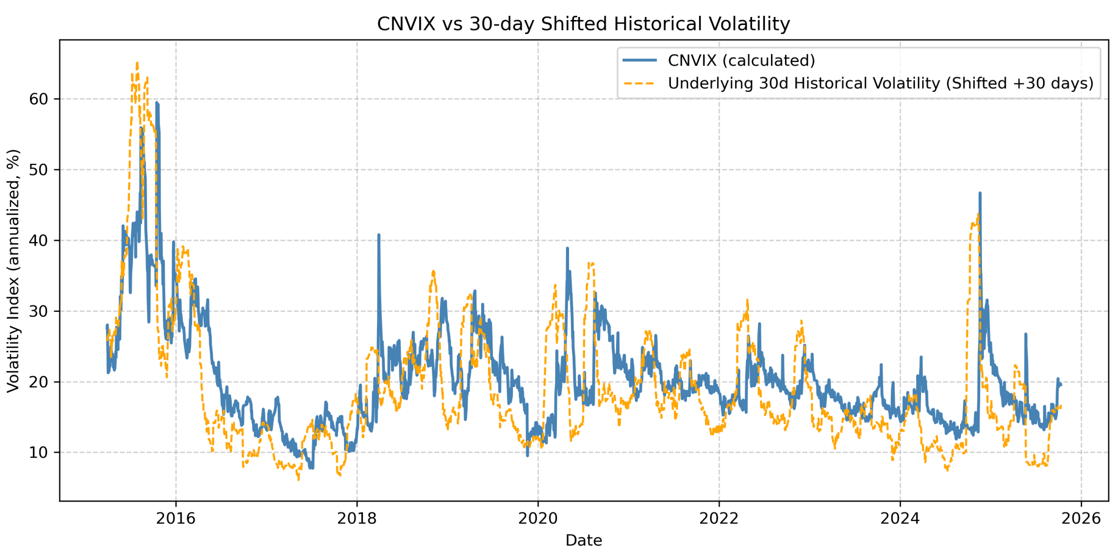


Figure 2: CNVIX vs. 30-day Shifted Realized Volatility

As illustrated in Figure 2, the CNVIX (blue solid line) closely tracks the 30-day-ahead realized volatility (orange dashed line) over most periods. The two series display strong co-movements, with CNVIX typically leading realized volatility, reflecting its nature as a forward-looking measure derived from option prices.

Notably, CNVIX tends to **overestimate** realized volatility during tranquil market periods and **spike earlier and higher** during periods of financial stress.

# 3 Statistical Experiments

## 3.1 Correlation Test between Implied and Realized Volatility

To quantitatively evaluate the forecasting ability of the CNVIX index, we compute the Pearson correlation coefficient between the 30-day implied volatility (CNVIX) and the corresponding 30-day-ahead realized volatility of the underlying 50ETF.

### Methodology

The Pearson correlation coefficient measures the strength of the linear relationship between two variables, defined as:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{t=1}^{T}(X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^{T}(X_t - \bar{X})^2}\sqrt{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}}. \tag{4}$$

Here $X_t$ denotes CNVIX at time $t$, and $Y_t$ represents the realized volatility observed 30 days later. The null hypothesis $H_0 : \rho = 0$ corresponds to no linear correlation.

### Empirical Result

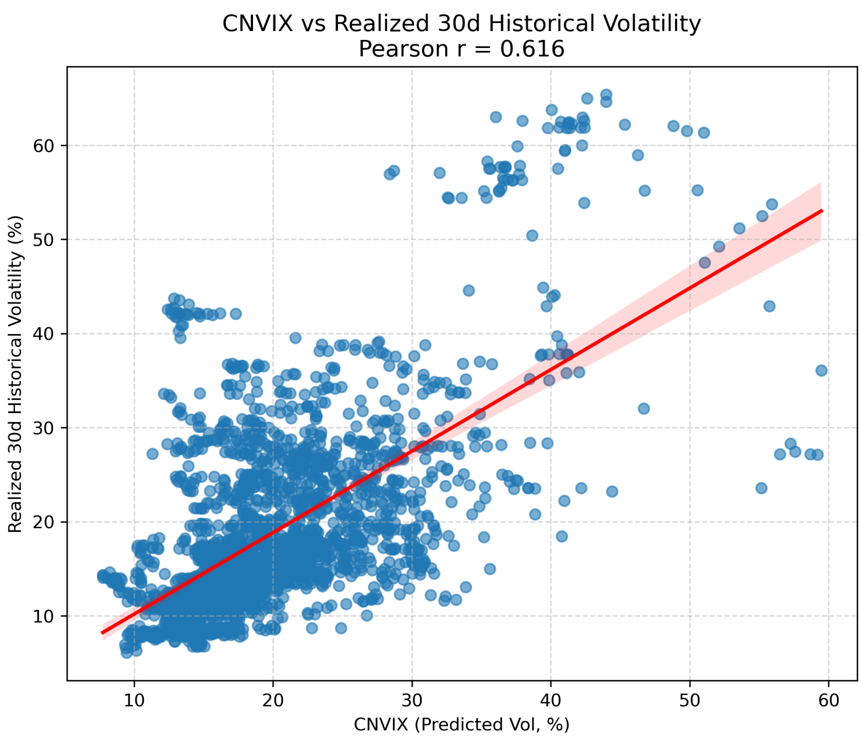


Figure 3: Scatter Plot of CNVIX vs. 30-day Realized Volatility

Table 1 reports the Pearson correlation coefficient and the corresponding p-value.

| Statistic | Pearson $r$ | $p$-value |
|---|---|---|
| Value | 0.6158 | $2.0391 \times 10^{-265}$ |

Table 1: Correlation between CNVIX and Realized 30-day Historical Volatility

This indicates an extremely strong and statistically significant positive relationship between CNVIX and future realized volatility.

As shown in Figure 3, each point represents one trading day, while the red line denotes the least-squares fitted trend. The dense clustering along the upward-sloping line reflects a clear positive linear dependence: periods with higher implied volatility (CNVIX) are generally followed by higher realized volatility in the subsequent month.

## 3.2 Distributional Comparison and Divergence Analysis

While the correlation test in Section 3.1 captures the linear association between CNVIX and the realized volatility, a deeper understanding requires examining their *distributional similarity*. This section compares the empirical distributions of CNVIX and 30-day realized volatility, and quantifies their divergence using the Kullback–Leibler (KL) measure.

### Methodology

The Kullback–Leibler divergence (KL divergence) measures how one probability distribution diverges from a reference distribution. For two continuous random variables $P$ (CNVIX) and $Q$ (realized volatility), it is defined as:

$$D_{\mathrm{KL}}(P\|Q) = \int p(x) \log \frac{p(x)}{q(x)} \, dx, \tag{5}$$

where $p(x)$ and $q(x)$ denote the corresponding probability density functions. In discrete implementation, we estimate $p(x)$ and $q(x)$ via normalized histograms over the same bin grid. Because the measure is asymmetric, we also compute the reverse divergence $D_{\mathrm{KL}}(Q\|P)$.

### Empirical Result

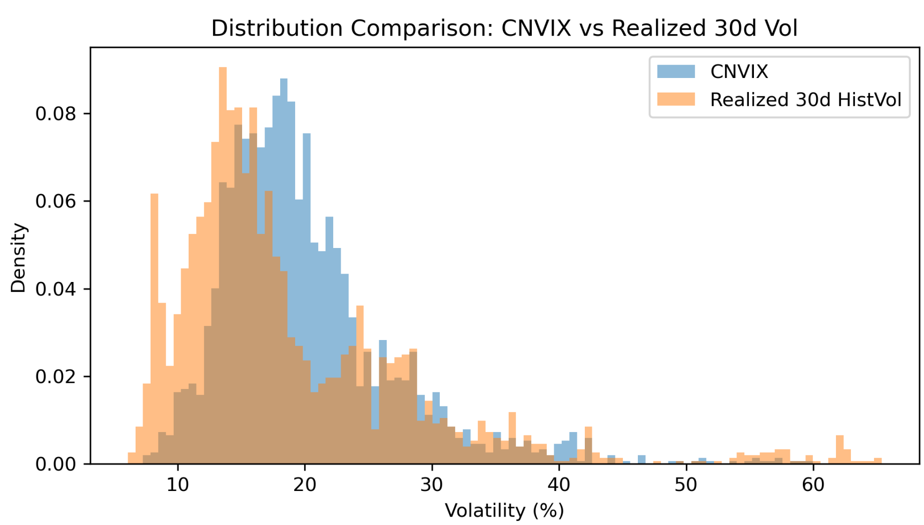


Figure 4: Distribution Comparison between CNVIX and 30-day Realized Volatility

Table 2 reports the computed KL divergence values.

| Measure | $D_{\mathrm{KL}}(P\|Q)$ | $D_{\mathrm{KL}}(Q\|P)$ |
|---|---|---|
| Value | 0.3197 | 0.7872 |

Table 2: KL Divergence between CNVIX and Realized Volatility Distributions

Figure 4 shows that the empirical distribution of CNVIX (blue) is located to the right of the realized volatility distribution (orange), indicating that CNVIX values are generally higher. This pattern is quantitatively confirmed by the positive KL divergence: CNVIX exhibits heavier upper tails, consistent with a systematic risk premium.

The KL divergence values are strictly non-negative, with smaller values indicating higher similarity between distributions. In this case, $D_{\mathrm{KL}}(P\|Q) = 0.32$ implies that CNVIX closely approximates the shape of realized volatility, while $D_{\mathrm{KL}}(Q\|P) = 0.79$ indicates that the realized volatility distribution is narrower and less dispersed than the implied one. Hence, CNVIX systematically overestimates future volatility—a well-documented phenomenon in volatility literature—reflecting the existence of a positive variance risk premium.

# 4 Future Research Directions

The empirical analysis above confirms that the CNVIX contains substantial information about future realized volatility and systematically embeds a variance risk premium. Several promising directions remain for deeper exploration.

## 4.1 Time-Series Dynamics and Model-Based Evaluation

The present study focuses on cross-sectional and distributional comparisons. Future work should incorporate the **time-series structure** of volatility risk, examining whether CNVIX innovations predict future volatility changes beyond contemporaneous correlation.

Specifically, applying econometric frameworks such as the **Newey–West heteroskedasticity-consistent (HAC) test** or **GARCH-family models** would enable a more rigorous dynamic assessment. For instance, a GARCH(1,1) or EGARCH specification could capture persistence and asymmetry in volatility, while the CNVIX level or innovations may serve as exogenous regressors to evaluate predictive power conditional on past volatility. Such modeling could clarify whether CNVIX merely co-moves with realized volatility or truly possesses incremental forecasting content.

## 4.2 Analysis of Extreme Observations

The scatter plot in Figure 3 shows several **extreme observations**—instances where CNVIX significantly over- or under-predicted subsequent realized volatility. Investigating these outliers can yield valuable insights into market behavior under stress or structural shifts.

Future research may:

- Identify and classify extreme points (e.g., top/bottom 2% deviations) and examine the associated market events or macroeconomic announcements.
- Review literature on volatility spikes and market dislocations to contextualize whether these extremes correspond to risk-aversion shocks, liquidity dry-ups, or option market distortions.
- Explore whether such deviations systematically coincide with changes in option-implied skewness, volume, or open interest, thus linking CNVIX inefficiency to underlying option market conditions.