Anomalies and False Rejections

Tarun Chordia Amit Goyal Alessio Saretto

ZIHAO MA

01

Multiple Hypotheses Testing

Multiple Hypotheses Testing

	H_0 not rejected	H_0 rejected	Total
H_0 true	T_0	F_1	S_0
H_0 false	F_0	T_1	S_1
	$S - R$	$R = F_1 + T_1$	S

Overall, F1 hypotheses are falsely rejected, and F1/R is the FDP. F0 hypotheses should have been rejected but are not, and F0/(S−R) is the false non-discovery proportion (FNDP).

There are three broad approaches differ in the statistical definition of what is actually controlled: familywise error rate (FWER), false discovery rate (FDR), and false discovery proportion (FDP).

1.1 Familywise error rate

The most basic idea in MHT is to control the FWER, defined as the probability of rejecting at least one of the true null hypotheses:

$$FWER = Prob(F1 \geq 1)$$

1.2 False discovery proportion

An alternative to controlling the “number” of false rejections is to control the “proportion” $F1/R$ of false rejections (FDP). Formally, a multiple testing procedure controls FDP at proportion γ and significance level α if

$$\text{Prob}\,(\text{FDP} \geq \gamma) \leq \alpha.$$

As such, FDP control guarantees that, in any application, the realized FDP cannot (statistically) exceed a particular threshold γ.

1.2 False discovery proportion

RSW Procedure

Althors implement the RSW method proposed by Romano and Wolf (2007) and Romano, Shaikh, and Wolf (2008) to control FDP. This procedure runs a sequence of k-FWER procedures with $k = 1, 2$, etc. At a generic step k, the total number of rejections is R_k. Step k corresponds to k-FWER. Therefore, at most k false rejections have taken place so far. This means that one additional rejection in the next step will make the FDP equal to $k/(R_k + 1)$. To control the FDP at the desired significance level γ, we undertake the next step only if $k/(R_k + 1) \leq \gamma$, or, equivalently, only if $R_k \geq k/\gamma - 1$. The last inequality determines the stopping condition of the algorithm.

1.3 False discovery rate

An alternative to controlling the tail behavior of FDP is to control its expected value, the FDR. Formally, a multiple testing method controls FDR at level δ if

$$FDR \equiv E(FDP) \leq \delta$$

where δ is a tolerance level imposed by the researcher.

The two main FDR control methods that we consider are the Benjamini and Hochberg (1995) method and its extension by Benjamini and Yekutieli (2001). These two methods are computationally easy to implement.

02 Monte Carlo Simulations

Real Data

VS

Simulation

2.1 Simulation details

2.1.1 Data-generating process.

Stocks:

There are N stocks whose returns are generated from a linear factor model:

$$R_{it} = \alpha_i + \beta_i' F_t + \epsilon_{it}. \tag{1}$$

Signals:

Authors draw an informative signal with probability π. After drawing a signal, its realizations corresponding to stock i in time period t are given by:

$$s_{it} = \begin{cases} \alpha_i + \eta_{it}^s & \text{if the signal is informative,} \\ \eta_{it}^s & \text{if the signal is uninformative.} \end{cases} \tag{2}$$

signal-to-noise ratio, $\pi \times \sigma_\alpha^2 / \sigma_\eta^2$.

2.1.2 Portfolios and cross-sectional regressions.

Authors analyze the simulated data in a similar way as they do with the real data. Every month, we sort stocks into deciles based on signal s and hold these portfolios for 1 month. The hedge portfolio goes long in decile 10 and short in decile 1. Authors run a time-series regression of the hedge portfolio returns, Rst, based on signal s, on factor returns Ft and record the t-statistic of the alpha, tα

$$R_{st} = \alpha_s + \beta_s' F_t + u_{st}. \qquad (3)$$

Authors obtain S portfolios and corresponding number of tαs. Every period t we also run a univariate cross-sectional regression of risk-adjusted returns, (Rit-βˆiFt), as in Brennan, Chordia, and Subrahmanyam (1998), on signal sit

$$R_{it} - \hat{\beta}_i' F_t = \lambda_{0t} + \lambda_{st} s_{it} + \psi_{it},$$

and compute the usual Fama and MacBeth (1973) t-statistic, tλs .

A. Signal-to-noise ratio						*B. Correlation between signals*					
σ_α	Bonf	Holm	BH	BHY	RSW	ρ	Bonf	Holm	BH	BHY	RSW
	Thresholds						Thresholds				
1.5	4.6	4.6	3.1	3.8	3.5	0	4.6	4.5	3.0	3.8	3.4
2.0	4.6	4.4	3.0	3.8	3.4	10	4.6	4.5	3.0	3.9	3.5
2.5	4.6	4.4	3.0	4.1	3.4	20	4.6	4.5	3.0	4.0	3.6
	Rejections						Rejections				
1.5	1.8	1.8	4.6	3.4	3.7	0	5.0	5.0	5.2	5.0	5.1
2.0	5.0	5.0	5.2	5.0	5.1	10	5.0	5.0	5.3	5.0	5.1
2.5	5.0	5.0	5.2	5.0	5.1	20	5.0	5.0	5.3	5.0	5.0
	E[FDP]						E[FDP]				
1.5	<0.1	<0.1	4.8	0.5	1.0	0	<0.1	<0.1	4.7	0.5	1.2
2.0	<0.1	<0.1	4.7	0.5	1.2	10	<0.1	<0.1	4.8	0.5	1.0
2.5	<0.1	0.2	4.7	0.5	1.2	20	<0.1	<0.1	4.8	0.5	0.8
	SD[FDP]						SD[FDP]				
1.5	0.1	0.1	1.0	0.4	0.5	0	<0.1	0.1	1.0	0.3	0.5
2.0	<0.1	0.1	1.0	0.3	0.5	10	<0.1	<0.1	2.3	0.5	0.8
2.5	<0.1	0.1	1.0	0.3	0.5	20	0.1	0.1	4.3	0.8	1.2
	E[FNDP]						E[FNDP]				
1.5	63.9	63.7	12.7	33.0	25.8	0	0.3	0.3	<0.1	<0.1	<0.1
2.0	0.3	0.3	<0.1	<0.1	<0.1	10	0.2	0.2	<0.1	<0.1	<0.1
2.5	0	0	0	0	0	20	0.3	0.3	<0.1	<0.1	<0.1

C. Proportion of informative signals						*D. Number of signals*					
π	Bonf	Holm	BH	BHY	RSW	S	Bonf	Holm	BH	BHY	RSW
	Thresholds						Thresholds				
0	4.6	4.5	3.0	3.8	3.4	5,000	4.4	4.4	3.0	4.1	3.5
10	4.6	4.5	2.8	3.5	3.2	10,000	4.6	4.5	3.0	3.8	3.4
20	4.6	4.5	2.6	3.3	2.8	100,000	5.0	5.0	3.0	3.7	3.4
30	4.6	4.5	2.4	3.2	2.6	500,000	5.2	5.2	3.0	3.7	3.4
	Rejections						Rejections				
0	5.0	5.0	5.2	5.0	5.1	5,000	5.0	5.0	5.2	5.0	5.0
10	10.0	10.0	10.5	10.0	10.2	10,000	5.0	5.0	5.2	5.0	5.1
20	19.9	20.0	20.8	20.1	20.4	100,000	4.9	4.9	5.2	5.0	5.1
30	29.9	29.9	31.1	30.1	30.6	500,000	4.9	4.9	5.2	5.0	5.1
	E[FDP]						E[FDP]				
0	<0.1	<0.1	4.7	0.5	1.2	5,000	<0.1	<0.1	4.7	0.5	0.9
10	<0.1	<0.1	4.4	0.5	1.5	10,000	<0.1	<0.1	4.7	0.5	1.2
20	<0.1	<0.1	3.9	0.4	1.8	100,000	<0.1	<0.1	4.7	0.4	1.2
30	<0.1	<0.1	3.5	0.4	2.1	500,000	<0.1	<0.1	4.7	0.4	1.2
	SD[FDP]						SD[FDP]				
0	<0.1	0.1	1.0	0.3	0.5	5,000	0.1	0.1	1.3	0.5	0.7
10	<0.1	<0.1	0.7	0.2	0.5	10,000	<0.1	0.1	1.0	0.3	0.5
20	<0.1	<0.1	0.4	0.1	0.4	100,000	<0.1	<0.1	0.4	0.1	0.3
30	<0.1	<0.1	0.3	0.1	0.4	500,000	<0.1	<0.1	0.3	0.1	0.3
	E[FNDP]						E[FNDP]				
0	0.3	0.3	<0.1	<0.1	<0.1	5,000	0.2	0.2	0	<0.1	<0.1
10	0.3	0.3	<0.1	<0.1	<0.1	10,000	0.3	0.3	0	<0.1	<0.1
20	0.3	0.2	<0.1	<0.1	<0.1	100,000	1.2	1.1	<0.1	<0.1	<0.1
30	0.3	0.2	0	<0.1	0	500,000	1.8	1.9	<0.1	<0.1	<0.1

Properties of different MHT methods

After comparison of three MHT Methods, Authors choose the RSW Method for 3 reasons: First, FWER control is not appropriate for finance applications as it lacks power and is mechanically affected by the number of hypotheses being tested. Second, RSW is in between the two FDR control methods in terms of both size and power. Third, as RSW is meant to control the tail behavior of FDP, it avoids the concerns that controlling FDR exposes a researcher to the possibility that the realized FDP varies significantly across applications.

2.2 Dual hurdles

Authors also use the practice of dual hurdles in this research method. Such practice helps in reducing the probability of false rejections for all testing methods. Coupled with MHT methods, the practice makes for a very effective tool to guard against false rejections.

		Threshold	Rejections		E[FDP]	
π	σ_α	RSW	RSW	SHT	RSW	SHT
		Portfolio alpha: t_α				
0	2.0	4.0	<0.1	4.9	—	—
5	1.5	3.5	3.7	9.6	1.1	48.7
5	2.0	3.4	5.1	9.7	1.2	48.2
5	2.5	3.4	5.1	9.7	1.2	48.4
		Fama-MacBeth: t_λ				
0	2.0	4.0	<0.1	5.0	—	—
5	1.5	3.2	5.1	9.8	2.7	48.9
5	2.0	3.2	5.1	9.8	2.7	48.9
5	2.5	3.2	5.1	9.8	2.8	48.9
		t_α and t_λ				
0	2.0	—	<0.1	1.5	—	—
5	1.5	—	3.7	6.3	0.2	22.0
5	2.0	—	5.0	6.4	0.2	21.8
5	2.5	—	5.0	6.4	0.2	21.9

03 Data and Trading Strategies

3.1 Data

Data source: Monthly returns and prices are obtained from CRSP. Annual accounting data come from the merged CRSP/Compustat files. Including All items included in the balance sheet, the income statement, the cash flow statement, and other miscellaneous items for the years 1972 to 2015. Authors choose 1972 as the beginning of sample as it corresponds to the first year of trading on Nasdaq that dramatically increased the number of stocks in the CRSP data set. Also, authors set a 6-month lag between the end of the fiscal year and the availability of accounting information.

3.2 Hedge portfolios

Authors sort firms into value-weighted deciles on June 30 of each year and rebalance these portfolios annually.

Authors consider long-short portfolios only. Thus, authors compute a hedge portfolio return that is long in decile ten and short in decile one.

For each trading strategy, we run a time-series regression of the corresponding hedge portfolio returns on the Fama and French (2015) five-factor model augmented with the momentum factor (Carhart 1997) and obtain the alpha as well as its heteroscedasticity adjusted t-statistic, $t\alpha$.

3.3 Fama-MacBeth regressions

Authors estimate the following FM regression each month:

$$\breve{R}_{it} - \hat{\beta}_i \breve{F}_t = \lambda_{0t} + \lambda_{1t} X_{it-1} + \lambda_{2t} Z_{it-1} + e_{it}, \quad (4)$$

where X is the variable that represents the signal and Z's are control variables. We calculate the FM coefficient λ1 as well its heteroscedasticity-adjusted t-statistic (tλ)

04

Results

Calibration

Calibrated parameters: Authors calibrate parameters related to the data-generating process for stocks and the researchers' advantage over econometricians in picking informative signals.

Ultimately, the calibration allows authors to characterize the proportion of false rejections under SHT, and MHT thresholds based on the RSW procedure. The proportion of false rejections is 45.3%, and thresholds are 3.8 and 3.4 for alpha and FM coefficient t-statistics, respectively.

A. Calibrated parameters		
σ_η	π	Ω
4.8	0.06	29.0
B. Target quantities		
	Data	Simulation
SnM rejections in $\mathcal{S}^{\mathcal{E}}$	97.9	96.5
SD(ψ) in $\mathcal{S}^{\mathcal{E}}$	17.2	17.5
$E(\lvert\text{sign. } \alpha_s\rvert)/E(\lvert\text{sign. } \alpha_i\rvert)$ in $\mathcal{S}^{\mathcal{E}}$	0.12	0.12
SnM rejections in $\mathcal{S}^{\mathcal{P}}$	27.0	30.9
C. Outcome quantities in the set $\mathcal{S}^{\mathcal{R}}$		
Alpha threshold, $\mathcal{T}_\alpha$		3.8
FM threshold, $\mathcal{T}_\lambda$		3.4
SnM rejections		42.5
False rejections		45.3

05

Robustness Checks

How the results changed by alternative assumptions, definitions…

5.1 Alternative assumptions about stock returns and signals

The time-series averages of the cross-sectional standard deviation, skewness, and kurtosis of estimated stock's alphas.

bootstrap factor returns by resampling them in stationary blocks with replacement directly from the observed six-factor monthly returns. The bootstrap preserves the higher-order moments and factor crashes observed in the data.

the cross-sectional average of residuals standard deviations, skewness, and kurtosis (again using rolling regressions for each stock).

the cross-sectional correlation, ρ, between strategies at 9% and repeat the calibration

5.2 Different factor models

four alternatives to the Fama and French (2015) five-factor model augmented by the momentum factor: a one-factor model with the market factor (capital asset pricing model, CAPM); the Fama and French (1993) three-factor model (FF3); the Barillas and Shanken (2018) six-factor model (BS); and the Hou, Xue and Zhang (2015) q-model (HXZ).

06 Conclusion

Authors apply MHT to obtain an estimate of the proportion of false discoveries and thresholds for alpha and FM coefficient t-statistics.

the proportion of false rejections under single hypothesis testing is about 45%, and thresholds are 3.8 and 3.4 for alpha and FM coefficient t-statistics, respectively.

THANKS