ØAMET2200 · Fall 2019
Worksheet 8

Instructor: Fenella Carpena

October 25, 2019

Exercise 1 Does the stock market efficiently use information in valuing stocks? The Efficient Markets Hypothesis (EMH), developed by Nobel-prize winner Eugene Fama, maintains that current stock prices fully reflect all available information. An implication of this hypothesis is that returns in the current period should not be systematically related to information known in earlier periods. Otherwise, we could use this information to predict stock returns, thus violating EMH.

As an analyst at an investment management company, you have been tasked with examining the validity of the EMH. You obtained a (fictional) dataset of 142 firms listed in the New York Stock Exchange, consisting of the following four variables, all of which are publicly available.

- Return: Total return from holding a firm's stock over a one-year period, from the January 2014 to December 2014. This is the response variable. Note that a return such has 31.4% is entered in the data as 31.4.
- DKR: A firm's debt to capital ratio in 2013.
- LNetIncome: Natural log of the net income for a firm in 2013.
- LSalary: Natural log of the total compensation for a firm's CEO in 2013.

Using these data, you estimated three regressions.

Regression # 1

```
. reg Return DKR ;
```

Source	SS	df	MS
Model	102.905767	1	102.905767
Residual	137036.078	140	978.829128
Total	137138.984	141	972.616906

Number of obs	= 142
F(1, 140)	= 0.11
Prob > F	= 0.7462
R-squared	= 0.0008
Adj R-squared	= -0.0064
Root MSE	= 31.286

Return	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
DKR	.0597546	.1842915	0.32	0.746	-.3045995	.4241088
_cons	4.467659	5.002147	0.89	0.373	-5.421856	14.35717

Regression # 2

```
. reg Return DKR LNetIncome ;
```

Source	SS	df	MS
Model	23917.3208	2	11958.6604
Residual	113221.663	139	814.544338
Total	137138.984	141	972.616906

Number of obs =	142
F(2, 139) =	14.68
Prob > F =	0.0000
R-squared =	0.1744
Adj R-squared =	0.1625
Root MSE =	28.54

Return	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
DKR	.062906	.1681171	0.37	0.709	-.2694913	.3953033
LNetIncome	4.551175	.8417071	5.41	0.000	2.886971	6.21538
_cons	-22.69277	6.786289	-3.34	0.001	-36.11047	-9.275073

Regression # 3

```
. reg Return DKR LNetIncome LSalary ;
```

Source	SS	df	MS
Model	24888.2688	3	8296.0896
Residual	112250.715	138	813.410978
Total	137138.984	141	972.616906

Number of obs =	142
F(3, 138) =	10.20
Prob > F =	0.0000
R-squared =	0.1815
Adj R-squared =	0.1637
Root MSE =	28.52

Return	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
DKR	.0699964	.1681254	0.42	0.678	-.2624385	.4024313
LNetIncome	9.621986	4.716846	2.04	0.043	.29535	18.94862
LSalary	-4.893646	4.479088	-1.09	0.276	-13.75016	3.96287
_cons	-14.27131	10.26663	-1.39	0.167	-34.57155	6.028928

(a) Which of the following statements is true about the correlation between DKR and LNetIncome? Explain your answer.

(i) It is positive.

(ii) It is negative.

(iii) It is zero.

(iv) There is not enough information to determine the sign of the correlation.

(b) Interpret the coefficient on LNetIncome in Regression # 2.

(c) Suppose that you use Regression # 3 to examine whether EMH holds (i.e. whether past information can predict future returns). What are the null and alternative hypotheses?

(d) Carry out the test in part (c) at the 5% level. Do you reject or fail to reject the null hypothesis?

(e) Interpret the result you obtained in part (d), in light of your task of examining the validity of EMH.

(f) Provide 3 reasons that suggest a collinearity problem in Regression # 3.

(g) Which of the following statements is true based on a comparison of Regression # 2 and Regression # 3? Explain your answer.

(i) DKR and LNetIncome are highly correlated.

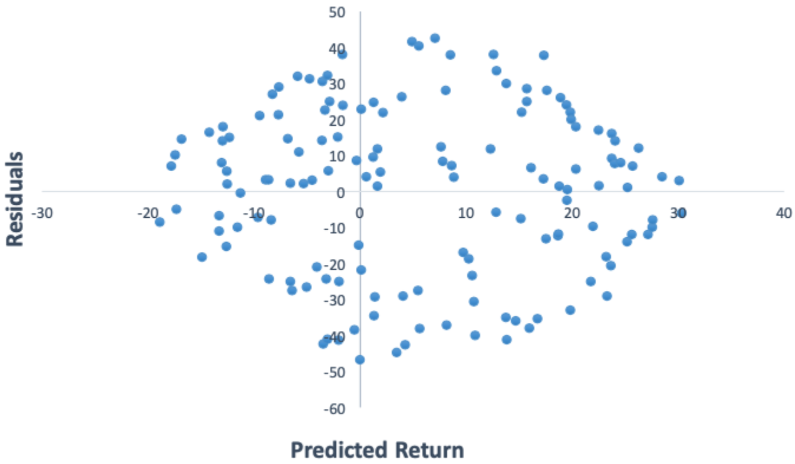
(ii) DKR and LSalary are highly correlated.

(iii) LNetIncome and LSalary are highly correlated.

(iv) All of the above.

(v) None of the above.

(h) Suppose that from Regression # 3, you obtained the following plot of the residuals and the predicted return. Does this graph provide evidence of a violation of the MRM conditions? Explain your answer.



Exercise 2 (Stata) In this exercise, we will replicate the analysis of Sony stock that was discussed during this week's lecture (Lecture 8). The datasets for this exercise are `sony_returns.dta` and `market_returns.dta`.

(a) Before we can implement the analysis, we must put together the above two datasets. Use the `merge` command to combine them into one dataset.

(b) Replicate the figures and the regression in slides 12-14. Do the figures show that the SRM conditions are satisfied?

(c) Replicate the analysis in slides 18-20, and compare the MRM results to the SRM results in part (b). What are the signs of that we have a collinearity problem in the MRM regression?

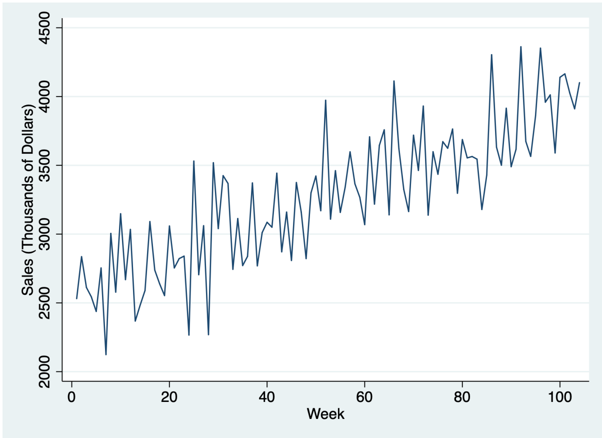
(d) Generate a new variable called `BigSmall` which is equal to `-1*SmallBig`. Then, add `BigSmall` to the MRM in part (c). You should obtain the same results as in slide 23, and Stata should tell you that the variable `BigSmall` was omitted from the regression. Why is this happening?

(e) Generate 25 variables that contain random numbers drawn from the standard normal distribution (the Stata function is `rnormal()`). Use the seed 2019. Using these new variables, replicate the analysis in slides 28-30. Are the regressions good models for predicting Sony returns?

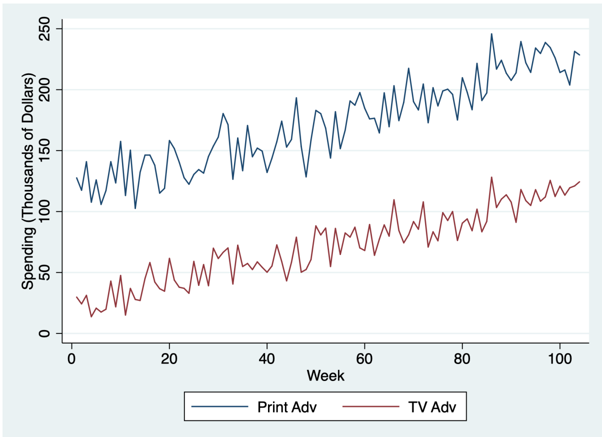
Exercise 3 (Budget Allocation) Collinearity among the predictors (i.e., explanatory variables) is common in many applications, particularly those that track the growth of a new business over time. The problem is worse when the business has steadily grown or fallen. Because the growth of a business affects many attributes of the business (such as assets, sales, number of employees and so forth), the simultaneous changes that take place make it hard to separate important factors from coincidences.

For this exercise, you're the manager who allocates advertising dollars. You have a fixed total budget for advertising, and you have to decide how to spend it. We've simplified things so that you have two choices: print ads or television.

The past two years have been a time of growth for your company, as you can see from the timeplot of weekly sales during this period. The data are in thousands of dollars, so you can see from the plot that weekly sales have grown from about \$2.5 million up to around \$4.5 million.



Other things have grown as well, namely, your expenditures for TV and print advertising. This timeplot shows the two of them over the same 104 weeks.



For this exercise, parts (a) to (c) are conceptual questions that do not require Stata. For the remaining parts, the dataset is `budget_allocation.xlsx`.

Motivation

(a) How are you going to decide how to allocate your budget between these two types of promotion?

Method

(b) Explain how you can use multiple regression to help decide how to allocate the advertising budget between print ads and television ads.

(c) Why would it not be enough to work with several, more easily understood simple regression models, such as sales on spending for television ads or sales on spending on print ads?

(d) Look at the scatterplot sales vs. each of the two explanatory variables (TV Adv and Print Adv). Do the relationships between the variables seem straight enough to fit a multiple regression?

(e) Do you anticipate that collinearity will affect the estimates and standard errors in the multiple regression? Use the correlation matrix of these variables to help construct your answer.

Mechanics Fit the multiple regression of Sales on TV Adv and Print Adv.

(f) Does the model satisfy the assumptions for the use of the MRM?

(g) Assuming that the model satisfies the conditions for the MRM,

  (i) Does the model as a whole explain statistically significant variation in Sales?

  (ii) Does each individual explanatory variable improve the fit of the model, beyond the variation explained by the other alone?

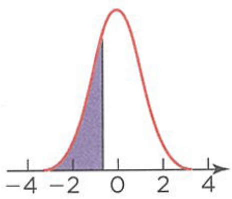
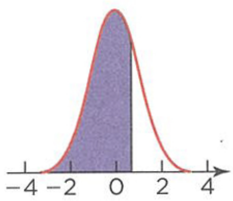
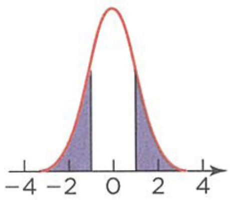
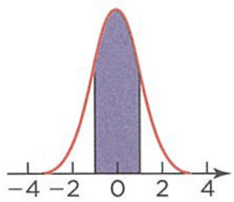
(h) Do the results from the multiple regression suggest a method for allocating your budget? Assume that your budget for the next week is $360,000.

(i) Does the fit of this model promise an accurate prediction of sales in the next week, accurate enough for you to think that you have the right allocation?

Message

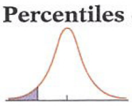
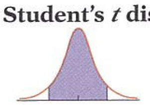
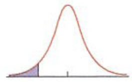
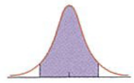
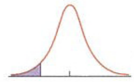
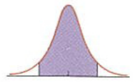
(j) Everyone at the budget meeting knows the information in the plots shown in the introduction to this exercise: Sales and both types of advertising are up. Make a recommendation with enough justification to satisfy their concerns.

(k) Identify any important concerns or limitations that you feel should be understood to appreciate your recommendation.

z	$P(Z \le -z)$	$P(Z \le z)$	$P(\lvert Z\rvert > z)$	$P(\lvert Z\rvert \le z)$
0	0.50	0.50	1	0
0.0502	0.48	0.52	0.96	0.04
0.1004	0.46	0.54	0.92	0.08
0.1510	0.44	0.56	0.88	0.12
0.2019	0.42	0.58	0.84	0.16
0.2533	0.40	0.60	0.80	0.20
0.3055	0.38	0.62	0.76	0.24
0.3585	0.36	0.64	0.72	0.28
0.4125	0.34	0.66	0.68	0.32
0.4677	0.32	0.68	0.64	0.36
0.4959	0.31	0.69	0.62	0.38
0.5244	0.30	0.70	0.60	0.40
0.5828	0.28	0.72	0.56	0.44
0.6433	0.26	0.74	0.52	0.48
0.6745	0.25	0.75	0.50	0.50
0.7063	0.24	0.76	0.48	0.52
0.7388	0.23	0.77	0.46	0.54
0.7722	0.22	0.78	0.44	0.56
0.8064	0.21	0.79	0.42	0.58
0.8416	0.20	0.80	0.40	0.60
0.8779	0.19	0.81	0.38	0.62
0.9154	0.18	0.82	0.36	0.64
0.9542	0.17	0.83	0.34	0.66
0.9945	0.16	0.84	0.32	0.68
1.0364	0.15	0.85	0.30	0.70
1.0803	0.14	0.86	0.28	0.72
1.1264	0.13	0.87	0.26	0.74
1.1750	0.12	0.88	0.24	0.76
1.2265	0.11	0.89	0.22	0.78
1.2816	0.10	0.90	0.20	0.80
1.3408	0.09	0.91	0.18	0.82
1.4051	0.08	0.92	0.16	0.84
1.4758	0.07	0.93	0.14	0.86
1.5548	0.06	0.94	0.12	0.88
1.6449	0.05	0.95	0.10	0.90
1.7507	0.04	0.96	0.08	0.92
1.8808	0.03	0.97	0.06	0.94
1.9600	0.025	0.975	0.05	0.95
2.0537	0.02	0.98	0.04	0.96
2.3263	0.01	0.99	0.02	0.98
2.5758	0.005	0.995	0.01	0.99
2.8070	0.0025	0.9975	0.005	0.995
3.0902	0.001	0.999	0.002	0.998
3.2905	0.0005	0.9995	0.001	0.999
3.7190	0.0001	0.9999	0.0002	0.9998
3.8906	0.00005	0.99995	0.0001	0.9999
4.2649	0.00001	0.99999	0.00002	0.99998
4.4172	0.000005	0.999995	0.00001	0.99999

T-TABLE **Percentiles of Student's *t* distribution.**

$df = 1$			$df = 2$			$df = 3$		
t	$P(T_1 \le -t)$	$P(-t \le T_1 \le t)$	t	$P(T_2 \le -t)$	$P(-t \le T_2 \le t)$	t	$P(T_3 \le -t)$	$P(-t \le T_3 \le t)$
3.078	0.1	0.8	1.886	0.1	0.8	1.638	0.1	0.8
6.314	0.05	0.9	2.920	0.05	0.9	2.353	0.05	0.9
12.71	0.025	0.95	4.303	0.025	0.95	3.182	0.025	0.95
31.82	0.01	0.98	6.965	0.01	0.98	4.541	0.01	0.98
63.66	0.005	0.99	9.925	0.005	0.99	5.841	0.005	0.99
318.3	0.001	0.998	22.33	0.001	0.998	10.21	0.001	0.998
636.6	0.0005	0.999	31.60	0.0005	0.999	12.92	0.0005	0.999
6366	0.00005	0.9999	99.99	0.00005	0.9999	28.00	0.00005	0.9999

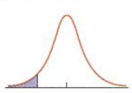
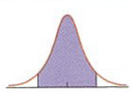
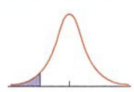
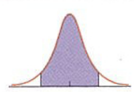
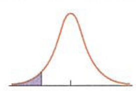
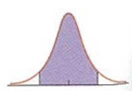
$df = 4$			$df = 5$			$df = 6$		
t	$P(T_4 \le -t)$	$P(-t \le T_4 \le t)$	t	$P(T_5 \le -t)$	$P(-t \le T_5 \le t)$	t	$P(T_6 \le -t)$	$P(-t \le T_6 \le t)$
1.533	0.1	0.8	1.476	0.1	0.8	1.440	0.1	0.8
2.132	0.05	0.9	2.015	0.05	0.9	1.943	0.05	0.9
2.776	0.025	0.95	2.571	0.025	0.95	2.447	0.025	0.95
3.747	0.01	0.98	3.365	0.01	0.98	3.143	0.01	0.98
4.604	0.005	0.99	4.032	0.005	0.99	3.707	0.005	0.99
7.173	0.001	0.998	5.893	0.001	0.998	5.208	0.001	0.998
8.610	0.0005	0.999	6.869	0.0005	0.999	5.959	0.0005	0.999
15.54	0.00005	0.9999	11.18	0.00005	0.9999	9.082	0.00005	0.9999

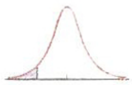
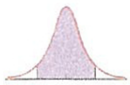
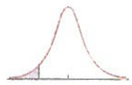
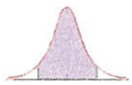
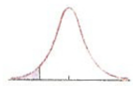
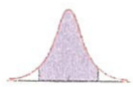
$df = 7$			$df = 8$			$df = 9$		
t	$P(T_7 \le -t)$	$P(-t \le T_7 \le t)$	t	$P(T_8 \le -t)$	$P(-t \le T_8 \le t)$	t	$P(T_9 \le -t)$	$P(-t \le T_9 \le t)$
1.415	0.1	0.8	1.397	0.1	0.8	1.383	0.1	0.8
1.895	0.05	0.9	1.860	0.05	0.9	1.833	0.05	0.9
2.365	0.025	0.95	2.306	0.025	0.95	2.262	0.025	0.95
2.998	0.01	0.98	2.896	0.01	0.98	2.821	0.01	0.98
3.499	0.005	0.99	3.355	0.005	0.99	3.250	0.005	0.99
4.785	0.001	0.998	4.501	0.001	0.998	4.297	0.001	0.998
5.408	0.0005	0.999	5.041	0.0005	0.999	4.781	0.0005	0.999
7.885	0.00005	0.9999	7.120	0.00005	0.9999	6.594	0.00005	0.9999

$df = 10$			$df = 11$			$df = 12$		
t	$P(T_{10} \le -t)$	$P(-t \le T_{10} \le t)$	t	$P(T_{11} \le -t)$	$P(-t \le T_{11} \le t)$	t	$P(T_{12} \le -t)$	$P(-t \le T_{12} \le t)$
1.415	0.1	0.8	1.397	0.1	0.8	1.383	0.1	0.8
1.895	0.05	0.9	1.860	0.05	0.9	1.833	0.05	0.9
2.365	0.025	0.95	2.306	0.025	0.95	2.262	0.025	0.95
2.998	0.01	0.98	2.896	0.01	0.98	2.821	0.01	0.98
3.499	0.005	0.99	3.355	0.005	0.99	3.250	0.005	0.99
4.785	0.001	0.998	4.501	0.001	0.998	4.297	0.001	0.998
5.408	0.0005	0.999	5.041	0.0005	0.999	4.781	0.0005	0.999
7.885	0.00005	0.9999	7.120	0.00005	0.9999	6.594	0.00005	0.9999

$df = 13$			$df = 14$			$df = 15$		
t	$P(T_{13} \le -t)$	$P(-t \le T_{13} \le t)$	t	$P(T_{14} \le -t)$	$P(-t \le T_{14} \le t)$	t	$P(T_{15} \le -t)$	$P(-t \le T_{15} \le t)$
1.350	0.1	0.8	1.345	0.1	0.8	1.341	0.1	0.8
1.771	0.05	0.9	1.761	0.05	0.9	1.753	0.05	0.9
2.160	0.025	0.95	2.145	0.025	0.95	2.131	0.025	0.95
2.650	0.01	0.98	2.624	0.01	0.98	2.602	0.01	0.98
3.012	0.005	0.99	2.977	0.005	0.99	2.947	0.005	0.99
3.852	0.001	0.998	3.787	0.001	0.998	3.733	0.001	0.998
4.221	0.0005	0.999	4.140	0.0005	0.999	4.073	0.0005	0.999
5.513	0.00005	0.9999	5.363	0.00005	0.9999	5.239	0.00005	0.9999

$df = 16$			$df = 17$			$df = 18$		
t	$P(T_{16} \le -t)$	$P(-t \le T_{16} \le t)$	t	$P(T_{17} \le -t)$	$P(-t \le T_{17} \le t)$	t	$P(T_{18} \le -t)$	$P(-t \le T_{18} \le t)$
1.337	0.1	0.8	1.333	0.1	0.8	1.33	0.1	0.8
1.746	0.05	0.9	1.740	0.05	0.9	1.734	0.05	0.9
2.120	0.025	0.95	2.110	0.025	0.95	2.101	0.025	0.95
2.583	0.01	0.98	2.567	0.01	0.98	2.552	0.01	0.98
2.921	0.005	0.99	2.898	0.005	0.99	2.878	0.005	0.99
3.686	0.001	0.998	3.646	0.001	0.998	3.610	0.001	0.998
4.015	0.0005	0.999	3.965	0.0005	0.999	3.922	0.0005	0.999
5.134	0.00005	0.9999	5.044	0.00005	0.9999	4.966	0.00005	0.9999

$df = 19$

t	$P(T_{19} \leq -t)$	$P(-t \leq T_{19} \leq t)$
1.328	0.1	0.8
1.729	0.05	0.9
2.093	0.025	0.95
2.539	0.01	0.98
2.861	0.005	0.99
3.579	0.001	0.998
3.883	0.0005	0.999
4.897	0.00005	0.9999

$df = 20$

t	$P(T_{20} \leq -t)$	$P(-t \leq T_{20} \leq t)$
1.325	0.1	0.8
1.725	0.05	0.9
2.086	0.025	0.95
2.528	0.01	0.98
2.845	0.005	0.99
3.552	0.001	0.998
3.850	0.0005	0.999
4.837	0.00005	0.9999

$df = 22$

t	$P(T_{22} \leq -t)$	$P(-t \leq T_{22} \leq t)$
1.321	0.1	0.8
1.717	0.05	0.9
2.074	0.025	0.95
2.508	0.01	0.98
2.819	0.005	0.99
3.505	0.001	0.998
3.792	0.0005	0.999
4.736	0.00005	0.9999

$df = 24$

t	$P(T_{24} \leq -t)$	$P(-t \leq T_{24} \leq t)$
1.318	0.1	0.8
1.711	0.05	0.9
2.064	0.025	0.95
2.492	0.01	0.98
2.797	0.005	0.99
3.467	0.001	0.998
3.745	0.0005	0.999
4.654	0.00005	0.9999

$df = 26$

t	$P(T_{26} \leq -t)$	$P(-t \leq T_{26} \leq t)$
1.315	0.1	0.8
1.706	0.05	0.9
2.056	0.025	0.95
2.479	0.01	0.98
2.779	0.005	0.99
3.435	0.001	0.998
3.707	0.0005	0.999
4.587	0.00005	0.9999

$df = 28$

t	$P(T_{28} \leq -t)$	$P(-t \leq T_{28} \leq t)$
1.313	0.1	0.8
1.701	0.05	0.9
2.048	0.025	0.95
2.467	0.01	0.98
2.763	0.005	0.99
3.408	0.001	0.998
3.674	0.0005	0.999
4.530	0.00005	0.9999

$df = 30$

t	$P(T_{30} \leq -t)$	$P(-t \leq T_{30} \leq t)$
1.31	0.1	0.8
1.697	0.05	0.9
2.042	0.025	0.95
2.457	0.01	0.98
2.75	0.005	0.99
3.385	0.001	0.998
3.646	0.0005	0.999
4.482	0.00005	0.9999

$df = 32$

t	$P(T_{32} \leq -t)$	$P(-t \leq T_{32} \leq t)$
1.309	0.1	0.8
1.694	0.05	0.9
2.037	0.025	0.95
2.449	0.01	0.98
2.738	0.005	0.99
3.365	0.001	0.998
3.622	0.0005	0.999
4.441	0.00005	0.9999

$df = 34$

t	$P(T_{34} \leq -t)$	$P(-t \leq T_{34} \leq t)$
1.307	0.1	0.8
1.691	0.05	0.9
2.032	0.025	0.95
2.441	0.01	0.98
2.728	0.005	0.99
3.348	0.001	0.998
3.601	0.0005	0.999
4.405	0.00005	0.9999

$df = 36$

t	$P(T_{36} \leq -t)$	$P(-t \leq T_{36} \leq t)$
1.306	0.1	0.8
1.688	0.05	0.9
2.028	0.025	0.95
2.434	0.01	0.98
2.719	0.005	0.99
3.333	0.001	0.998
3.582	0.0005	0.999
4.374	0.00005	0.9999

$df = 40$

t	$P(T_{40} \leq -t)$	$P(-t \leq T_{40} \leq t)$
1.303	0.1	0.8
1.684	0.05	0.9
2.021	0.025	0.95
2.423	0.01	0.98
2.704	0.005	0.99
3.307	0.001	0.998
3.551	0.0005	0.999
4.321	0.00005	0.9999

$df = 50$

t	$P(T_{50} \leq -t)$	$P(-t \leq T_{50} \leq t)$
1.299	0.1	0.8
1.676	0.05	0.9
2.009	0.025	0.95
2.403	0.01	0.98
2.678	0.005	0.99
3.261	0.001	0.998
3.496	0.0005	0.999
4.228	0.00005	0.9999

$df = 60$

t	$P(T_{60} \leq -t)$	$P(-t \leq T_{60} \leq t)$
1.296	0.1	0.8
1.671	0.05	0.9
2.000	0.025	0.95
2.390	0.01	0.98
2.660	0.005	0.99
3.232	0.001	0.998
3.460	0.0005	0.999
4.169	0.00005	0.9999

$df = 75$

t	$P(T_{75} \leq -t)$	$P(-t \leq T_{75} \leq t)$
1.293	0.1	0.8
1.665	0.05	0.9
1.992	0.025	0.95
2.377	0.01	0.98
2.643	0.005	0.99
3.202	0.001	0.998
3.425	0.0005	0.999
4.110	0.00005	0.9999

$df = 100$

t	$P(T_{100} \leq -t)$	$P(-t \leq T_{100} \leq t)$
1.290	0.1	0.8
1.660	0.05	0.9
1.984	0.025	0.95
2.364	0.01	0.98
2.626	0.005	0.99
3.174	0.001	0.998
3.390	0.0005	0.999
4.053	0.00005	0.9999

$df = 125$

t	$P(T_{125} \leq -t)$	$P(-t \leq T_{125} \leq t)$
1.288	0.1	0.8
1.657	0.05	0.9
1.979	0.025	0.95
2.357	0.01	0.98
2.616	0.005	0.99
3.157	0.001	0.998
3.370	0.0005	0.999
4.020	0.00005	0.9999

$df = 150$

t	$P(T_{150} \leq -t)$	$P(-t \leq T_{150} \leq t)$
1.287	0.1	0.8
1.655	0.05	0.9
1.976	0.025	0.95
2.351	0.01	0.98
2.609	0.005	0.99
3.145	0.001	0.998
3.357	0.0005	0.999
3.998	0.00005	0.9999

$df = \infty$

t	$P(Z \leq -t)$	$P(-t \leq Z \leq t)$
1.282	0.1	0.8
1.645	0.05	0.9
1.960	0.025	0.95
2.326	0.01	0.98
2.576	0.005	0.99
3.090	0.001	0.998
3.291	0.0005	0.999
3.891	0.00005	0.9999