Derivation of confidence intervals for direction-of-change model

Stephan Berke

September 2025

1 Model framework

The specification follows the approach of Christoffersen and Diebold (2006), who propose a binary response model estimated in a rolling-window forecasting framework using information on the conditional mean and the volatility of returns. This baseline model is extended by introducing confidence intervals around the predicted probabilities, constructed via the delta method. This allows for formal inference against the null hypothesis of random guessing.

1.1 Binary response models

Binary response models are parametric models used when the dependent variable y takes on two possible outcomes:

$$y = \begin{cases} 1 & \text{if the event occurs,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

These models estimate the conditional probability of y = 1 given a set of explanatory variables X:

$$P(y = 1|X) = F(X\beta), \tag{2}$$

where $F()$ is the cumulative distribution function (CDF).

In the probit model, the error term is assumed to follow a normal, while in the Logit case, it follows a logistic distribution :

$$\begin{aligned} F_{Probit}(X\beta) &= \Phi(X\beta) \\ F_{Logit}(X\beta) &= \Lambda(X\beta), \end{aligned} \tag{3}$$

where $\Phi()$ denotes the standard normal CDF and $\Lambda()$ the logistic CDF.

1.2 Approach by Christoffersen

Christoffersen and Diebold (2006) propose a binary response framework models for forecasting the direction of returns. The model assumes the following structure of returns:

$$R_{t+1} = \mu_t + \sigma_{t+1}\epsilon_{t+1}, \tag{4}$$

where μ is the conditional mean, σ_{t+1} the conditional standard deviation and $\varepsilon \sim N(0,1)$ is a standard normal innovation.

The binary response variable is defined as the sign of the next-day return:

$$y_t = \begin{cases} 1 & \text{if } R_{t+1} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The goal is to estimate the probability of observing a positive return:

$$\begin{aligned} P(R_{t+1} > 0) &= 1 - P(R_{t+1} \leq 0) \\ &= 1 - P(\epsilon_{t+1} \leq \frac{-\mu_t}{\sigma_{t+1}}) \\ &= 1 - F_\epsilon(\frac{-\mu_t}{\sigma_{t+1}}) \\ &= \Phi(\frac{\mu_t}{\sigma_{t+1}}). \end{aligned} \tag{6}$$

This serves as the baseline equation of the Christoffersen and Diebold (2006) model. In this framework, the predictive probability depends on the conditional mean and volatility. Forecasting volatility is a central component of this framework. In this approach, a standard $AR(1)$ realized variance forecast is used (see Chapter 3.1). Depending on modeling preferences and data characteristics, other forecasting routines, such as HAR models, risk metrics, or GARCH-type models can be applied as well.

The conditional mean is estimated using averages of the previous T days log returns in a rolling window:

$$\mu_t = \frac{1}{T}\sum_{i=1}^{T} r_i. \tag{7}$$

The choice of the window length T introduces a key modeling parameter, as both the estimated mean and volatility depend on it. It is therefore recommended to evaluate model performance across different window sizes to assess the sensitivity of results.

1.2.1 Derivation of confidence intervals

Confidence intervals complement point predictions by quantifying the uncertainty associated with forecasted probabilities. In statistical forecasting, it is standard practice to reflect the inherent uncertainty of model estimates. The strength of this approach is its ability to account for time-varying uncertainty, driven primarily by changes in volatility. This enables a formal testing framework based on the null hypothesis that the predicted probability corresponds to a random guess, defined by a probability of 0.5:

$$H_0 : Pr(R_{t+1} > 0) = 0.5. \tag{8}$$

Under this framework, only predictions whose confidence intervals exclude a probability the value of 0.5 are considered statistically significant and therefore indicative of meaningful directional predictability.

There exist multiple approaches in the literature for constructing confidence intervals for functions of maximum likelihood estimators. One of the most commonly used techniques is the delta method, which employs a first-order Taylor expansion to approximate the variance of nonlinear functions (see, e.g., Casella and Berger (2002)). The following derivation follows the work of Xu and Long (2005), who proposed a routine for constructing confidence intervals for predicted probabilities. A step-by-step derivation is presented in the following.

First, let the previously defined prediction function of interest be:

$$G(\theta) = \Phi(\frac{\mu_t}{\sigma_{t+1}}), \tag{9}$$

where $\theta \in (\mu_t, \sigma_{t+1})$.

The function $G(\hat{\theta})$ can be approximated by a first-order Taylor series expansion to derive its variance:

$$G(\hat{\theta}) \approx G(\theta) + (\hat{\theta} - \theta)' G'(\theta), \tag{10}$$

where $G'(\theta)$ denotes the gradient with respect to θ. Subtracting $G(\theta)$ yields:

$$G(\hat{\theta}) - G(\theta) \approx (\hat{\theta} - \theta)' G'(\theta). \tag{11}$$

Under the standard regularity conditions and assuming asymptotic normality:

$$\sqrt{n}[G(\hat{\theta}) - G(\theta)] \xrightarrow{d} N(0, Var(G(\hat{\theta}))), \tag{12}$$

where the variance of the function is:

$$\begin{aligned} Var(G(\hat{\theta})) &= Var(G(\theta) + (\hat{\theta} - \theta)'G'(\theta)) \\ &= Var((\hat{\theta} - \theta)'G'(\theta)) \qquad \text{(since } G(\theta) \text{ is a constant)} \\ &= Var(\hat{\theta}\, G'(\theta)) \\ &= G'(\theta)^T Var(\hat{\theta}) G'(\theta). \end{aligned} \tag{13}$$

Hence, the gradient of the objective functions needs to be derived, as well as the variances of the estimates. The gradient can be defined as:

$$G'(\theta) = \frac{\partial G(\theta)}{\partial \theta} = [\frac{\partial \Phi(\mu_t/\sigma_{t+1})}{\partial \mu_t} \quad \frac{\partial \Phi(\mu_t/\sigma_{t+1})}{\partial \sigma_{t+1}}]^T, \tag{14}$$

where

$$\frac{\partial \Phi(\mu_t/\sigma_{t+1})}{\partial \mu_t} = \phi(\frac{\mu_t}{\sigma_{t+1}})\frac{1}{\sigma_{t+1}}, \tag{15}$$

and

$$\frac{\partial \Phi(\mu_t/\sigma_{t+1})}{\partial \sigma_{t+1}} = -\phi(\frac{\mu_t}{\sigma_{t+1}})\frac{\mu_t}{\sigma_{t+1}^2}. \tag{16}$$

Here, ϕ denotes the normal probability density function (PDF) as the derivative of the cumulative distribution function.

The variance of $G(\hat{\theta})$ is stated as the variance-covariance matrix:

$$Var(G(\hat{\theta})) = \begin{bmatrix} \text{Var}(\mu_t) & \text{Cov}(\mu_t, \sigma_{t+1}) \\ \text{Cov}(\mu_t, \sigma_{t+1}) & \text{Var}(\sigma_{t+1}) \end{bmatrix}, \tag{17}$$

and requires deriving the sample variances and covariances.

1. **Variance of the sample mean:**

   Let $\mu_t = \frac{1}{T}\sum_{i=1}^{T} r_i$ denote the mean of the past T log returns. Assuming the returns r_i are i.i.d. with variance σ_r^2, the variance of the sample mean is:

   $$\text{Var}(\mu_t) = \frac{\sigma_r^2}{T}. \tag{18}$$

   In empirical applications, σ_r^2 is estimated using the sample variance of the returns in the rolling window.

2. **Variance of the volatility forecast:**

   Recall the AR(1) specification for the square root of realized volatility:

   $$X_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2), \tag{19}$$

where $X_t = \sqrt{RV_t}$. After estimating $\hat{\alpha}$ and $\hat{\beta}$, the residuals are computed as:

$$\hat{\varepsilon}_{t+1} = X_{t+1} - (\hat{\alpha} + \hat{\beta} X_t). \tag{20}$$

The variance of the one-step ahead forecast is given by the residual variance:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{\varepsilon}_{t+1}^2. \tag{21}$$

Under the model assumptions, this represents the conditional variance of the forecast:

$$\mathrm{Var}(X_{t+1} \mid \mathcal{F}_t) = \hat{\sigma}_\varepsilon^2. \tag{22}$$

3. **Covariance of the sample mean and volatility forecast**

   The covariance between μ_t and σ_{t+1} is assumed to be zero:

$$Cov(\mu_t, \sigma_{t+1}) = 0. \tag{23}$$

   This assumption simplifies the application of the delta method. In future, the model can be extended by incorporating return-volatility interactions.

The resulting confidence intervals take the form:

$$CI = G(\theta) \pm z_{\alpha/2} \sqrt{G'(\theta)^T V(\hat{\theta}) G'(\theta)}. \tag{24}$$

This allows for statistical inference on the predicted probability. Specifically, the null hypothesis that the estimated probability corresponds to a random guess of 0.5 can be tested using the test statistic:

$$TS = \frac{G(\theta) - 0.5}{\sqrt{\mathrm{Var}(G(\hat{\theta}))}}. \tag{25}$$

For completion, the p-value can be stated as:

$$p = 2(1 - \Phi(|TS|)). \tag{26}$$

This framework enables the differentiation between statistically significant and non-significant forecasts, allowing for a statistical evaluation framework.

References

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd). Duxbury.

Christoffersen, P. F., & Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, *52*(8), 1273–1287.

Xu, J., & Long, J. S. (2005). Using the delta method to construct confidence intervals for predicted probabilities, rates, and discrete changes. *The Stata Journal*, *5*(4), 537–559.