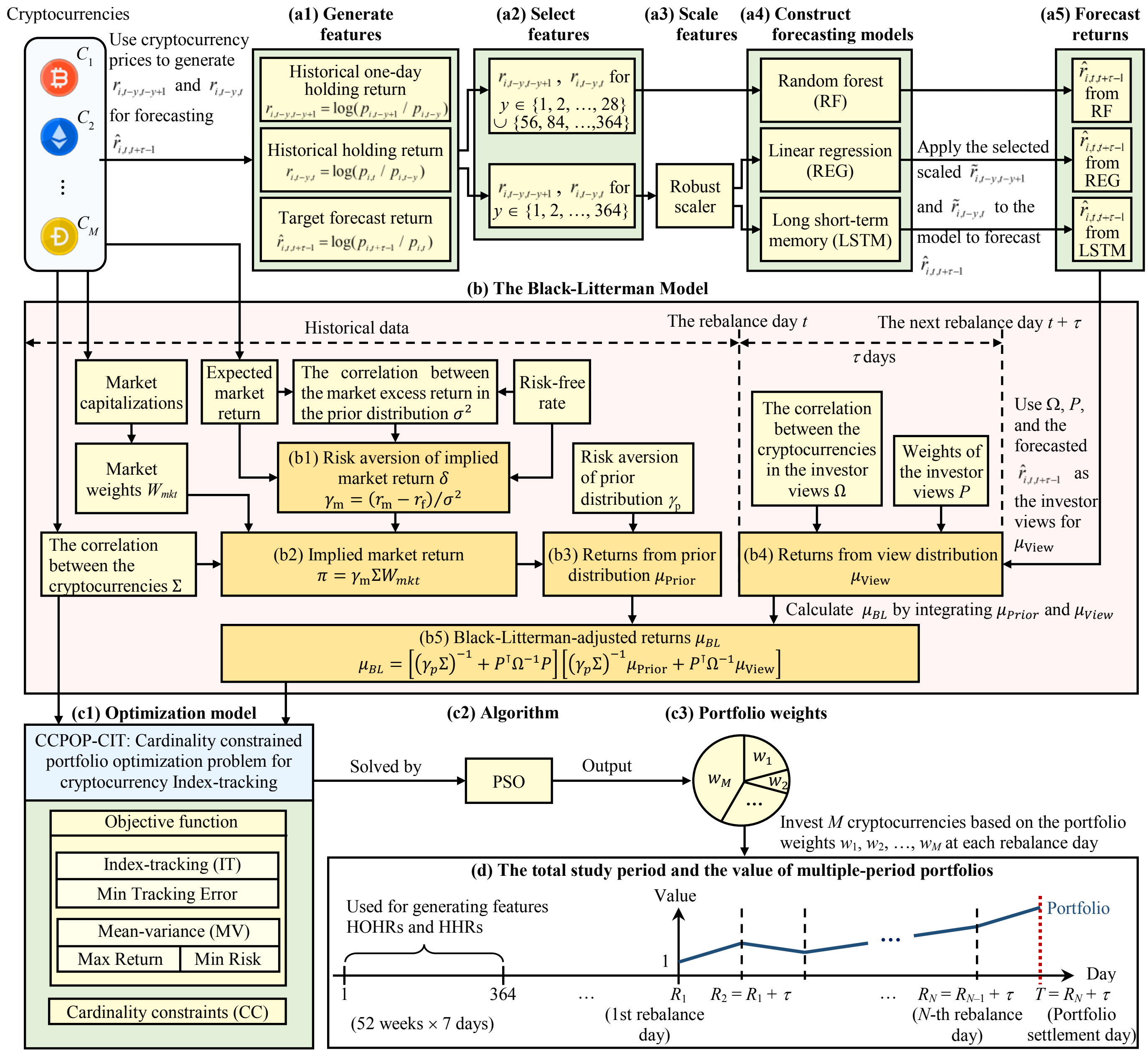


Figure 1. The research framework for CCPOP-CIT.

2. Related Works

This section provides a comprehensive discussion on various Portfolio Optimization Problems (POP), focusing on model construction, return forecasting, the application of the BL model, and optimization algorithms.

	Day	80 features																Label	
		HOHR								HHR								TFR	
		$y = 1$	$y = 2$	…	$y = 28$	$y = 56$	$y = 84$	…	$y = 364$	$y = 1$	$y = 2$	…	$y = 28$	$y = 56$	$y = 84$	…	$y = 364$		
Training set	$t-119$	$r_{i,t-120,t-119}$	$r_{i,t-121,t-120}$	…	$r_{i,t-147,t-146}$	$r_{i,t-175,t-174}$	$r_{i,t-203,t-202}$	…	$r_{i,t-483,t-482}$	$r_{i,t-120,t-119}$	$r_{i,t-121,t-119}$	…	$r_{i,t-147,t-119}$	$r_{i,t-175,t-119}$	$r_{i,t-203,t-119}$	…	$r_{i,t-483,t-119}$	$\hat{r}_{i,t-119,t-120+\tau}$	
	$t-118$	$r_{i,t-119,t-118}$	$r_{i,t-120,t-119}$	…	$r_{i,t-146,t-145}$	$r_{i,t-174,t-173}$	$r_{i,t-202,t-201}$	…	$r_{i,t-482,t-481}$	$r_{i,t-119,t-118}$	$r_{i,t-120,t-118}$	…	$r_{i,t-146,t-118}$	$r_{i,t-174,t-118}$	$r_{i,t-202,t-118}$	…	$r_{i,t-482,t-118}$	$\hat{r}_{i,t-118,t-119+\tau}$	
	⋮	⋮								⋮								⋮	
	$t-30$	$r_{i,t-31,t-30}$	$r_{i,t-32,t-31}$	…	$r_{i,t-58,t-57}$	$r_{i,t-86,t-85}$	$r_{i,t-114,t-113}$	…	$r_{i,t-394,t-393}$	$r_{i,t-31,t-30}$	$r_{i,t-32,t-30}$	…	$r_{i,t-58,t-30}$	$r_{i,t-86,t-30}$	$r_{i,t-114,t-30}$	…	$r_{i,t-394,t-30}$	$\hat{r}_{i,t-30,t-31+\tau}$	
Forecasting window	$t-29$	$r_{i,t-30,t-29}$	$r_{i,t-31,t-30}$	…	$r_{i,t-57,t-56}$	$r_{i,t-85,t-84}$	$r_{i,t-113,t-112}$	…	$r_{i,t-393,t-392}$	$r_{i,t-30,t-29}$	$r_{i,t-31,t-29}$	…	$r_{i,t-57,t-29}$	$r_{i,t-85,t-29}$	$r_{i,t-113,t-29}$	…	$r_{i,t-393,t-29}$	$\hat{r}_{i,t-29,t-30+\tau}$	Pre-test samples
	$t-28$	$r_{i,t-29,t-28}$	$r_{i,t-30,t-29}$	…	$r_{i,t-56,t-55}$	$r_{i,t-84,t-83}$	$r_{i,t-112,t-111}$	…	$r_{i,t-392,t-391}$	$r_{i,t-29,t-28}$	$r_{i,t-30,t-28}$	…	$r_{i,t-56,t-28}$	$r_{i,t-84,t-28}$	$r_{i,t-112,t-28}$	…	$r_{i,t-392,t-28}$	$\hat{r}_{i,t-28,t-29+\tau}$	
	⋮	⋮								⋮								⋮	
	t	$r_{i,t-1,t}$	$r_{i,t-2,t-1}$	…	$r_{i,t-28,t-27}$	$r_{i,t-56,t-55}$	$r_{i,t-84,t-83}$	…	$r_{i,t-364,t-363}$	$r_{i,t-1,t}$	$r_{i,t-2,t}$	…	$r_{i,t-28,t}$	$r_{i,t-56,t}$	$r_{i,t-84,t}$	…	$r_{i,t-364,t}$	$\hat{r}_{i,t,t+\tau-1}$	Test sample

Figure 2. Training set and forecasting window for predicting TFR using RF.

4.2.2. Linear regression (REG)

By giving a regression coefficient to each feature, REG can be used to estimate the value of the forecasting target. First, this study takes all the features HOHR $r_{i,t-y,t-y+1}$ and HHR $r_{i,t-y,t}$ the year before day t, , i.e., $y \in \{1, 2, \ldots, 364\}$, as illustrated in Figure 3. Different from RF using selective feature sampling, REG utilizes continuous time-series data to learn patterns within the sequence, obtaining the overall trend of the complete time-series data.

Then, these features are scaled by robust scaler as follows:

$$\tilde{r}_{i,t-y,t-y+1} = (r_{i,t-y,t-y+1} - Q_2) / (Q_3 - Q_1) \quad (15)$$

$$\tilde{r}_{i,t-y,t} = (r_{i,t-y,t} - Q_2) / (Q_3 - Q_1) \tag{16}$$

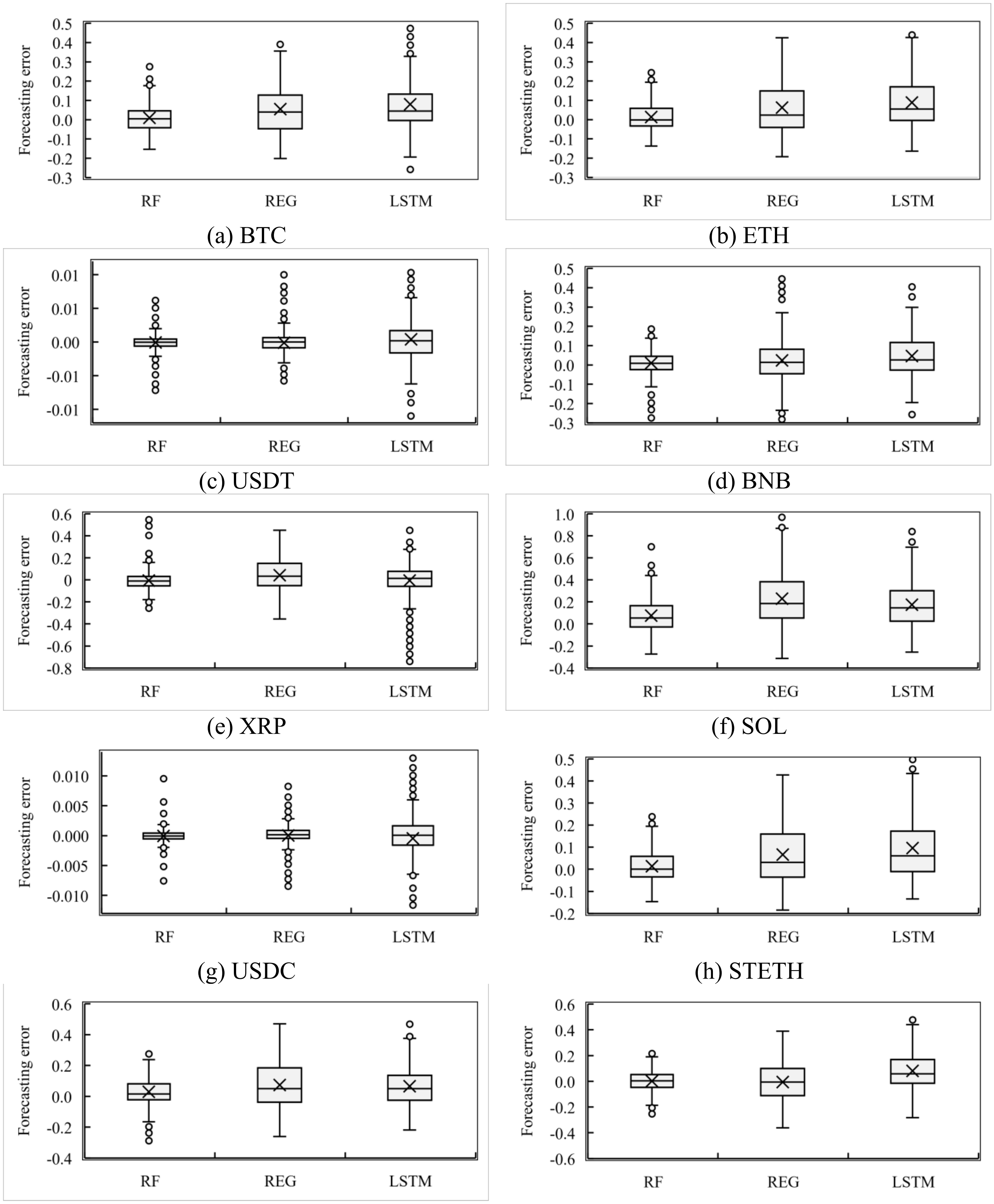
for each $y \in \{1, \ldots, 364\}$, where Q_1, Q_2, and Q_3 are the first, second, and third quartiles of $\{r_{i,t-y,t-y+1}, r_{i,t-y,t} \mid \forall\ y \in \{1, \ldots, 364\}\}$, respectively. Then, this study uses the features to obtain the regression coefficients β_q of all the scaled features through the following regression equation:

$$\begin{aligned} \hat{r}_{i,t,t+\tau-1} = \beta_0 + \beta_1 \cdot \tilde{r}_{i,t-1,t} + \beta_2 \cdot \tilde{r}_{i,t-2,t-1} + \cdots + \beta_{364} \cdot \tilde{r}_{i,t-364,t-363} \\ + \beta_{365} \cdot \tilde{r}_{i,t-1,t} + \beta_{366} \cdot \tilde{r}_{i,t-2,t} + \cdots + \beta_{728} \cdot \tilde{r}_{i,t-364,t} + \rho \end{aligned} \tag{17}$$

where ρ is the estimation error. Then, it can be used to forecasts the TFR $\hat{r}_{i,t,t+\tau-1}$.

	Day	80 features								Label	
		HOHR				HHR				TFR	
		$y = 1$	$y = 2$	…	$y = 364$	$y = 1$	$y = 2$	…	$y = 364$		
Training set	$t-119$	$r_{i,t-120,t-119}$	$r_{i,t-121,t-120}$	…	$r_{i,t-483,t-482}$	$r_{i,t-120,t-119}$	$r_{i,t-121,t-119}$	…	$r_{i,t-483,t-119}$	$\hat{r}_{i,t-119,t-120+\tau}$	
	$t-118$	$r_{i,t-119,t-118}$	$r_{i,t-120,t-119}$	…	$r_{i,t-482,t-481}$	$r_{i,t-119,t-118}$	$r_{i,t-120,t-118}$	…	$r_{i,t-482,t-118}$	$\hat{r}_{i,t-118,t-119+\tau}$	
	⋮	⋮				⋮				⋮	
	$t-30$	$r_{i,t-31,t-30}$	$r_{i,t-32,t-31}$	…	$r_{i,t-394,t-393}$	$r_{i,t-31,t-30}$	$r_{i,t-32,t-30}$	…	$r_{i,t-394,t-30}$	$\hat{r}_{i,t-30,t-31+\tau}$	
Forecasting window	$t-29$	$r_{i,t-30,t-29}$	$r_{i,t-31,t-30}$	…	$r_{i,t-393,t-392}$	$r_{i,t-30,t-29}$	$r_{i,t-31,t-29}$	…	$r_{i,t-393,t-29}$	$\hat{r}_{i,t-29,t-30+\tau}$	Pre-test samples
	$t-28$	$r_{i,t-29,t-28}$	$r_{i,t-30,t-29}$	…	$r_{i,t-392,t-391}$	$r_{i,t-29,t-28}$	$r_{i,t-30,t-28}$	…	$r_{i,t-392,t-28}$	$\hat{r}_{i,t-28,t-29+\tau}$	
	⋮	⋮				⋮				⋮	
	t	$r_{i,t-1,t}$	$r_{i,t-2,t-1}$	…	$r_{i,t-364,t-363}$	$r_{i,t-1,t}$	$r_{i,t-2,t}$	…	$r_{i,t-364,t}$	$\hat{r}_{i,t,t+\tau-1}$	Test sample

Figure 3. Training set and forecasting window for predicting TFR using REG or LSTM.



(a) BTC (b) ETH

(c) USDT (d) BNB

(e) XRP (f) SOL

(g) USDC (h) STETH

(i) ADA (j) DOGE

Figure 4. Box plots of forecasting error results of the 10 cryptocurrencies of the portfolios from REG, LSTM, and RF.

Table 6. The realized fitness and various performance results of the portfolios from the proposed method for the CCPOP-CIT and the SPCBLC cryptocurrency index.

γ_p	Forecasting model	Realized fitness	Excess return (ER)	Volatility (VOL)	Sharpe ratio (SR)	Maximum drawdown (MDD)	Calmar ratio (CR)
0.1	RF	**0.2114**	**0.9087**	0.3976	**2.0940**	0.2527	**3.5957**
0.1	REG	**0.2065**	**1.3354**	0.4479	**2.9817**	0.3037	**4.3965**
0.1	LSTM	**0.2227**	**0.7770**	0.4540	**1.7115**	0.2966	2.6194
0.3	RF	0.2359	0.7250	0.4049	1.7904	0.2194	3.3042
0.3	REG	0.2335	0.6901	0.4050	1.7037	0.2447	2.8200
0.3	LSTM	0.2499	0.6120	0.3878	1.5779	0.2156	**2.8388**
0.5	RF	0.2513	0.6614	0.3852	1.7173	0.2139	3.0921
0.5	REG	0.2519	0.6809	0.3973	1.7138	0.2177	3.1279
0.5	LSTM	0.2603	0.5906	0.4045	1.4602	0.2253	2.6211
0.7	RF	0.2658	0.5976	0.4022	1.4858	0.2251	2.6551
0.7	REG	0.2543	0.6590	**0.3914**	1.6838	0.2118	3.1118
0.7	LSTM	0.2741	0.5216	0.3831	1.3613	0.2361	2.2095
0.9	RF	0.2703	0.5061	**0.3623**	1.3969	**0.1764**	2.8693
0.9	REG	0.2555	0.6043	0.3939	1.5341	**0.2062**	2.9308
0.9	LSTM	0.2970	0.4794	**0.3769**	1.2720	**0.1793**	2.6740
SPCBLC		-	1.0036	0.4605	2.1794	0.2364	4.2457

7.4. Analysis of investment weights

Figure 5 shows the investment weights of the 10 cryptocurrencies of the portfolios from the return forecasting models of REG, LSTM, and RF. Note that the lower bound of investment weight $w_i^{\min}$ is set to 0.01; and the upper bound of investment weight $w_i^{\max}$ is set as the maximum market capitalization ratio of cryptocurrencies and is recalculated at each portfolio rebalance with the highest value of 0.6067 and the lowest value of 0.4752. From Figure 5, it can be observed that all of the portfolios have complied with $w_i^{\max}$.

In addition, regarding the investment weights and portfolio performance, REG and RF have higher investment returns, and often have cryptocurrency investment weights close to the upper bound, in which there are one or two cryptocurrencies with investment weights higher than 0.4 in each optimization. On the other hand, LSTM tends to diversify its investment, with most cryptocurrencies with investment weights ranging from 0.01 to 0.2. According to the results in Table 6, the REG and RF portfolios with concentrated investments outperform the LSTM portfolios with diversified investments with respect to return, risk and risk-adjusted returns.

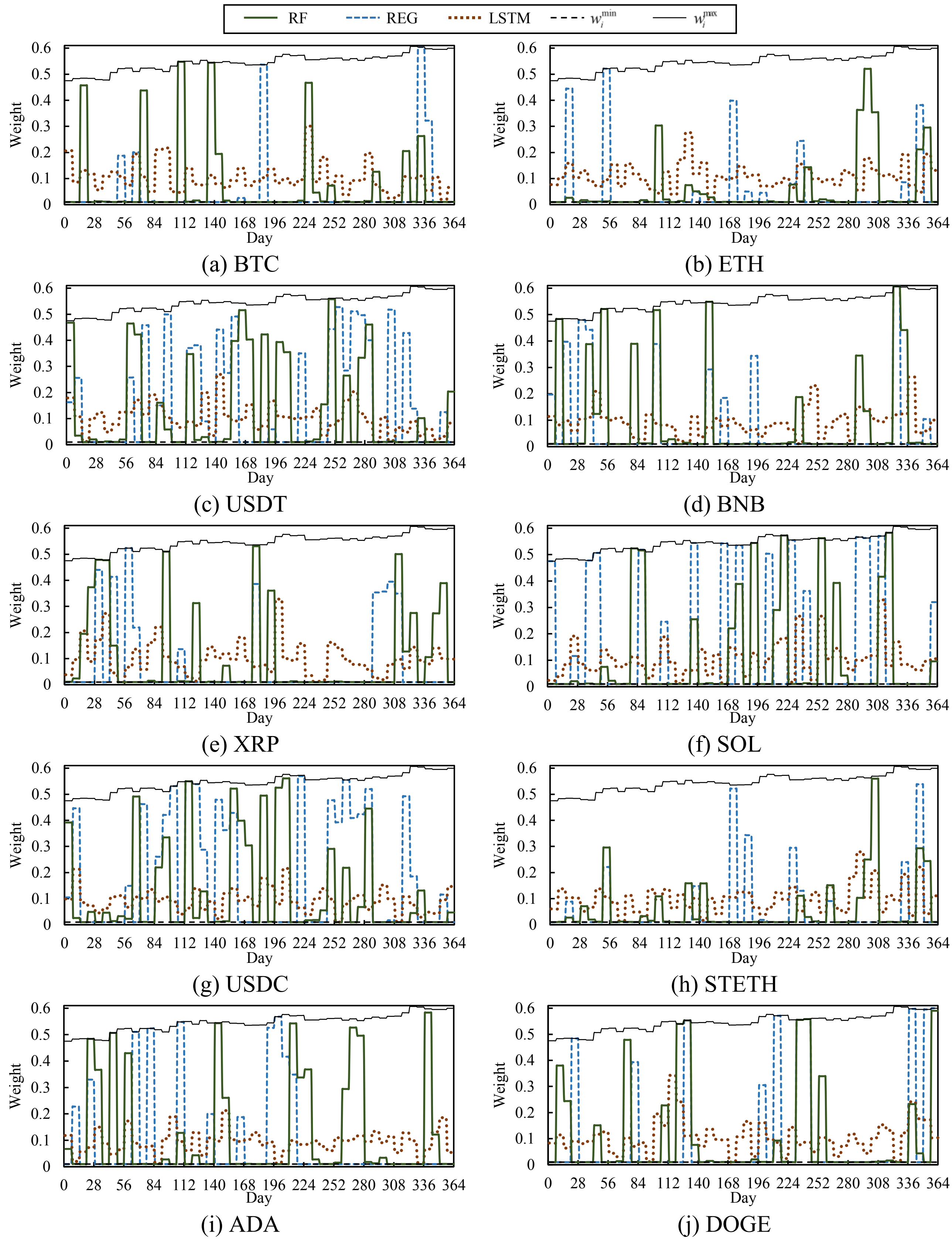
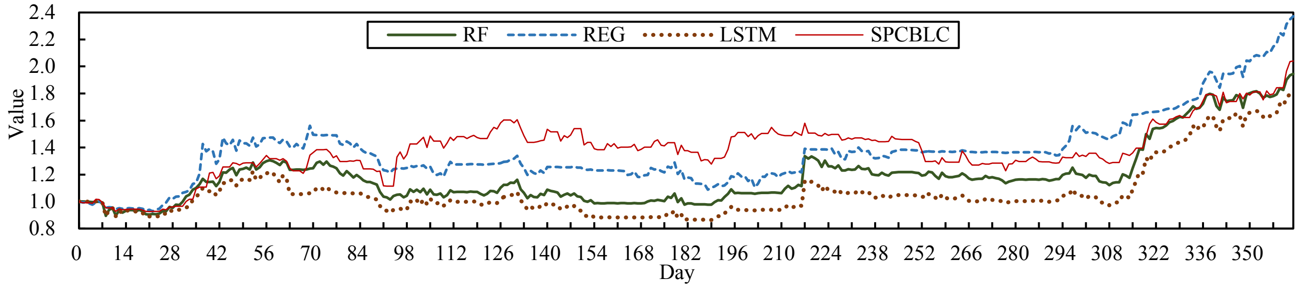
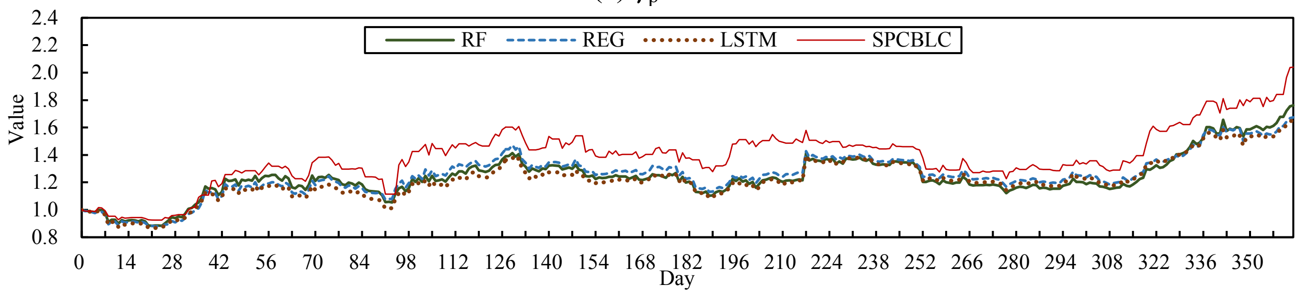


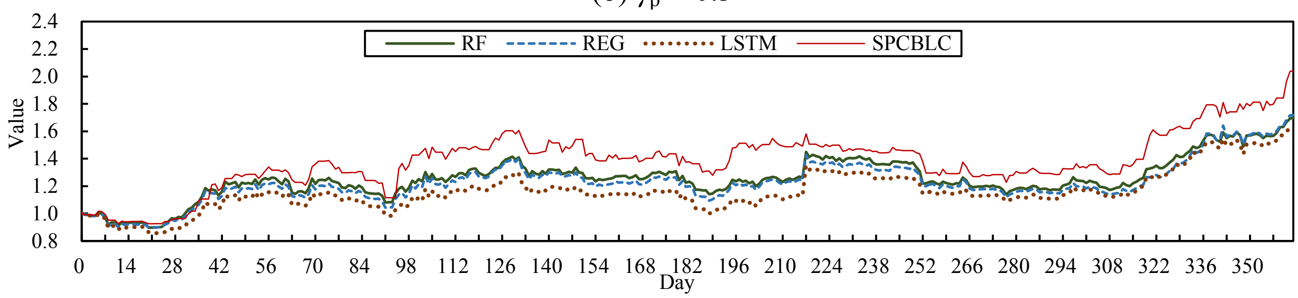
Figure 5. Investment weights of the top 10 market-cap-ranked cryptocurrencies of the portfolios from REG, LSTM, and RF.



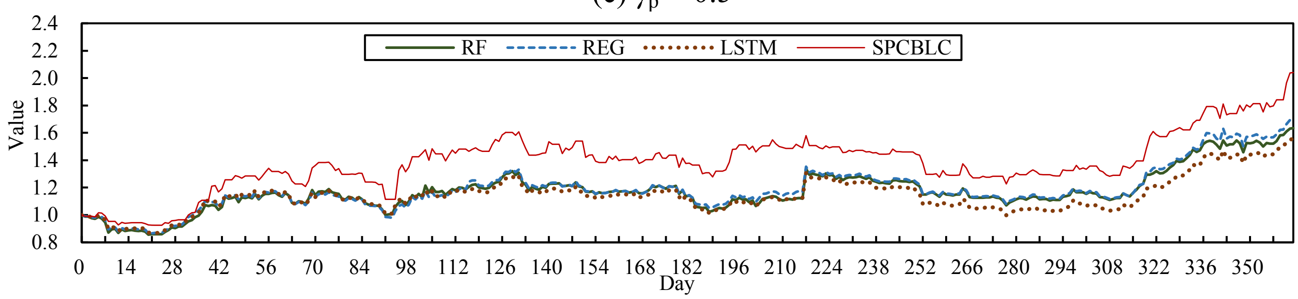
(a) $\gamma_p = 0.1$



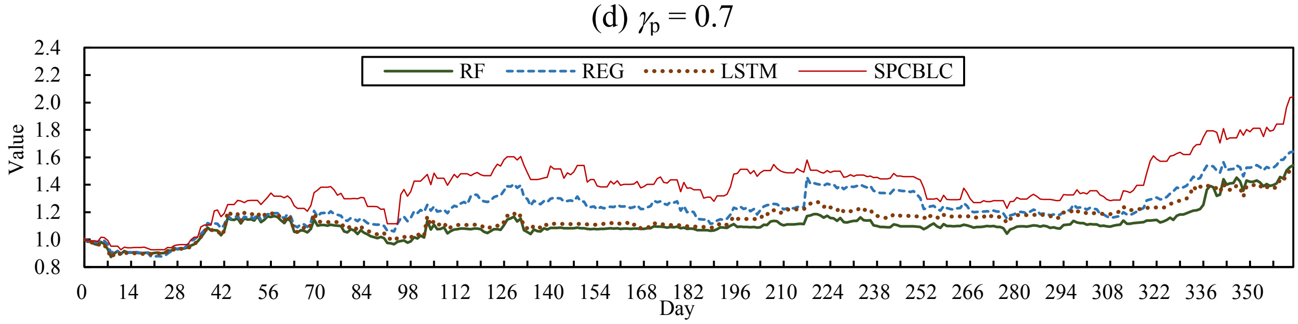
(b) $\gamma_p = 0.3$



(c) $\gamma_p = 0.5$



(d) $\gamma_p = 0.7$



(e) $\gamma_p = 0.9$

Figure 6. The dynamics of the value of portfolios using different return forecast models under different risk aversion parameters γ_p.