# Income and Democracy

Bachelor's/Master's Thesis submitted

to

**Prof. Dr. Nikolaus Hautsch**

Humboldt-Universität zu Berlin
School of Business and Economics
Institute for Statistics and Econometrics
Chair of Econometrics

by

**your name**
(your matriculation number)

in partial fulfillment of the requirements
for the degree of
**Bachelor/Master of Science**
Berlin, September 30, 2007

## Abstract

This is the template for a thesis at the Chair of Econometrics of Humboldt–Universität zu Berlin. A popular approach to write a thesis or a paper is the IMRAD method (Introduction, Methods, Results and Discussion). This approach is not mandatory! You can find more information about formal requirements in the booklet 'Hinweise zur Gestaltung der äußeren Form von Diplomarbeiten' which is available in the office of studies.

The abstract should not be longer than a paragraph of around 10 to 15 lines (or about 150 words). The abstract should contain a concise description of the econometric/economic problem you analyse and of your results. This allows the busy reader to obtain quickly a clear idea of the thesis content.

# Contents

## List of Abbreviations

| | | | |
|---|---|---|---|
| CPI | Consumer Price Index | ETF | Equity Traded Funds |
| ETH | Eat the Horse | XLM | Xetra Liquidity |

# List of Figures

# List of Tables

# 1 Introduction: Sharon, Soichi

- What is the subject of the study? Describe the economic/econometric problem.
- What is the purpose of the study (working hypothesis)?
- What do we already know about the subject (literature review)? Use citations: *Gallant (1987) shows that... Alternative Forms of the Wald test are considered (Breusch and Schmidt, 1988).*
- What is the innovation of the study?
- Provide an overview of your results.
- Outline of the paper:
  *The paper is organized as follows. The next section describes the model under investigation. Section* **??** *describes the data set and Section* **??** *presents the results. Finally, Section* **??** *concludes.*
- The introduction should not be longer than 4 pages.

# 2 Statistical Theory: Hyerin

Describe the statistical theory that you implement and the corresponding algorithm(s) that was (were) designed. Avoid describing how the design was produced, rather, show different versions of the design if relevant.

- Panel Data analysis
- Fixed Effects
- Clustering by group
- Instrumental Variable

# 3 Implementation

Show the implementation. The goal of this section is to show and explain the most important parts of the code. Listing the code with highlighting and possibly line numbering is essential. Explain the code by referring to line numbers, function calls and variable names. Leave out trivial parts (initialization, parameter-tuning, etc...).

- PLM
- CLSE
- Stargazer
- Figure

# 4 Empirical Study

## 4.1 Data description: Soichi

- Describe the data and its quality.
- How was the data sample selected?
- Provide descriptive statistics such as:
  - time period,
  - number of observations, data frequency,
  - mean, median,
  - min, max, standard deviation,
  - skewness, kurtosis, Jarque–Bera statistic,
  - time series plots, histogram.
- Allows the reader to judge whether the sample is biased or to evaluate possible impacts of outliers, for example.

## 4.2 Regression Specification: Hyerin

In this section, we discuss the causal effect of income on democracy. The relation between democracy scores and income per capita is estimated using the following simple econometric model:

$$d_{i,t} = \alpha d_{i,t-1} + \gamma y_{i,t-1} + X'_{i,t-1}\beta + \mu_t + \delta_i + u_{i,t} \tag{1}$$

The democracy score of coutry $i$ in time period $t$ is denoted by $d_{i,t}$. To capture persistency of it, the lagged value of this variable, $d_{i,t}$, is included as a regressor in the model. $y_{i,t-1}$ denotes the lagged value of country $i$'s log income per capita and the estimated coefficient, $\gamma$, measures the causal effect of income per capita on democracy. If an increase in income per capita leads to an higher score of democracy, then the coefficient on the lagged value of log income per capita $\gamma$ should be positive. A full set of coutry dummies and time effects $\delta_i$ $\mu_t$. All other potential covariates are denoted in the vector formation, $X'_{i,t-1}$. To control for country-specific factors a full set of country dummies, $\delta_i$, is included in the right-hand side. In addition, to capture common shocks over all sample countries, it introduces $\mu_t$ as a full set of time effects. An error term is denoted as $u_{i,t}$.

## 4.3 Regression Estimates

### 4.3.1 Fixed Effects: Soichi

**Table 1:** FIXED EFFECTS RESULTS USING FREEDOM HOUSE MEASURE OF DEMOCRACY

| | Base sample, 1960-2000 | | | | |
|---|---|---|---|---|---|
| | Five-year data | | | Ten-year data | Twenty-year data |
| | Pooled OLS (1) | Fixed effects OLS (2) | Fixed effects OLS (3) | Fixed effects OLS (4) | Fixed effects OLS (5) |
| | *Dependent variable is democracy* | | | | |
| Democracy$_{t-1}$ | 0.706*** (0.035) | 0.379*** (0.051) | | -0.025 (0.088) | -0.581*** (0.198) |
| Log GDP per capita$_{t-1}$ | 0.072*** (0.010) | 0.010 (0.035) | 0.054 (0.046) | 0.053 (0.066) | -0.030 (0.156) |
| Observations | 945 | 945 | 958 | 457 | 192 |
| $R^2$ | 0.725 | 0.242 | 0.118 | 0.122 | 0.452 |

*Notes:*

- Organize material and present results.
- Use tables, figures (but prefer visual presentation):
  - Tables and figures should supplement (and not duplicate) the text.
  - Tables and figures should be provided with legends.

    *Figure 1 shows how to include and reference graphics. The graphic must be labelled before. Files must be in* `.eps` *format.*

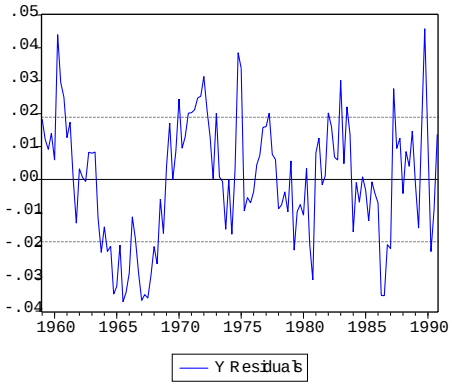


**Figure 1:** Estimated residuals from model XXX. ...

- Tables and graphics may appear in the text or in the appendix, especially if there are many simulation result tabulated, but is also depends on the study and number of tables resp.figures. The key graphs and tables must appear in the text!

### 4.3.2 Instrumental Variable: Hyerin

Since the fixed effects estimation does not necessarily examine the causal effect of income on democracy, we need a instrumental variable to estimate the causal relation betweem them.

**Table 2:** FIXED EFFECTS RESULTS USING FREEDOM HOUSE MEASURE OF DEMOCRACY: TWO-STAGE LEAST SQUARES WITH SAVINGS RATE INSTRUMENT

| | Base sample, 1960-2000 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All countries | | | | | | | |
| | Pooled OLS (1) | Fixed effects OLS (2) | Fixed effects OLS (3) | Fixed effects 2SLS (4) | Fixed effects 2SLS (5) | Fixed effects 2SLS (6) | Fixed effects 2SLS (7) | Fixed effects 2SLS (8) |
| Panel A | *Dependent variable is democracy* | | | | | | | |
| $\text{Democracy}_{t-1}$ | | | 0.359 (0.054) | | 0.363 (0.056) | | | |
| $\text{Log GDP per capita}_{t-1}$ | 0.233*** (0.013) | 0.044 (0.051) | 0.009 (0.038) | -0.035 (0.094) | -0.020 (0.081) | -0.036 (0.191) | -0.074 (0.113) | 0.016 (0.095) |
| $\text{Labor share}_{t-1}$ | | | | | | 0.250 (0.199) | | |
| Panel B | *First stage for log GDP per capita*$_{t-1}$ | | | | | | | |
| $\text{Democracy}_{t-1}$ | | | | | 0.144** (0.066) | | [0.24] | |
| $\text{Labor share}_{t-1}$ | | | | | | 0.329* (0.187) | | |
| $\text{Savings rate}_{t-2}$ | | | | 1.356*** (0.277) | 1.343*** (0.270) | 1.202*** (0.315) | 1.173*** (0.254) | 1.022*** (0.218) |
| $\text{Savings rate}_{t-3}$ | | | | | | | | 0.720*** (0.182) |
| Observations | 891 | 900 | 766 | 900 | 891 | 471 | 733 | 796 |
| $R^2$ | 0.226 | 0.115 | 0.225 | 0.571 | 0.571 | 0.725 | 0.541 | 0.536 |

*Notes:* This table summarizes the codefficients of each cross-sectional regression. All regression model includes year dummies to capture country-invariant factors. Except for column 1, country dummies are included in the regressions. The robust standrad errors clustered by country are summarized in parentheses.

***Significant at the 1 percent level

**Significant at the 5 percent level

*Significant at the 10 percent level

### 4.4 Robustness Tests: Sharon

**Table 3:** Fixed effects results with alternative samples and additional control variables

| | Five-year data | | | |
|---|---|---|---|---|
| | Balanced panel 1970-2000 | Base sample, 1960-2000, without former socialist countries | Base sample, 1960-2000 | |
| | Fixed effects OLS (1) | Fixed effects OLS (2) | Fixed effects OLS (3) | Fixed effects OLS (4) |
| | *Dependent variable is democracy* | | | |
| Democracy$_{t-1}$ | 0.283*** | 0.362*** | 0.353*** | 0.351*** |
| | (0.058) | (0.052) | (0.053) | (0.055) |
| Log GDP per capita$_{t-1}$ | -0.031 | 0.005 | 0.015 | -0.001 |
| | (0.049) | (0.035) | (0.041) | (0.049) |
| Log population$_{t-1}$ | | | -0.109 | -0.042 |
| | | | (0.100) | (0.108) |
| Education$_{t-1}$ | | | | -0.007 |
| | | | | (0.020) |
| Age Structure$_{t-1}$ | | | [0.05] | [0.19] |
| Observations | 630 | 908 | 863 | 676 |
| Countries | 90 | 128 | 142 | 95 |
| $R^2$ | 0.215 | 0.221 | 0.241 | 0.235 |

*Notes:*

Discuss results:

- Do the results support or do they contradict economic theory ?
- What does the reader learn from the results?
- Try to give an intuition for your results.
- Provide robustness checks.
- Compare to previous research.

## 5 Conclusions

- Give a short summary of what has been done and what has been found.
- Expose results concisely.
- Draw conclusions about the problem studied. What are the implications of your findings?
- Point out some limitations of study (assist reader in judging validity of findings).

**Table 4:** FIXED EFFECTS RESULTS USING AN ALTERNATIVE DEPENDENT VARIABLE: POLITY MEASURE

| | Base sample, 1960-2000 | | | | |
|---|---|---|---|---|---|
| | Five-year data | | | Ten-year data | Twenty-year data |
| | Pooled OLS (1) | Fixed effects OLS (2) | Fixed effects OLS (3) | Fixed effects OLS (4) | Fixed effects OLS (5) |
| Democracy$_{t-1}$ | 0.749*** | 0.449*** | | 0.060 | -0.516*** |
| | (0.034) | (0.063) | | (0.091) | (0.165) |
| Log GDP per capita$_{t-1}$ | 0.053*** | -0.006 | -0.011 | 0.007 | -0.126 |
| | (0.010) | (0.039) | (0.055) | (0.070) | (0.164) |
| Observations | 854 | 854 | 880 | 419 | 168 |
| $R^2$ | 0.772 | 0.396 | 0.248 | 0.257 | 0.544 |

*Notes:*

- Suggest issues for future research.

## References

Breusch, T. S. and P. Schmidt (1988): "Alternative Forms of the Wald test: How Long is a Piece of String," *Communications in Statistics, Theory and Methods*, 17, 2789–2795.

Gallant, A. R. (1987): *Nonlinear Statistical Models*, New York: John Wiley & Sons.

# A Source Code Listing

```
require("sandwich")
require("lmtest")
require("stargazer")



clse = function(reg) {
  # index(reg, "id") returns the id or entity variable vector
  G = length(unique(index(reg,"id")))
  N = length(index(reg,"id"))
  dfa = (G/(G - 1))*(N-1)/reg$df.residual   # note Bluhm multiplies this by
      finite-sample df adjustment
  rob = sqrt(diag(dfa*vcovHC(reg, type = "HC0",
                             cluster = "group", adjust = T)))
  return(rob)
}

stargazer(F_pols, F_fols, F_folswod,se=list(clse(F_pols),clse(F_fols),clse(F_
    folswod)), title="Panel regression, clustered SEs", type="text", column.
    labels=c("Pooled OLS", "Fixed Effects OLS", "Fixed Effects OLS"), df=FALSE,
     digits=3)
```

```
#rm(list=ls())
getwd()
#setwd()
setwd("C:/myR/Dem_data")


#Import data set
One=read.csv("Annual_panel.csv")
Five=read.csv("5yr_panel.csv")
Ten=read.csv("10yr_panel.csv")
Twenty=read.csv("20yr_panel.csv")

require(plm)
require(lmtest)
require(tseries)
require(car)
#Base sample with Five-year data
#T2. C1
```

```
F_pols = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+factor(year), data =
    Five, subset = sample == "1", index = c("code","year"), model = "pooling")
summary(F_pols)


#Test for multicollinearity by using pooled data
#VIF(Variance inflation factior)
vif(F_pols)

#Breusch-Godfrey test
pbgtest(F_pols)

#Breusch-Pagan test for homoskedastisity
plmtest(F_pols, type="bp")


#T2. C2
F_fols = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+factor(country)+factor
    (year), data = Five, subset = sample == "1",index = c("country","year"),
    model = "within", effect = "individual")
summary(F_fols)

#Breusch-Godfrey test
pbgtest(F_fols)

#Breusch-Pagan test for homoskedastisity
plmtest(F_fols, type="bp")


#T2. C5
F_folswod = plm(fhpolrigaug ~ lag(lrgdpch)+factor(year)+factor(country), data =
     Five, subset = sample =="1", index = c("code","year"), model = "within",
    effect = "individual")
summary(F_folswod)

#Breusch-Godfrey test
pbgtest(F_folswod)

#Breusch-Pagan test for homoskedastisity
plmtest(F_folswod, type="bp")

```

```
55
56 #T2. C6
57 O_fols = plm(fhpolrigaug ~ lag(fhpolrigaug, 1:5)+lag(lrgdpch, 1:5)+factor(
       country)+factor(year), data = One, subset = sample == "1",index = c("
       country","year"), model = "within", effect = "individual")
58 summary(O_fols)
59
60 #Breusch-Godfrey test
61 pbgtest(O_fols)
62
63 #Breusch-Pagan test for homoskedastisity
64 plmtest(O_fols, type="bp")
65
66
67
68 #T2. C7
69 T_fols = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+factor(country)+factor
       (year), data = Ten, subset = sample == "1",index = c("country","year"),
       model = "within", effect = "individual")
70 summary(T_fols)
71
72 #Breusch-Godfrey test
73 pbgtest(T_fols)
74
75 #Breusch-Pagan test for homoskedastisity
76 plmtest(T_fols, type="bp")
77
78
79 #T2. C9
80 Tw_fols = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+factor(country)+
       factor(year), data = Twenty, subset = sample == "1",index = c("country","
       year"), model = "within", effect = "individual")
81 summary(Tw_fols)
82
83 #Breusch-Godfrey test
84 pbgtest(Tw_fols)
85
86 #Breusch-Pagan test for homoskedastisity
87 plmtest(Tw_fols, type="bp")
```

```
1 #T3. C1
2 Pooled_pols_p = plm(polity4 ~ lag(polity4)+lag(lrgdpch)+factor(year), data =
```

```
    Five, subset = sample == "1", index = c("code","year"), model = "pooling")
summary(Pooled_pols_p)

#Test for multicollinearity by using pooled data
#VIF(Variance inflation factior)
vif(Pooled_pols_p)

#Breusch-Godfrey test
pbgtest(Pooled_pols_p)

#Breusch-Pagan test for homoskedastisity
plmtest(Pooled_pols_p, type="bp")


#T3. C2
FE_fols_p_2 = plm(polity4 ~ lag(polity4)+lag(lrgdpch)+factor(country)+factor(
    year), data = Five, subset = sample == "1",index = c("country","year"),
    model = "within", effect = "individual")
summary(FE_fols_p_2)

#Breusch-Godfrey test
pbgtest(FE_fols_p_2)

#Breusch-Pagan test for homoskedastisity
plmtest(FE_fols_p_2, type="bp")


#T3. C5
FE_folswod_p = plm(polity4 ~ lag(lrgdpch)+factor(year)+factor(country), data =
    Five, subset = sample =="1", index = c("code","year"), model = "within",
    effect = "individual")
summary(FE_folswod_p)

#Breusch-Godfrey test
pbgtest(FE_folswod_p)

#Breusch-Pagan test for homoskedastisity
plmtest(FE_folswod_p, type="bp")


#T3. C6
```

```
39 FE_fols_p_6 =  plm(polity4 ~ lag(polity4, 1:5)+lag(lrgdpch, 1:5)+factor(country
       )+factor(year), data = One, subset = sample == "1",index = c("country","
       year"), model = "within", effect = "individual")
40 summary(FE_fols_p_6)
41
42 #Breusch-Godfrey test
43 pbgtest(FE_fols_p_6)
44
45 #Breusch-Pagan test for homoskedastisity
46 plmtest(FE_fols_p_6, type="bp")
47
48
49
50 #T2. C7
51 FE_fols_p_7 = plm(polity4 ~ lag(polity4)+lag(lrgdpch)+factor(country)+factor(
       year), data = Ten, subset = sample == "1",index = c("country","year"),
       model = "within", effect = "individual")
52 summary(FE_fols_p_7)
53
54 #Breusch-Godfrey test
55 pbgtest(FE_fols_p_7)
56
57 #Breusch-Pagan test for homoskedastisity
58 plmtest(FE_fols_p_7, type="bp")
59
60
61
62 #T2. C9
63 FE_fols_p_9 = plm(polity4 ~ lag(polity4)+lag(lrgdpch)+factor(country)+factor(
       year), data = Twenty, subset = sample == "1",index = c("country","year"),
       model = "within", effect = "individual")
64 summary(FE_fols_p_9)
65
66 #Breusch-Godfrey test
67 pbgtest(FE_fols_p_9)
68
69 #Breusch-Pagan test for homoskedastisity
70 plmtest(FE_fols_p_9, type="bp")
71
72
73 library(stargazer)
```

```
74 stargazer(Pooled_pols_p, FE_fols_p_2, FE_folswod_p, FE_fols_p_7, FE_fols_p_9,
       type="html", style = "aer",
75             se=list(clse(Pooled_pols_p),clse(FE_fols_p_2),clse(FE_folswod_p),
                   clse(FE_fols_p_7),clse(FE_fols_p_9)),
76             column.labels = c("Five-year data", "Ten-year data", "Twenty-year
                   data"), column.separate = c(3,1,1),
77             no.space = TRUE, dep.var.labels="Dependent variable is democracy",
                   omit = c("factor", "Constant"), omit.stat = c("f"),
78             covariate.labels=c("Democracy_{t-1}", "Log GDP per capita_{t-1}"),
79             object.names = TRUE,
80             title="Table 3-Fixed Effects Results Using Polity Measure of
                   Democracy", align=TRUE, out = "Table 3.htm")

 1 #T4.C1
 2 F_fols_b = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+factor(year)+factor(
       country), data = Five, subset = samplebalancefe == "1", index = c("code","
       year"), model = "within", effect = "individual")
 3 summary(F_fols_b)
 4
 5 #Breusch-Godfrey test
 6 pbgtest(F_fols_b)
 7
 8 #Breusch-Pagan test for homoskedastisity
 9 plmtest(F_fols_b, type="bp")
10
11
12 #T4.C3
13 F_fols_c = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+factor(year)+factor(
       country), data = Five, subset= sample == "1" & socialist !="1", index = c("
       code","year"), model = "within", effect = "individual")
14 summary(F_fols_c)
15
16 #Breusch-Godfrey test
17 pbgtest(F_fols_c)
18
19 #Breusch-Pagan test for homoskedastisity
20 plmtest(F_fols_c, type="bp")
21
22
23 #T4.C5
24 #reg fhpolrigaug L.fhpolrigaug L.lrgdpch L.lpop L.medage L.age_veryyoung L.age_
       young L.age_midage L.age_old  yr* cd* if sample==1, cluster(code)
```

```
25
26 F_fols_add = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+ lag(lpop)+lag(
       medage)+ lag(age_veryyoung)+ lag(age_young)+lag(age_midage)+lag(age_old)+
       factor(year)+factor(country), data = Five, subset = sample == "1", index =
       c("code", "year"), model="within", effect="individual")
27 summary(F_fols_add)
28
29 #Breusch-Godfrey test
30 pbgtest(F_fols_add)
31
32 #Breusch-Pagan test for homoskedastisity
33 plmtest(F_fols_add, type="bp")
34
35
36 #T4.C7
37 #  reg fhpolrigaug L.fhpolrigaugL.education L.lrgdpch L.lpop L.medage L.age_
       veryyoung L.age_young L.age_midage L.age_old  yr* cd* if sample==1, cluster
       (code)
38 F_fols_we = plm(fhpolrigaug ~ lag(fhpolrigaug)+lag(lrgdpch)+ lag(lpop)+lag(
       education)+lag(medage)+ lag(age_veryyoung)+ lag(age_young)+lag(age_midage)+
       lag(age_old)+factor(year)+factor(country), data = Five, subset = sample ==
       "1", index = c("code", "year"), model="within", effect="individual")
39 summary(F_fols_we)
40
41 #Breusch-Godfrey test
42 pbgtest(F_fols_we)
43
44 #Breusch-Pagan test for homoskedastisity
45 plmtest(F_fols_we, type="bp")
46
47
48
49 library(stargazer)
50 stargazer(F_fols_b, F_fols_c, F_fols_add, F_fols_we, type="html", style = "aer"
       ,
51           se=list(clse(F_fols_b),clse(F_fols_c),clse(F_fols_add), clse(F_fols_
                 we)),
52           column.labels = c("Balanced panel, 1970-2000", "Base sample,
                 1960-2000, without former socialist countries", "Base sample,
                 1960-2000"), column.separate = c(1,1,2),
53           no.space = TRUE, dep.var.labels="Dependent variable is democracy
```

```
    using five-year data and fixed effects OLS", omit = c("factor", "
    Constant", "age"), omit.stat = c("f"),
54  covariate.labels=c("Democracy_{t-1}", "Log GDP per capita_{t-1}", "
    Log population_{t-1}", "Education_{t-1}"),
55  title="Table 4-Fixed Effects Results Using Freedom House Measure of
    Democracy: Robustness Checks", align=TRUE, out = "Table 4.htm")
```

# B Figures

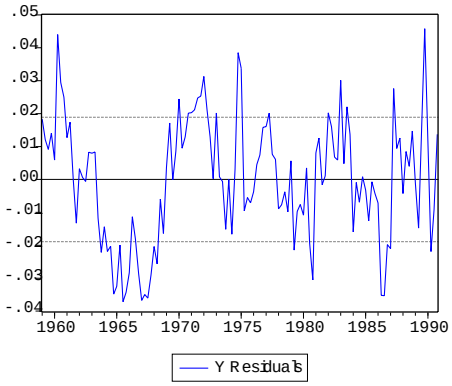


**Figure 2:** Estimated residuals (2) from model XXX. ...

# C Tables

| | 3m | 6m | 1yr | 2yr | 3yr | 5yr | 7yr | 10yr | 12yr | 15yr |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 3.138 | 3.191 | 3.307 | 3.544 | 3.756 | 4.093 | 4.354 | 4.621 | 4.741 | 4.878 |
| Median | 3.013 | 3.109 | 3.228 | 3.490 | 3.680 | 3.906 | 4.117 | 4.420 | 4.575 | 4.759 |
| Min | 1.984 | 1.950 | 1.956 | 2.010 | 2.240 | 2.615 | 2.850 | 3.120 | 3.250 | 3.395 |
| Max | 5.211 | 5.274 | 5.415 | 5.583 | 5.698 | 5.805 | 5.900 | 6.031 | 6.150 | 6.295 |
| StD | 0.915 | 0.919 | 0.935 | 0.910 | 0.876 | 0.825 | 0.803 | 0.776 | 0.768 | 0.762 |

**Table 5:** Detailed descriptive statistics of location and dispersion for 2100 observed swap rates for the period from February 15, 1999 to March 2, 2007. Swap rates measured as 3.12 (instead of 0.0312).