Christian-Albrechts-Universität zu Kiel
Wirtschafts- und Sozialwissenschaftliche Fakultät

# Applied Econometrics of FX Markets

Stefan Reitz, QBER University of Kiel

# Aims of the Course (Content)

Enable students to conduct their own empirical research!

- Learn how to produce estimation results in empirical work
- Learn how to interpret estimation results
- Learn to avoid pitfalls in empirical work
- Learn the programming language R

# Tutorial

- Introduction to statistical programming language
- Exercises in R
- Implementation of econometric models from the lecture
- Registration required in OLAT
- You have to bring your own notebook
- Please bring also multi-plugs/extension cables

# Organization

- Lecture: Mo, 2.15 - 3.45pm **and** Th, 8.15 - 9.45am
- Lecturer: Stefan Reitz (stefan.reitz@qber.uni-kiel.de)
- Tutorial: Mo, 2.15 - 3.45pm **or** Th, 8.15 - 9.45am starting May 26th
- Tutor: Jannis Poggensee (jannis.poggensee@qber.uni-kiel.de)
- 6 credits for passing written exam

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: TAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

# Linear Regression

- In regressions we try to identify the dependence of a variable y from an exogenous variable x.

# Linear Regression

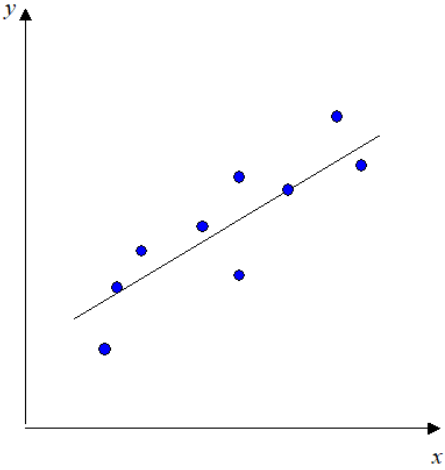
- In regressions we try to identify the dependence of a variable y from an exogenous variable x.
- Different from the concept of correlation!

# Linear Regression

- In regressions we try to identify the dependence of a variable y from an exogenous variable x.
- Different from the concept of correlation!
- The y variable is assumed to be random or "stochastic" in some way, i.e. to have a probability distribution. The x variables are, however, assumed to have fixed ("non- stochastic") values in repeated samples, or at least to be exogenous.

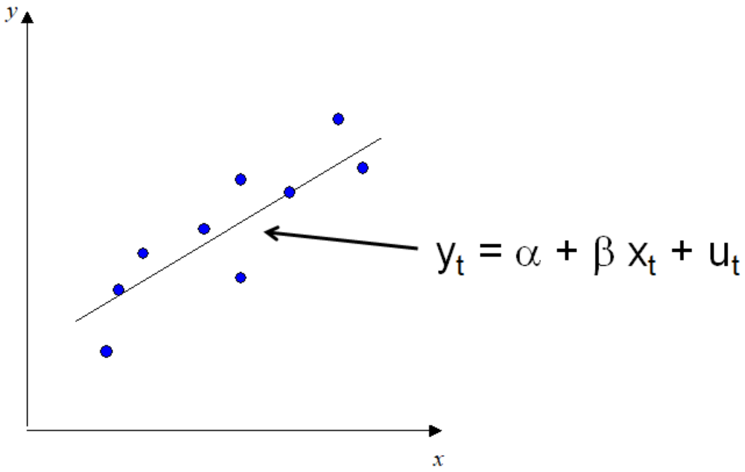
# Linear Regression

- From graphical inspection we may find that the variable y is positively affected by variable x.

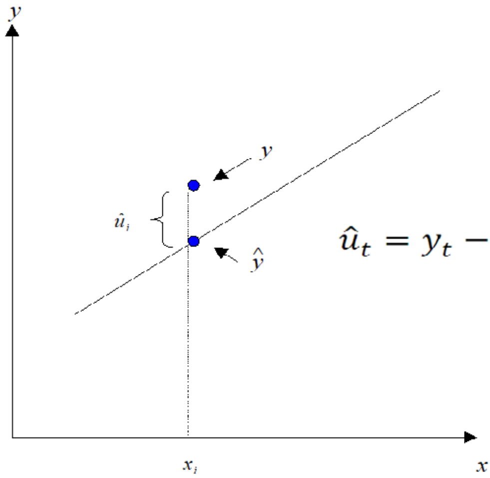


# Linear Regression

- To describe the dependence of y from x we may assume that the true relationship is linear!



- The method of ordinary least squares (OLS) chooses $\alpha$ and $\beta$ so that the (vertical) distances between the data points and the fitted line minimize a particular loss function:




$$\hat{u}_t = y_t - (\hat{\alpha} + \hat{\beta} x_t)$$

# Linear Regression

- The method of ordinary least squares (OLS) chooses $\alpha$ and $\beta$ so that the (vertical) distances between the data points and the fitted line minimize a particular loss function:

$$\hat{u}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$$

# Linear Regression

- The method of ordinary least squares (OLS) chooses $\alpha$ and $\beta$ so that the (vertical) distances between the data points and the fitted line minimize a particular loss function:

$$\hat{u}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$$

- In fact, OLS calculates coefficients such that the residual sum of squares (RSS) is minimal:

$$\min_{\alpha,\beta} RSS = \min_{\alpha,\beta} \sum_{i=1}^{N} u_i^2 = \min_{\alpha,\beta} \sum_{i=1}^{N} (y_i - (\alpha + \beta x_i))^2$$

$$= \sum_{i=1}^{N} (y_i - (\hat{\alpha} + \hat{\beta} x_i))^2$$

# Linear Regression

- The OLS formulas for estimating $\alpha$ and $\beta$ are:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}} \qquad \text{and} \qquad \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

# Linear Regression

- The OLS formulas for estimating $\alpha$ and $\beta$ are:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}} \qquad \text{and} \qquad \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

- The nominator of $\hat{\beta}$ is equivalent to the empirical covariance of x and y

# Linear Regression

- The OLS formulas for estimating $\alpha$ and $\beta$ are:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}} \qquad \text{and} \qquad \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

- The nominator of $\hat{\beta}$ is equivalent to the empirical covariance of x and y
- The denominator of $\hat{\beta}$ is equivalent to the empirical variance of x

- Some intuition for $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2 \frac{(y_i - \overline{y})}{(x_i - \overline{x})}}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

# Linear Regression

- Some intuition for $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2 \frac{(y_i - \overline{y})}{(x_i - \overline{x})}}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

- The estimator $\hat{\beta}$ is a weighted average

# Linear Regression

- Some intuition for $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2 \frac{(y_i - \overline{y})}{(x_i - \overline{x})}}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

- The estimator $\hat{\beta}$ is a weighted average
- First term is the weight calculated as the contribution of a single observation to total variance of x

# Linear Regression

- Some intuition for $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2 \frac{(y_i - \overline{y})}{(x_i - \overline{x})}}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

- The estimator $\hat{\beta}$ is a weighted average
- First term is the weight calculated as the contribution of a single observation to total variance of x
- Second term is the slope of that observation

# Outline

1. Basic Econometric Concepts I
   - Linear Regression
   - Hypothesis Testing
   - Goodness of Fit
   - Application: The Keynesian Phillips-Curve

# Hypothesis Testing

- We also want to know how "good" our estimates of $\alpha$ and $\beta$ are!
- Logic:
  - The sample of observations should be seen as one random draw from the underlying population

# Hypothesis Testing

- We also want to know how "good" our estimates of $\alpha$ and $\beta$ are!
- Logic:
  - The sample of observations should be seen as one random draw from the underlying population
  - Another draw most likely produces a different set of estimated coefficients, although the underlying population is still the same

# Hypothesis Testing

- We also want to know how "good" our estimates of $\alpha$ and $\beta$ are!
- Logic:
  - The sample of observations should be seen as one random draw from the underlying population
  - Another draw most likely produces a different set of estimated coefficients, although the underlying population is still the same
  - In fact, a large number of draws would deliver an entire distribution of coefficients

# Hypothesis Testing

- We also want to know how "good" our estimates of $\alpha$ and $\beta$ are!
- Logic:
  - The sample of observations should be seen as one random draw from the underlying population
  - Another draw most likely produces a different set of estimated coefficients, although the underlying population is still the same
  - In fact, a large number of draws would deliver an entire distribution of coefficients
  - If the standard deviation of this distribution is very large, the current estimate would not be reliable

# Hypothesis Testing

- The precision of the estimates is given by its standard errors (square root of variance):

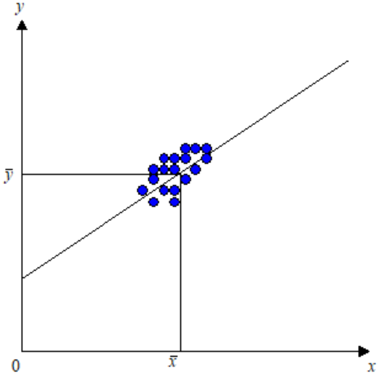
$$SE(\alpha^*) = s\sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N}(x_i - \overline{x})^2}}$$

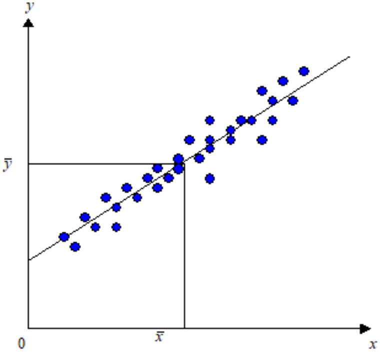
$$SE(\beta^*) = s\sqrt{\frac{1}{\sum_{i=1}^{N}(x_i - \overline{x})^2}}$$

- The standard deviation of the error term is estimated as:

$$s = \sqrt{\frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2}}$$

- Consider what happens if $\sum_{i=1}^{N}(x_i - \overline{x})^2$ is small or large:



# Hypothesis Testing

- Estimates of $\beta$ are c.p. more reliable ...
  - ...the smaller the standard deviation of the error term (the smaller the impact of noise!)
  - ...the larger the variance of the explaining variable (a new observation, even an outlier, would not change $\hat{\beta}$ much)

# Hypothesis Testing

- Estimates of $\beta$ are c.p. more reliable ...
  - ...the smaller the standard deviation of the error term (the smaller the impact of noise!)
  - ...the larger the variance of the explaining variable (a new observation, even an outlier, would not change $\hat{\beta}$ much)

- Estimates of $\alpha$ are, in addition, more reliable ...
  - ...the smaller the sum of squared $x_i$!
    Why?

# Hypothesis Testing

- The precision of estimates may be used to test economic hypotheses

# Hypothesis Testing

- The precision of estimates may be used to test economic hypotheses
- The null hypothesis is the statement derived from economics that is actually being tested.

# Hypothesis Testing

- The precision of estimates may be used to test economic hypotheses
- The null hypothesis is the statement derived from economics that is actually being tested.
- For example, if theory suggests that $\beta$ is equal to $\beta_0$, we could test:

$$\begin{aligned} H_0 &: \quad \beta = \beta_0 \\ H_1 &: \quad \beta \neq \beta_0 \end{aligned}$$

This would be known as a two sided test.

# Hypothesis Testing

- Assuming normality of residuals, standard normal variates can be constructed from $\alpha$ and $\beta$:

$$\frac{\hat{\alpha}-\alpha}{\sqrt{Var(\alpha)}} \sim \mathcal{N}(0,1) \qquad \text{and} \qquad \frac{\hat{\beta}-\beta}{\sqrt{Var(\beta)}} \sim \mathcal{N}(0,1)$$

# Hypothesis Testing

- Assuming normality of residuals, standard normal variates can be constructed from $\alpha$ and $\beta$:

$$\frac{\hat{\alpha}-\alpha}{\sqrt{Var(\alpha)}} \sim \mathcal{N}(0,1) \qquad \text{and} \qquad \frac{\hat{\beta}-\beta}{\sqrt{Var(\beta)}} \sim \mathcal{N}(0,1)$$

- However theoretical variances of parameter estimates are unknown and have to be estimated. Using estimated standard errors yields $t$-test statistics:

$$\frac{\hat{\alpha}-\alpha_0}{\widehat{se(\alpha)}} \sim t_{N-2} \qquad \text{and} \qquad \frac{\hat{\beta}-\beta_0}{\widehat{se(\beta)}} \sim t_{N-2}$$

- We need to choose a **significance level**, often denoted $\alpha$. This is also sometimes called the size of the test and it determines the region where we will reject or not reject the null hypothesis that we are testing.

# Hypothesis Testing

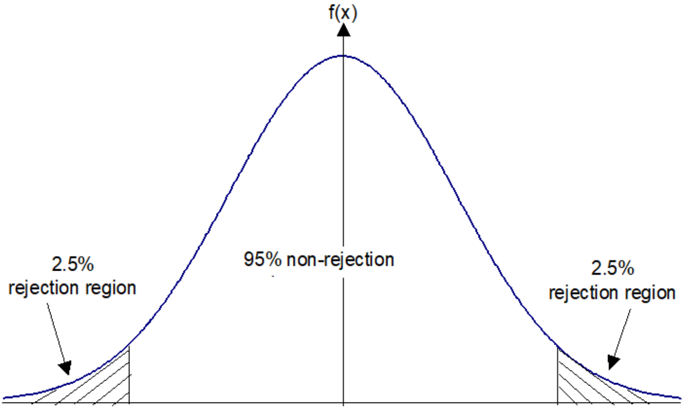
- We need to choose a **significance level**, often denoted $\alpha$. This is also sometimes called the size of the test and it determines the region where we will reject or not reject the null hypothesis that we are testing.
- Intuitive explanation is that we would only expect a result as extreme as this or more extreme say 5% of the time as a consequence of chance alone.

# Hypothesis Testing

- We need to choose a **significance level**, often denoted $\alpha$. This is also sometimes called the size of the test and it determines the region where we will reject or not reject the null hypothesis that we are testing.
- Intuitive explanation is that we would only expect a result as extreme as this or more extreme say 5% of the time as a consequence of chance alone.
- Conventional to use a 5% size of test, but 10% and 1% are also commonly used.

# Hypothesis Testing

- Given the significance level, we can calculate rejection and non-rejection areas!



# Hypothesis Testing

- Use the t-tables to obtain a critical value or values with which to compare the test statistic.

# Hypothesis Testing

- Use the t-tables to obtain a critical value or values with which to compare the test statistic.
- Finally perform the test. If the test statistic lies in the rejection region then reject the null hypothesis ($H_0$), else do not reject $H_0$.

# Hypothesis Testing

- Use the t-tables to obtain a critical value or values with which to compare the test statistic.
- Finally perform the test. If the test statistic lies in the rejection region then reject the null hypothesis ($H_0$), else do not reject $H_0$.
- In the limit, a t-distribution with an infinite number of degrees of freedom is a standard normal, i.e. $t_\vartheta \overset{\vartheta\to\infty}{\to} \mathcal{N}(0,1)$

# Hypothesis Testing

- Examples from statistical tables:

| Significance level | $\mathcal{N}(0,1)$ | $t_{40}$ | $t_4$ |
|---|---|---|---|
| 50% | 0 | 0 | 0 |
| 5% | 1.64 | 1.68 | 2.13 |
| 2.5% | 1.96 | 2.02 | 2.78 |
| 1% | 2.57 | 2.70 | 4.60 |

# Hypothesis Testing

The well-known t-ratio!

- If the test is

$$\begin{aligned} H_0 : \quad & \beta_i = 0 \\ H_1 : \quad & \beta_i \neq 0 \end{aligned}$$

  i.e. a test that the population coefficient is zero against a two-sided alternative, this is known as a t-ratio test:

$$\text{test stat} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

- The ratio of the coefficient to its SE is known as the *t*-ratio or *t*-statistic.

# Hypothesis Testing

Pitfalls in hypothesis testing!

- We usually reject $H_0$ if the test statistic is statistically significant at a chosen significance level.

# Hypothesis Testing

Pitfalls in hypothesis testing!

- We usually reject $H_0$ if the test statistic is statistically significant at a chosen significance level.
- There are two possible errors we could make:
    1. Rejecting $H_0$ when it was really true. This is called a type I error. (size of a test)

# Hypothesis Testing

Pitfalls in hypothesis testing!

- We usually reject $H_0$ if the test statistic is statistically significant at a chosen significance level.
- There are two possible errors we could make:
  1. Rejecting $H_0$ when it was really true. This is called a type I error. (size of a test)
  2. Not rejecting $H_0$ when it was in fact false. This is called a type II error. (power of a test)

# Outline

1. Basic Econometric Concepts I
   - Linear Regression
   - Hypothesis Testing
   - Goodness of Fit
   - Application: The Keynesian Phillips-Curve

# Goodness of Fit

Goodness of fit statistics

- We would also like some measure of how well our regression model actually fits the data.
- The most common goodness of fit statistic is known as $R^2$. One way to define $R^2$ is to say that it is the square of the correlation coefficient between y and $\hat{y}$.

# Goodness of Fit

- For another explanation, recall that what we are interested in doing is explaining the variability of y about its mean value
- The total sum of squares, TSS:

$$TSS = \sum_{i=1}^{N} (y_i - \overline{y})^2$$

# Goodness of Fit

- For another explanation, recall that what we are interested in doing is explaining the variability of y about its mean value
- The total sum of squares, TSS:

$$TSS = \sum_{i=1}^{N} (y_i - \overline{y})^2$$

- We can split the TSS into two parts, the part which we have explained (known as the explained sum of squares, ESS) and the part which we did not explain using the model (the RSS).

$$TSS = \sum_{i=1}^{N} (y_i - \overline{y})^2 = \sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{N} \hat{u}_i^2 = ESS + RSS$$

# Goodness of Fit

- Goodness of fit is:

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

# Goodness of Fit

- $R^2$ never falls if more regressors are added to the regression.

# Goodness of Fit

- $R^2$ never falls if more regressors are added to the regression.
- $R^2$ will always be at least as high for regressions with more variables.

# Goodness of Fit

- $R^2$ never falls if more regressors are added to the regression.
- $R^2$ will always be at least as high for regressions with more variables.
- In order to get around these problems, a modification is often made which takes into account the loss of degrees of freedom associated with adding extra variables.

# Goodness of Fit

- $R^2$ never falls if more regressors are added to the regression.
- $R^2$ will always be at least as high for regressions with more variables.
- In order to get around these problems, a modification is often made which takes into account the loss of degrees of freedom associated with adding extra variables.
- This is known as $\overline{R}^2$, or adjusted $R^2$:

$$\overline{R}^2 = 1 - \left( \frac{N-1}{N-K}(1-R^2) \right)$$

with K number of regressors.

# Outline

1. Basic Econometric Concepts I
   - Linear Regression
   - Hypothesis Testing
   - Goodness of Fit
   - Application: The Keynesian Phillips-Curve

# Application: The Keynesian Phillips-Curve

R-Example to illustrate parameter estimation and hypothesis testing.

Christian-Albrechts-Universität zu Kiel

Wirtschafts- und Sozialwissenschaftliche Fakultät

# Applied Econometrics of FX Markets

Stefan Reitz, QBER University of Kiel

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Linear Regression

- This lecture gives an introduction into R programming
- As an example for illustration we perform an empirical analysis of the Phillips curve for the US economy
- Codes and data are provided on the OLAT platform
- Please check the related movie 'Lecture2.mp4'

# Applied Econometrics of FX Markets

## 2. Basic Econometric Concepts II

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

1. Basic Econometric Concepts II
   - Multiple Linear Regression
   - Underlying Assumptions
   - Diagnostic Testing
   - Robust Standard Errors

# Multiple Linear Regression

- In the first session we derived formulas to identify the dependence of a variable y from an single exogenous variable x.
- In general, however, we have an entire set of exogenous variables $x_1, \ldots, x_K$, resulting in the regression model

$$y_i = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_K x_{i,K} + u_i$$

  for observation $i = 1 \ldots N$ with $u_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
- We already made use of this in our application for the Phillips curve

# Multiple Linear Regression

We can formulate the multiple regression model in vector notation:

$$y = X\beta + u$$

with

- Data: $y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \qquad X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,K} \\ \vdots & & \\ x_{N,1} & \cdots & x_{N,K} \end{pmatrix}$
- Vector of coefficients: $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$
- Error term: $u \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I_N)$

# Multiple Linear Regression

- The OLS-estimator minimizing least squares between regression line and data points can be formulated as:

$$\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y$$

- For K=1 we would get the estimators we know from last session:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}} \qquad \text{and} \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

# Multiple Linear Regression

- The variance of the coefficient estimators is given by

$$Var(\hat{\beta}) = \sigma^2(X^\top X)^{-1}$$

- With an appropriate variance estimator like $\hat{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N} \hat{e}_i^2$ we can form t-statistics:

$$test\ stat = \frac{\hat{\beta}_i - \beta_i}{\sqrt{(\hat{\sigma}^2(X^\top X)^{-1})_{i,i}}} \sim t_{N-2}$$

to test $H_O:\ \hat{\beta}_i = \beta_i$.

# Outline

# Underlying Assumptions

Regression model:

$$y_i = \alpha + \beta x_i + u_i$$

- In the classical linear regression model we observe data for $x$, but since $y$ also depends on an error term $u$, we must be specific about how the $u$ is generated generated.
- We usually make the following assumptions:

# Underlying Assumptions

1. The errors have zero mean:

$$E(u_i) = 0$$

for every i.

# Underlying Assumptions

1. The errors have zero mean:

$$E(u_i) = 0$$

   for every i.

2. The errors are uncorrelated:

$$Cov(u_i, u_j) = 0$$

   for $i \neq j$.

# Underlying Assumptions

1. The errors have zero mean:

$$E(u_i) = 0$$

   for every i.

2. The errors are uncorrelated:

$$Cov(u_i, u_j) = 0$$

   for $i \neq j$.

3. The variances of the errors is constant and finite for every i (homoscedasticity):

$$Var(u_i) = \sigma^2 < \infty$$

   for every i.

# Underlying Assumptions

1. The errors have zero mean:

$$E(u_i) = 0$$

   for every i.

2. The errors are uncorrelated:

$$Cov(u_i, u_j) = 0$$

   for $i \neq j$.

3. The variances of the errors is constant and finite for every i (homoscedasticity):

$$Var(u_i) = \sigma^2 < \infty$$

   for every i.

4. No relationship between $u$ and any regressor $x_1, \ldots, x_K$:

$$Cov(u_i, x_{i,j}) = 0$$

   for every i and j.

# Underlying Assumptions

- We implicitly assumed (1) to (3) already by requiring $u \sim \mathcal{N}(0, \sigma^2 I_N)$.
- Endogeneity problem, very common in economics

$$y_i = \alpha_1 + \beta_1 x_i + u_i$$
$$x_i = \alpha_2 + \beta_2 y_i + \epsilon_i$$

- If $\beta_2 \neq 0$ (implying a reverse causality) then $Cov(u_i, x_i) \neq 0$. (Insert for $y_i$ in the second equation to see this!)

# Underlying Assumptions

- Quite often, one or more of the assumptions (1) to (4) do not hold.
- We can adjust our empirical framework to relax assumptions
- This comes at cost of higher data requirements and/or involvement of more complicated computations
- How do we know that a assumption is not fulfilled?

  $\rightarrow$ Diagnostic Testing!

# Outline

1. Basic Econometric Concepts II
   - Multiple Linear Regression
   - Underlying Assumptions
   - Diagnostic Testing
   - Robust Standard Errors

# Diagnostic Testing: Autocorrelation

- OLS assumes that $Cov(u_i, u_j) = 0$ which means that the errors are uncorrelated
- If not, the estimated coefficients are still unbiased, but the standard errors are biased
- This clearly affects statistical significance of parameter estimates
- The Durbin Watson statistic gives a first impression

# Diagnostic Testing: Autocorrelation

Durbin Watson (DW) statistic

- If the error term is autocorrelated then $u_t = \beta u_{t-1} + \nu_t$
- The DW statistic is calculated as

$$DW = \frac{\sum_{t=2}^{T}(u_t - u_{t-1})^2}{\sum_{t=2}^{T} u_t^2}$$

$$DW = \frac{\sum_{t=2}^{T}(u_t^2 + u_{t-1}^2 - 2u_t u_{t-1})}{\sum_{t=2}^{T} u_t^2}$$

$$DW = \frac{Var(u_t) + Var(u_t) - 2COV(u_t, u_{t-1})}{Var(u_t)}$$

$$DW = 2 - 2\beta = 2(1 - \beta)$$

# Diagnostic Testing: Autocorrelation

Durbin Watson (DW) statistic

- If the error term is autocorrelated then $u_t = \beta u_{t-1} + \nu_t$
- The DW statistic $DW = 2(1 - \beta)$ is bounded

$$\beta = -1 \rightarrow DW = 4$$

$$\beta = 0 \rightarrow DW = 2$$

$$\beta = +1 \rightarrow DW = 0$$

- The DW statistic gives a first impression, no formal test
- The DW statistic only considers first order autocorrelation

# Diagnostic Testing: Autocorrelation

A simple test of autocorrelated errors

- If $Cov(u_i, u_j) = 0$ then the correlation coefficient

$$\rho(u_i, u_j) = \frac{Cov(u_i, u_j)}{\sqrt{Var(u_i)}\sqrt{Var(u_j)}} = 0$$

- If $\rho$ is statistically different from zero, we could argue that the assumption does not hold

# Diagnostic Testing: Autocorrelation

A simple test of autocorrelated errors

- The Ljung-Box-Q test uses the fact that

$$Q = N(N+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{N-k} \sim \chi^2$$

- If $Q > \chi^2_{1-\alpha,h}$ then reject the null hypothesis of no autocorrelation

# Diagnostic Testing: Heteroscedasticity

A simple test of heteroscedastic errors

- If $Cov(u_i^2, u_j^2) = 0$ then the correlation coefficient

$$\rho(u_i^2, u_j^2) = \frac{Cov(u_i^2, u_j^2)}{\sqrt{Var(u_i^2)}\sqrt{Var(u_j^2)}} = 0$$

- If $\rho$ is statistically different from zero, we could argue that the assumption does not hold

# Diagnostic Testing: Heteroscedasticity

A simple test of heteroscedastic errors

- The Ljung-Box-Q test uses the fact that

$$Q = N(N+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{N-k} \sim \chi^2$$

- If $Q > \chi^2_{1-\alpha,h}$ then reject the null hypothesis of no heteroscedasticity

# Diagnostic Testing: Non-Normality

- Standard tests rely on the assumption that residuals are normally distributed
- Are they?
- A popular test exploits that in case of normally distributed random variables the skewness (centralized third moment) is zero and the kurtosis (centralized fourth moment) is three.

# Diagnostic Testing: Non-Normality

Jarque-Bera Test

$$JB = \frac{N}{6}\left(Sk^2 + \frac{(Ku-3)^2}{4}\right) \sim \chi^2_2$$

with

- Skewness: $Sk = \frac{\frac{1}{N}\sum_{i=1}^{N}(u_i - \overline{u})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(u_i - \overline{u})^2\right)^{3/2}}$
- Kurtosis: $Ku = \frac{\frac{1}{N}\sum_{i=1}^{N}(u_i - \overline{u})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(u_i - \overline{u})^2\right)^{2}}$

If $JB > \chi^2_2$ then reject the null hypothesis of normality

# Outline

1. Basic Econometric Concepts II
   - Multiple Linear Regression
   - Underlying Assumptions
   - Diagnostic Testing
   - Robust Standard Errors

# Robust Standard Errors

- What if we observe significant departures from no-autocorrelation and normality?
- In general, parameter estimates remain unbiased, but their standard errors do not!
- First best solution is to provide a model that accounts for these issues (eg. non-overlapping data or ARMA/ARCH/GARCH (we will see future lectures)
- Second best solution: correct covariance matrix to provide 'robust' standard errors

# Robust Standard Errors

- **Heteroscedasticity-consistent (HC) standard errors** model the covariance matrix of residuals more appropriately and are therefore "robust" with respect to heteroscedasticity.
- We therefore have $u \sim \mathcal{N}(0, \Sigma)$.
- It's often assumed that $\Sigma = diag(\sigma_1^2, \ldots, \sigma_N^2)$.
- We could then estimate individual variances with $\hat{\sigma}_i^2 = \hat{u}_i^2$ and build standard errors from

$$\widehat{Var(\hat{\beta})}^{HC} = (X^\top X)^{-1}(X^\top diag(\hat{u}_1^2, \ldots, \hat{u}_N^2)X)(X^\top X)^{-1}$$

that are robust with respect to heteroscedasticity.

# Robust Standard Errors

- **HAC standard errors** additionally try to overcome the problem of autocorrelation.
- Popular are Newey-West standard errors:

$$\widehat{Var(\hat{\beta})}^{NW} = \frac{1}{N}\sum_{n=1}^{N} \hat{u}_i X_i X_i^\top + \frac{1}{N}\sum_{l=1}^{L}\sum_{n=l+1}^{N} w_l \hat{u}_i \hat{u}_{i-l}(X_i X_{i-l}^\top + X_{i-l} X_i^\top)$$

  with $X_i$ being the i-th row of X and $w_l = 1 - \frac{l}{L+1}$.
- $w_l$ can be understood as weighting giving disturbances farther apart from each other less weight.

# Applied Econometrics of FX Markets

## 4. Basic Time Series Models of Exchange Rates

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

1. Basic Time Series Models
   - Time Series and their Properties
   - Moving Averages
   - Autoregressive Processes
   - ARMA Models

# Time series Properties: Groundwork

Assume a variable $x_t$ follows an autoregressive process :

$$x_t = \phi x_{t-1} + u_t$$

# Time series Properties: Groundwork

Assume a variable $x_t$ follows an autoregressive process :

$$x_t = \phi x_{t-1} + u_t$$

Backward iteration gives:

$$x_t = \phi(\phi x_{t-2} + u_{t-1}) + u_t$$

$$x_t = \phi^2 x_{t-2} + \phi u_{t-1} + u_t$$

$$x_t = \phi^2(\phi x_{t-3} + u_{t-2}) + \phi u_{t-1} + u_t$$

... and so on
Gives:

$$x_t = \phi^K u_{t-K} + ... + \phi^2 u_{t-2} + \phi u_{t-1} + u_t$$

# Time series Properties: Groundwork

Interpretation:

- Observations of a time series are just a function of past shocks (or news!)
- Forward looking impact of a single shock $\rightarrow$ impulse response function
- Role of $\phi$: persistence of shocks!
- Mean reversion (calculate first difference):

$$x_t = \phi^K u_{t-K} + ... + \phi^2 u_{t-2} + \phi u_{t-1} + u_t$$

$$x_{t-1} = \phi^{K-1} u_{t-(K+1)} + ... + \phi^2 u_{t-3} + \phi u_{t-2} + u_{t-1}$$

$$x_t - x_{t-1} = (\phi - 1)(\phi^{K-1} u_{t-K} + ... + \phi u_{t-2} + u_{t-1}) + u_t$$

$$x_t - x_{t-1} = (\phi - 1) x_{t-1} + u_t$$

# Time series Properties: Groundwork

Interpretation:

- The unconditional expectation of x is a long-run (historical) average independent of a given point in time

$$E[x] = \frac{1}{N}\sum x_t = E[f(u_t)] = 0$$

- The conditional expectation of $x_{t+1}$ is based on current conditions summarized in the information variable $\Omega_t$

$$\begin{aligned} E_t[x_{t+1}|\Omega_t] =& E_t[\phi x_t|\Omega_t] + E_t[u_{t+1}|\Omega_t] \\ =& \phi x_t \end{aligned}$$

# Time series Properties: Groundwork

Interpretation:

- For $\phi = 1$ no mean reversion, shocks do not die out!
- Time series is called "nonstationary" or "has a unit root"
- Variance is not finite anymore!

$$V(x_t) = \phi^{2K}\sigma^2_{t-K} + ... + \phi^{2*2}\sigma^2_{t-2} + \phi^2\sigma^2_{t-1} + \sigma^2_t$$

$$V(x_t) = \frac{1}{1-\phi^2}\sigma^2$$

# Time Series Properties

We call a time series $x_t$ (weakly) stationary if

- second moments are are finite:

$$E(x_t^2) < \infty \qquad \text{for all t}$$

- the unconditional mean is constant over time:

$$E(x_t) = \mu \qquad \text{for all t}$$

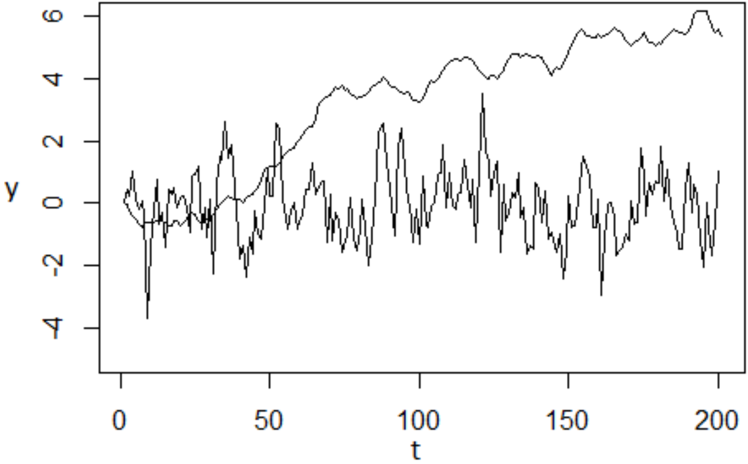
- the autocovariance function

$$\gamma(s, t) := Cov(x_s, x_t)$$

only depends on the distance $\tau = |s - t|$ of time periods.

# Time Series

Figure: A stationary and a non-stationary time series



# Outline

1. Basic Time Series Models
   - Time Series and their Properties
   - Moving Averages
   - Autoregressive Processes
   - ARMA Models

# White Noise

- We model time series by combining several building blocks to mimic statistical features of our data
- The most elementary one is known as (strict) white noise that is an independently and identically distributed (iid) time series $\epsilon_t$ with mean 0 and variance $\sigma^2$.
- White noise processes are stationary.
- If $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ we call $\epsilon_t$ Gaussian white noise.

# Moving Averages

- A way to build a time series model with non-zero mean and autocorrelation from white noises is to use a moving average
- A moving average of order q (MA(q)) $y_t$ is defined by

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

where $\epsilon_t$ is white noise.

# Moving Averages

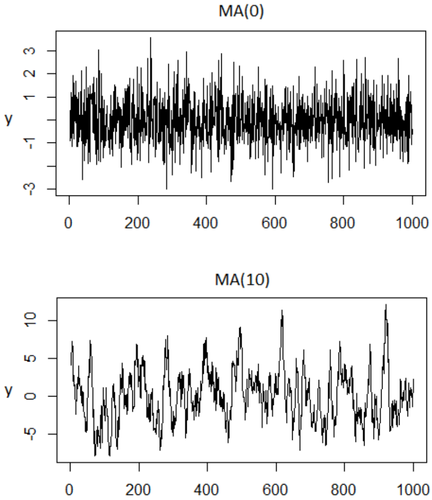
Properties of a MA(q)-process $y_t$:

- $y_t$ is stationary
- $E(y_t) = \mu \qquad$ for all t
- $\gamma(\tau) = Cov(y_{t+\tau}, y_t) = \begin{cases} \sigma^2 \sum_{i=0}^{q-\tau} \theta_i \theta_{i+\tau} & \tau \leq q \\ 0 & \tau > q \end{cases}$

  for all t and $\tau \geq 0$

# Moving Averages

Figure: Realizations of a MA(0)- and a MA(10)-process



# Outline

1. Basic Time Series Models
   - Time Series and their Properties
   - Moving Averages
   - Autoregressive Processes
   - ARMA Models

# Autoregressive Processes

- Autocorrelation of MA(q)-processes is zero after q lags
- But most time series we observe in economics and finance show smoothly declining autocorrelations
- This feature can be achieved with another popular building block for time series: Autoregressive (AR) processes.
- An autoregressive process of order p (AR(p)) $y_t$ is defined by

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t$$

where $\epsilon_t$ is white noise.

# AR(1)-process

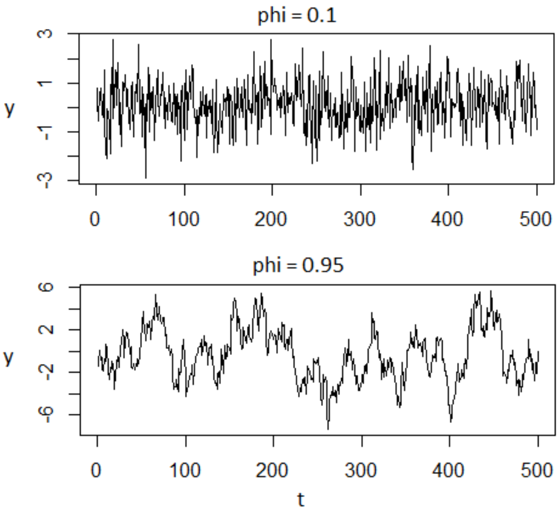
- Let's have a closer look at the AR(1) process:

$$y_t = c + \phi y_{t-1} + \epsilon_t$$

- AR(1) processes are stationary iff $|\phi| < 1$.
- $\phi$ is often called persistence parameter:
  - For $\phi$ close to 1 shocks have a persistent impact on future values of the series
  - For $\phi$ close to 0 shocks die out very quickly

# AR(1)-process

Figure: Realizations of two AR(1)-processes



# AR(1)-process

Properties of stationary (!) AR(1)-processes:

- $E(y_t) = \frac{c}{1-\phi}$ for all t
- $\gamma(\tau) = Cov(y_{t+\tau}, y_t) = \frac{\phi^\tau}{1-\phi^2}\sigma^2$ for all t and $\tau \geq 0$
- $\gamma(\tau) \neq 0$ for every $\tau$ and $\phi \neq 0$, but $\lim_{\tau\to 0}\gamma(\tau) = 0$.

# Outline

1. Basic Time Series Models
   - Time Series and their Properties
   - Moving Averages
   - Autoregressive Processes
   - ARMA Models

# ARMA Models

- The classisal workhorse model in univariate time series analysis combines AR(p)- and MA(q)-processes.
- An ARMA(p,q)-processes is defined by

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

where $\epsilon_t$ is white noise.

# ARMA Models

- Suppose we have time series data $y_1, \ldots, y_p$.
- How to estimate the parameters $\varphi = (c, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$?
- If $q = 0$, we have an AR(p) model that can be rewritten as linear model that we already know from lecture 2:

$$\begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & y_p & y_{p-1} & \cdots & y_1 \\ 1 & y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{pmatrix} \begin{pmatrix} c \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_T \end{pmatrix}$$

- $\rightarrow$ Use OLS!

# ARMA Models

- If $q \neq 0$, we can not use OLS since lagged error terms are not observable.
- But we can use Maximum Likelihood (ML) estimation instead!
- The idea is to choose parameters $\hat{\varphi}^{ML}$ such that the likelihood function is maximized.
- For given parameter values, the likelihood tells you the probability of your observations $y = (y_1, \ldots, y_T)^\top$:

$$\begin{aligned}\mathcal{L}(\varphi) &= P_\varphi(y_1, \ldots, y_T) \\ &= \prod_{t=1}^{T} P\varphi(y_t)\end{aligned}$$

# ARMA Models

- We may assume that $\epsilon_t$ is Gaussian white noise
- For an ARMA(p,q)-process we have then

$$\mathcal{L}(\varphi) = (2\pi)^{-T/2}|\Gamma(\varphi)|^{-1/2} exp\left\{-\frac{1}{2}(y - E(y))^\top \Gamma(\varphi)^{-1}(y - E(y))\right\}$$

  where $\Gamma(\varphi) = Var(y) = E(yy^\top) - E(y)E(y)^\top$.
- For estimation we can numerically compute

$$\hat{\varphi}^{ML} = \arg\max_{\varphi} \mathcal{L}(\varphi)$$

# Model Selection

- We know how to estimate parameters of an ARMA(p,q)-model.
- But how should we choose p and q?
- Idea: Choose p and q such that the estimated residual variance $\hat{\sigma}^2_{p,q}$ is small.
- Problem: $\hat{\sigma}^2_{p,q}$ is non-increasing in p and q
  $\rightarrow$ trade-off between model fit and model complexity

# Model Selection

Popular information criteria for model selection:

- AIC: $\ln \hat{\sigma}^2_{p,q} + (p+q)\frac{2}{T}$
- BIC: $\ln \hat{\sigma}^2_{p,q} + (p+q)\frac{\ln T}{T}$
- HQ: $\ln \hat{\sigma}^2_{p,q} + (p+q)\frac{2\ln(\ln T)}{T}$

# Applied Econometrics of FX Markets

## 4. Modeling Trends

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

1. Modeling Trends
   - Testing for unit roots
   - Testing for unit roots in exchange rates
   - Testing purchasing power parity

# Testing for unit roots

- Time series behavior of AR-Processes
- Assume that

$$q_t = \alpha + \beta q_{t-1} + u_t$$

- The non-stochastic part is a simple difference equation

$$q_t = \alpha + \beta q_{t-1}$$

- The solution has two parts, the adjustment dynamics and a long-run equilibrium

# Testing for unit roots

- The solution of adjustment dynamics are found setting $\alpha = 0$ and assuming
$$q_t = Ab^t$$
- That means
$$Ab^t = \beta Ab^{t-1}$$
- or
$$b = \beta$$
- Thus,
$$q_t^{ad} = A\beta^t$$

# Testing for unit roots

- The solution of the long-run equilibrium is found setting $q_t = q_{t-1} = q_{lr}$

$$q_{lr} = \alpha + \beta q_{lr}$$

- That means

$$q_{lr} = \frac{\alpha}{1-\beta}$$

- The complete solution is

$$q_t = q_t^{ad} + q_{lr}$$

- Thus,

$$q_t = A\beta^t + \frac{\alpha}{1-\beta}$$

# Testing for unit roots

- The constant factor A can be determined by computing an initial shock at time t = 0!

$$q_0 = A\beta^0 + \frac{\alpha}{1-\beta}$$

- That means

$$A = q_0 - \frac{\alpha}{1-\beta}$$

- And

$$q_t = (q_0 - \frac{\alpha}{1-\beta})\beta^t + \frac{\alpha}{1-\beta}$$

- Interpret $\beta$!

# Consequences for empirical analysis

- What happens if we run regressions using nonstationary variables? Consider the regression

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \tag{1}$$

- OLS assumptions require that both variables are stationary and errors have mean zero and finite variance
- In the presence of nonstationary variables we get **spurious regressions** (Granger and Newbold, 1974)
- Regression output looks good, but coefficients and standard errors are biased

# Consequences for empirical analysis

- Granger and Newbold performed a simulation study using two nonstationary variables completely unrelated to each other where error terms are independent by definition

$$y_t = y_{t-1} + \varepsilon_{yt}$$

$$x_t = x_{t-1} + \varepsilon_{xt}$$

- A regression of (1) clearly makes no sense!
- However, the authors found that $\beta_1 = 0$ was rejected at the 5% level in 75% of the cases

# Intuition

- Nonstationary errors in (1) imply permanent deviations from the equilibrium, which is hard to interpret in terms of a economic model
- Abstracting from the constant term, eq. (1) says that

$$\epsilon_t = y_t - \alpha_1 x_t$$

- Obviously, that means

$$\epsilon_t = \sum_{i=1}^{t} \varepsilon_{yt} - \alpha_1 \sum_{i=1}^{t} \varepsilon_{xt} \tag{2}$$

# Intuition

- Error variance increases to infinity as $t \to \infty$
- Expected error is non-zero!

$$E_t(\epsilon_{t+i}) = \epsilon_t$$

- Errors are strongly autocorrelated (Updated of eq. (2) shows that two consecutive errors share a large number of shocks)
- Any t-test, F-test or $R^2$ are unreliable
- Problems do not disappear in large samples! Phillips (1986) showed that $\beta_1 = 0$ is falsely rejected more often as sample size increases

# As a result

- Before running OLS-type regressions we should test stationarity of included variables. If the variable q has to be tested, a regression

$$\Delta q_t = \alpha + (\beta - 1)q_{t-1} + u_t \tag{3}$$

  is performed. Nonstationarity is then rejected if $(\beta - 1)$ is significantly negative.
- Of course, eq. (3) is misspecified under $H_0$!
- Dickey and Fuller (1979) calculated critical values for different models (3) and different sample sizes.

# Outline

1. Modeling Trends
   - Testing for unit roots
   - Testing for unit roots in exchange rates
   - Testing purchasing power parity

# Testing for unit roots in exchange rates

- See R excercise!

# Outline

1. Modeling Trends
   - Testing for unit roots
   - Testing for unit roots in exchange rates
   - Testing purchasing power parity

# Tests for unit roots in real exchange rates

- Real exchange rate is expected to be constant under PPP:

$$Q_t = S_t P_t^* / P_t = 1$$

- Real exchange rate in logs: $q_t = s_t + p_t^* - p_t = 0$
- Deviations from PPP ($q_t \neq 0$) should at least be corrected over time

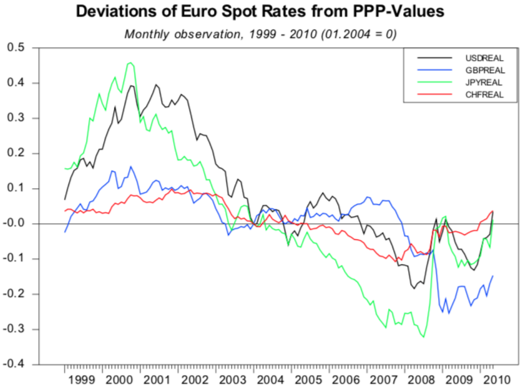
$$\Delta q_t = \gamma_0 + \gamma_1 q_{t-1} + \sum_{i=1}^{n} \gamma_{i+1} \Delta q_{t-i} + \epsilon_t$$

- This is true if $\gamma_1$ is significantly negative!

# Regression analysis

- See R excercise!

# Regression analysis




Data sources: IMF-IFS; ECB

# Applied Econometrics of FX Markets

## 5. Testing Parity Conditions

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

# UIP Testing

- Assume a risk neutral agent considers investing domestic liquidity on foreign capital markets. $S_t$ denotes the exchange rate as amount of home currency exchanged for one unit of foreign currency. The interest paid on domestic funds is denoted by $i_t$, the interest paid on foreign markets is denoted by $i_t^*$ .
- The expectation of profits $G_{t+1}$ from exporting X units of home currency is

$$\mathbb{E}_t(G_{t+1}) = X\left(\frac{(1+i_t^*)\mathbb{E}_t(S_{t+1})}{S_t}\right) - X(1+i_t)$$

and X may be negative.

# UIP Testing

- An international capital market equilibrium may be defined as a situation where funds are not moved across borders
- This implies that expected profits from exporting funds must be zero!

$$
\begin{aligned}
\mathbb{E}_t(G_{t+1}) \overset{!}{=} 0 \quad &\Leftrightarrow \quad \left(\frac{(1+i_t^*)\mathbb{E}_t(S_{t+1})}{S_t}\right) = (1+i_t) \\
&\Leftrightarrow \quad S_t = \left(\frac{(1+i_t^*)\mathbb{E}_t(S_{t+1})}{(1+i_t)}\right)
\end{aligned}
$$

# UIP Testing

- When writing in logs we find that

$$s_t = \mathbb{E}_t(s_{t+1}) + (i_t^* - i_t)$$

- Or in a more familiar way,

$$\mathbb{E}_t(s_{t+1}) - s_t = i_t - i_t^*$$

- This dynamic equilibrium condition is called:

Uncovered Interest Parity (UIP) .

# UIP Testing

- Assuming Rational Expectations (RE)

$$\mathbb{E}_t(s_{t+1}) = s_{t+1} + u_{t+1}$$

we can construct the linear regression equation

$$s_{t+1} - s_t = i_t - i_t^* + u_{t+1}$$

- This equation can be used to test the joint hypothesis of UIP and RE.

# Linear Regression

- We can test the joint hypothesis by estimating the regression model

$$s_{t+1} - s_t = \alpha + \beta(i_t - i_t^*) + u_{t+1}$$

- If UIP and RE holds we should find $\hat{\alpha} \approx 0$ and $\hat{\beta} \approx 1$.
- Literature: Fama (1984) and Engel (2016)

# Linear Regression

- Data: (Source: Thomsen-Reuters Eikon)
  - $S$ : EUR/USD exchange rate
  - $i, i^*$ : 3-month Libor interest rates
- Practical Implementation in R:
  - Compute log-changes of exchange rates (in percentages)

$$y \quad <- \quad 100 * (log(S[4:T]) - log(S[1:(T-3)]))$$

  - Compute interest rate differentials

$$x \quad <- \quad (i - i\_star)[1:(T-1)]/4$$

  - Fit linear regression model:

```
mod <- lm(y~x)
summary(mod)
```

# Linear Regression

Move over to R!

# Applied Econometrics of FX Markets

## 5. Testing Parity Conditions

Stefan Reitz

# Course Outline

# Outline

1. Testing Parity Conditions
   - Uncovered Interest Parity
   - The News Approach
   - Covered Interest Parity

# UIP Testing - Recap

- Assume a risk neutral agent considers investing domestic liquidity on foreign capital markets. $S_t$ denotes the exchange rate as amount of home currency exchanged for one unit of foreign currency. The interest paid on domestic funds is denoted by $i_t$, the interest paid on foreign markets is denoted by $i_t^*$ .
- The expectation of profits $G_{t+1}$ from exporting X units of home currency is

$$\mathbb{E}_t(G_{t+1}) = X\left(\frac{(1+i_t^*)\mathbb{E}_t(S_{t+1})}{S_t}\right) - X(1+i_t)$$

and X may be negative.

# UIP Testing - Recap

- An international capital market equilibrium may be defined as a situation where funds are not moved across borders
- This implies that expected profits from exporting funds must be zero!

$$\mathbb{E}_t(G_{t+1}) \overset{!}{=} 0 \quad \Leftrightarrow \quad \left(\frac{(1+i_t^*)\mathbb{E}_t(S_{t+1})}{S_t}\right) = (1+i_t)$$
$$\Leftrightarrow \quad S_t = \left(\frac{(1+i_t^*)\mathbb{E}_t(S_{t+1})}{(1+i_t)}\right)$$

# UIP Testing - Recap

- When writing in logs we find that

$$s_t = \mathbb{E}_t(s_{t+1}) + (i_t^* - i_t)$$

- Or in a more familiar way,

$$\mathbb{E}_t(s_{t+1}) - s_t = i_t - i_t^*$$

- This dynamic equilibrium condition is called:

Uncovered Interest Parity (UIP) .

# UIP Testing - Recap

- Assuming Rational Expectations (RE)

$$\mathbb{E}_t(s_{t+1}) = s_{t+1} + u_{t+1}$$

we can construct the linear regression equation

$$s_{t+1} - s_t = i_t - i_t^* + u_{t+1}$$

- This equation can be used to test the joint hypothesis of UIP and RE.

# Linear Regression - Recap

- We can test the joint hypothesis by estimating the regression model

$$s_{t+1} - s_t = \alpha + \beta(i_t - i_t^*) + u_{t+1}$$

- If UIP and RE holds we should find $\hat{\alpha} \approx 0$ and $\hat{\beta} \approx 1$.
- Literature: Fama (1984) and Engel (2016)

# Outline

1. Testing Parity Conditions
   - Uncovered Interest Parity
   - The News Approach
   - Covered Interest Parity

# The News Approach

- The agent may consider cross-border investment for more than one period. Think of a 3-month contract that might be prolonged several times.
- Assuming Rational Expectations we can solve

$$s_t = \mathbb{E}_t(s_{t+1}) + (i_t^* - i_t)$$

forward to obtain

$$s_t = \sum_{k=0}^{K} \mathbb{E}_t(i_{t+k}^* - i_{t+k}) + \mathbb{E}_t(s_{t+K})$$

(see lecture/tutorial "Foreign Exchange Markets")

- Thus, the current level of the exchange rate is determined by (i) the sum of future expected short-term interest differentials and (ii) its expected long-run level.

# The News Approach

The first difference of the spot rate is

$$\Delta s_t = \sum_{k=0}^{K} \mathbb{E}_t(i^*_{t+k} - i_{t+k}) - \sum_{k=0}^{K} \mathbb{E}_{t-1}(i^*_{t-1+k} - i_{t-1+k})$$
$$+ \mathbb{E}_t(s_{t+K}) - \mathbb{E}_{t-1}(s_{t-1+K})$$

or, equivalently,

$$\begin{aligned}\Delta s_t = & -(i^*_{t-1} - i_{t-1}) + (i^*_t - i_t) - \mathbb{E}_{t-1}(i^*_t - i_t) \\ & + \sum_{k=1}^{K-1} (\mathbb{E}_t(i^*_{t+k} - i_{t+k}) - \mathbb{E}_{t-1}(i^*_{t+k} - i_{t+k})) + \mathbb{E}_t(i^*_{t+K} - i_{t+K}) \\ & + (\mathbb{E}_t - \mathbb{E}_{t-1})(s_{t-1+k}) + \mathbb{E}_t(\Delta s_{t+K})\end{aligned}$$

# The News Approach

The spot rate is driven by

- lagged interest differential
- time t information about current differential
- time t information about future expected differentials
- time t information about the long-run level of the exchange rate
- expectation about the new final return (Due to a given investment horizon K)
- Literature
  News approach: Anderson et al. (2003)
  Present value representation: Engel and West (2006)

# The News Approach

- To make things a little more comfortable assume that

$$i_t^* - i_t = \phi(i_{t-1}^* - i_{t-1}) + u_t$$

where $\phi$ is a persistence parameter and $u_t$ is an error term.

# The News Approach

- To make things a little more comfortable assume that

$$i_t^* - i_t = \phi(i_{t-1}^* - i_{t-1}) + u_t$$

  where $\phi$ is a persistence parameter and $u_t$ is an error term.
- Since the one-step-ahead expectation is

$$\mathbb{E}_{t-1}(i_t^* - i_t) = \phi(i_{t-1}^* - i_{t-1}),$$

  the error term $u_t$ represents the news shock.

# Expectations of expectations

For $x_t = i_t^* - i_t$:

$$
\begin{aligned}
\mathbb{E}_{t-1} x_{t+1} &= \mathbb{E}_{t-1}(\mathbb{E}_t x_{t+1}) \\
&= \mathbb{E}_{t-1}(\phi x_t) \\
&= \phi(\phi x_t) \\
&= \phi^2 x_t
\end{aligned}
$$

That means $\mathbb{E}_t x_{t+i} = \phi^i x_t$

# The level of the spot rate:

From eq. (1) for $K \to \infty$:

$$
\begin{aligned}
s_t &= \sum_{k=0}^{\infty} \phi^k x_t + \mathbb{E}_t(s_{t+\infty}) \\
&= x_t \sum_{k=0}^{\infty} \phi^k + \mathbb{E}_t(s_{t+\infty}) \\
&= x_t \frac{1}{1-\phi} + \mathbb{E}_t(s_{t+\infty}) \\
&= (\phi x_{t-1} + u_t) \frac{1}{1-\phi} + \mathbb{E}_t(s_{t+\infty}) \\
&= \frac{\phi}{1-\phi} x_{t-1} + \frac{1}{1-\phi} u_t + \mathbb{E}_t(s_{t+\infty})
\end{aligned}
$$

# The change of the spot rate:

For $\Delta s_t = s_t - s_{t-1}$:

$$\begin{aligned}
\Delta s_t =& \sum_{k=0}^{\infty} \phi^k x_t + \mathbb{E}_t(s_{t+\infty}) - \sum_{k=0}^{\infty} \phi^k x_{t-1} + \mathbb{E}_{t-1}(s_{t+\infty}) \\
=& \sum_{k=0}^{\infty} \phi^k (\phi x_{t-1} + u_t) - \sum_{k=0}^{\infty} \phi^k x_{t-1} + (\mathbb{E}_t - \mathbb{E}_{t-1})(s_{t+\infty}) \\
=& \frac{1}{1-\phi}(\phi x_{t-1} + u_t) - \frac{1}{1-\phi} x_{t-1} + (\mathbb{E}_t - \mathbb{E}_{t-1})(s_{t+\infty}) \\
=& (\frac{\phi}{1-\phi} - \frac{1}{1-\phi}) x_{t-1} + \frac{1}{1-\phi} u_t + (\mathbb{E}_t - \mathbb{E}_{t-1})(s_{t+\infty}) \\
=& - x_{t-1} + \frac{1}{1-\phi} u_t + (\mathbb{E}_t - \mathbb{E}_{t-1})(s_{t+\infty})
\end{aligned}$$

# The News Approach

Thus we have for $x_t = i_t^* - i_t$:

$$s_t = \frac{\phi}{1-\phi}(i_{t-1}^* - i_{t-1}) + \frac{1}{1-\phi}u_t + E_t(s_{t+\infty})$$

$$\Delta s_t = (i_{t-1} - i_{t-1}^*) + \frac{1}{1-\phi}u_t + (\mathbb{E}_t - \mathbb{E}_{t-1})(s_{t+\infty})$$

This can be used as another starting point for empirical analysis!

# The News Approach

Move over to R!

# Outline

1. Testing Parity Conditions
   - Uncovered Interest Parity
   - The News Approach
   - Covered Interest Parity

# Covered Interest Parity

- The Covered Interest Parity (CIP) is a no-arbitrage condition stating that a forward rate $F_{t,t+1}$ assigned in period t for exchange in period t+1 should be equal today's exchange rate adjusted by interest rates:

$$F_{t,t+1} = S_t \frac{1 + i_t}{1 + i_t^*}$$

(Derivation in lecture/tutorial "Foreign Exchange Markets"

- Taking logarithms we find a linear relationship:

$$f_{t,t+1} - s_t = i_t - i_t^*$$

# Covered Interest Parity

- An idea to test whether CIP holds is to run a regression

$$f_{t,t+1} - s_t = \alpha + \beta(i_t - i_t^*) + u_t$$

  and test $H_0 : \alpha = 0, \beta = 1$.
- When choosing the data:
  - Home and foreign asssets must be comparable (same maturity, default risk...)
  - Transaction costs should be taken into account (use bid and ask quotes!)
  - There must be no barriers like capital controls
- But what could still be a problem with this empirical setup?

# Covered Interest Parity

But what could still be a problem with this empirical setup?

- Under $H_0$ we have

$$f_{t,t+1} - s_t = i_t - i_t^* + u_t$$

- Since $u_t \neq 0$ in general, CIP does not hold under $H_0$!
- Positive/Negative $u_t$ would offer an arbitrage opportunity.
- However: CIP holds on average under $H_0$ because $E(u_t) = 0$.
- So if you reject $H_0$, you would reject CIP on average and in particular reject CIP holding continuously!
- But the test still has weak power (high type II error)

# Covered Interest Parity

- Another way to test CIP is to calculate deviations to CIP (the possible arbitrage profit)

$$rx_{t+1} = f_{t,t+1} - s_t - i_t + i_t^*$$

  and test whether they differ from zero.

- Empirical evidence:
  - CIP deviations have been extremely rare before the financial crisis in 2007
    (Akram, Rime and Sarno (Journal of International Economics 2008))
  - Since the financial crisis in 2007 we find significant and persistent deviations to CIP
    (Du, Wenxin, Tepper and Verdelhan (Journal of Finance 2018))

# Covered Interest Parity

- Why does CIP not hold in recent periods?
- Still active research field.
- Idea:
  - Financial institutions face stricter regulatory balance sheet constraints (have to hold more equity) after the crisis. CIP trades exploiting arbitrage would enlarge the balance sheet and would decrease equity ratios. Hence arbitrage oppotunties may be left unexploited.

# Applied Econometrics of FX Markets

## 6. Modeling Volatility

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Recap:

- Variance of a variable $x_t$ is defined as

$$V_t(x_{t+1}) = E_t(x_{t+1} - E_t(x_{t+1}))^2$$

- The expected value of the variable comes from the economic or econometric model
- Assume we currently rely on an AR(1) model for $x_t$

$$x_{t+1} = \phi x_t + u_{t+1}$$

so that

$$E_t(x_{t+1}) = \phi x_t$$

- As a result

$$V_t(x_{t+1}) = E_t(x_{t+1} - \phi x_t)^2$$
$$V_t(x_{t+1}) = E_t(u_{t+1})^2 = \sigma_u^2$$

# Recap:

- Fitted residuals $\hat{u}_t$ should have constant variance due to our OLS assumptions
- This implies that the time series of $u_t^2$ should not exhibit any systematic variation
- Is that the case for high frequency time series of financial markets such as FX?

# Recap:

- We have argued economically that UIP is an adequate model for the EUR-USD exchange rate change:

$$s_t - s_{t-1} = (i_{t-1} - i^*_{t-1}) + u_t$$

- Variance of the exchange rate return is

$$V_{t-1}(s_t - s_{t-1}) = E_{t-1}((s_t - s_{t-1}) - E_{t-1}(s_t - s_{t-1}))^2$$

- The expected value of the return comes from the interest differential so that

$$E_{t-1}(s_t - s_{t-1}) = i_{t-1} - i^*_{t-1}$$

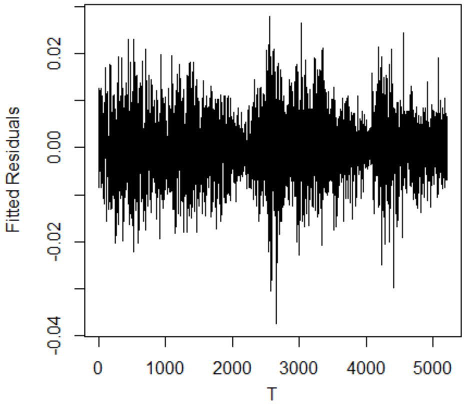
- As a result

$$V_{t-1}(s_t - s_{t-1}) = E_t(u_{t+1})^2 = \sigma_u^2$$

- See example in R!

# Recap:

Figure: AR(1)-Residuals of log EUR-USD FX-Rate




- Problem: Autocorrelated volatility or also known as volatility clustering

# Outline

1. Modeling Volatility
   - GARCH Models
   - GARCH-M models

# GARCH models

- GARCH models were introduced to deal with autocorrelated volatility
- GARCH means **G**eneral **A**uto**R**egressive **C**onditional **H**eteroskedasticity
- The idea is to model residuals $u_t$ with an own (non-linear) time series model

$$u_t = \varepsilon_t \sqrt{h_t}$$
$$h_t = c + a u_{t-1}^2 + b h_{t-1}$$

where $\varepsilon_t$ is i.i.d. noise with mean 0 and variance 1.

- How does this make volatility autocorrelated?

# GARCH models

- Conditional variance of $s_t$:

$$\begin{aligned} Var_{t-1}(s_t) &= E_{t-1}((s_t - E_{t-1}(s_t))^2) \\ &= E_{t-1}(u_t^2) \\ &= h_t E_{t-1}(\varepsilon_t^2) \\ &= h_t \\ &= c + a u_{t-1}^2 + b h_{t-1} \end{aligned}$$

- Volatility follows an ARMA(1,1) process
- This reflects the fact that trading on financial markets evolves in "heat waves".

# GARCH Models

- Unconditional variance:

$$E(u_t^2) = E(\varepsilon_t^2)(c + au_{t-1}^2 + bh_{t-1})$$

- Since $E(\varepsilon_t^2) = 1$ and $u_t = \varepsilon_t\sqrt{h_t}$ we find $E(u_t^2) = E(h_t)$
- In equilibrium, if $E(u_t^2) = E(u_{t+j}^2)$ for all j
- it follows that

$$\begin{aligned} E(u_t^2) &= c + aE(u_t^2) + bE(u_t^2) \\ &= \frac{c}{1 - a - b} \end{aligned}$$

# GARCH Models

- GARCH residuals can be also used in general linear models (and not only time series).
- An UIP model with heteroskedastic errors:

$$s_t - s_{t-1} = \beta_0 + \beta_1(i_{t-1} - i^*_{t-1}) + u_t$$

$$u_t = \varepsilon_t \sqrt{h_t}$$

$$h_t = c + a u^2_{t-1} + b h_{t-1}$$

- Again, we expect $\beta_0 = 0, \ \beta_1 = 1$
- What do we expect for c,a,b?

# GARCH Estimation

- The model above cannot be estimated using ordinary least squares
- To deal with the nonlinear recursive nature of the model the standard choice is (Quasi-) Maximum Likelihood

$$\mathcal{L}(\beta) = \prod_{t=1}^{N} f_{No}(y_t, x_t, \beta) \rightarrow \text{max!}$$

# Estimation

- Log of the Likelihood is a monotonic transformation and simplifies calculations

$$\mathcal{L}(\beta) = \sum_{t=1}^{N} \ln f_{No}(y_t, x_t, \beta) \to \max!$$

$$\mathcal{L}(\beta) = \sum_{t=1}^{N} -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_t^2 - \frac{u_t^2}{2\sigma_t^2}$$

- The parameter estimates are then calculated by numerical maximization of the log likelihood

# Outline

1. Modeling Volatility
   - GARCH Models
   - GARCH-M models

# Applied Econometrics of FX Markets

## 6. Modeling Volatility

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

1. Modeling Volatility
   - GARCH Models
   - GARCH-M models

# Outline

1. Modeling Volatility
   - GARCH Models
   - GARCH-M models

# GARCH-M models

- Advantage of Maximum Likelihood is its flexibility
- It allows not only for time-varying volatility, but also for repercussions of volatility on the mean of the model

$$y_t = \alpha + \beta x_t + \gamma h_{t-1} + u_t$$

$$u_t = \varepsilon_t \sqrt{h_t} \qquad \textit{and} \qquad h_t = c + \alpha \varepsilon_{t-1}^2 + b h_{t-1}$$

# Risk adjusted uncovered interest parity

- Utility function (mean variance approach!)

$$E(U(W_t)) = x(E(s_{t+1}) - s_t + (i_t^* - i_t)) - \frac{\eta x^2 \sigma_t^2}{2} \rightarrow \text{max!}$$

- First order condition is

$$x = \frac{E(s_{t+n}) - s_t + (i_t^* - i_t)}{\eta \sigma_t^2}$$

# Risk adjusted uncovered interest parity

- Rearranging gives

$$E(s_{t+1}) - s_t = (i_t - i_t^*) + x\eta\sigma_t^2$$

- Assuming Rational Expectations

$$s_{t+1} - s_t = (i_t - i_t^*) + x\eta\sigma_t^2 + u_{t+1} \qquad \rho_t = x\eta\sigma_t^2$$

- The time-varying risk premium $\rho_t$ depends on
  - the exposure x
  - the CARA coefficient $\eta$
  - the exchange rate volatility

# An empirical model of FX risk premia

- As far the contribution of the exposure x is concerned...
- ... Frankel (1986) estimated that in order to raise the risk premium by 1% the share of dollar assets in the world portfolio must rise by 50%
- ... it is suggested that portfolio shares do not exhibit substantial volatility
- Most researchers rule out foreign currency exposures as a major driving force of FX premia
- More promising: Exchange rate volatility

# An empirical model of FX premia

- A model incorporating GARCH effects in the return equation has been introduced by Engle, Lilien and Robins (1987)
- The so-called GARCH-M model:

$$s_t - s_{t-1} = \beta_0 + \beta_1(i_{t-1} - i^*_{t-1}) + \beta_2 h_t + u_t$$

$$u_t = \varepsilon_t \sqrt{h_t}$$

$$h_t = c + \alpha\varepsilon^2_{t-1} + bh_{t-1}$$

- Of course, we expect $\beta_0 = 0, \beta_1 = 1$ *and* $\beta_2 > 0$!
- The parameter estimates are again calculated by numerical maximization of the log likelihood

# Applied Econometrics of FX Markets

## 7. Modeling Nonlinearities I: Markov-Switching

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Recap - GARCH-M model of FX premia

- A model incorporating GARCH effects in the return equation has been introduced by Engle, Lilien and Robins (1987)
- The so-called GARCH-M model:

$$s_t - s_{t-1} = \beta_0 + \beta_1(i_{t-1} - i^*_{t-1}) + \beta_2 h_t + u_t$$

$$u_t = \varepsilon_t \sqrt{h_t}$$

$$h_t = c + \alpha\varepsilon^2_{t-1} + bh_{t-1}$$

- Of course, we expect $\beta_0 = 0, \beta_1 = 1 \; and \; \beta_2 > 0$!
- The parameter estimates are calculated by numerical maximization of the log likelihood

# Outline

1. Modeling Nonlinearities: Markov-switching models
   - Segmented means and volatilities
   - Application to UIP

# Segmented means and volatilities

- Influential paper by Engel and Hamilton (1990)
- The stochastic process of exchange rate return may be very simple if we allow the process to switch between two or more regimes
- For instance, it might be the case that in some periods of the sample the return is driven by

$$\Delta s_t \sim N(\mu_{1t}, h_{1t})$$

while in other periods it is driven by

$$\Delta s_t \sim N(\mu_{2t}, h_{2t})$$

# Segmented means and volatilities

- If the $\mu_{it}$ are a positive and negative constant then this model represents the time series of returns as a sequence of ups and downs depending on the prevailing **regime** in a given point in time
- This is in contrast to standard time series representation where it is assumed that the return is always drawn from the same single distribution
- Here: A two stage procedure:
  - First step: Randomly pick a regime (eg. up or down)!
  - Second step: Draw a return from the respective distribution!
- How do we model such a time series ex post?

# Segmented means and volatilities

- First step: The probability to pick a given regime should only depend on the last regime of the process:

$$Pr(I_t = 1 | I_{t-1} = 1) = P$$

$$Pr(I_t = 2 | I_{t-1} = 1) = 1 - P$$

$$Pr(I_t = 2 | I_{t-1} = 2) = Q$$

$$Pr(I_t = 1 | I_{t-1} = 2) = 1 - Q$$

- The regime indicator $I_t$ thus follows a first order Markov process and this is why we call the model Markov RSm
- Transition probabilities P and Q (explaining and forecasting fx)

# Segmented means and volatilities

- To fit the Markov RSm to the data we have to infer the regime of the process in each point of time
- Because we cannot be entirely sure about this we are looking for the (conditional) probability that the observed return is drawn from regime 1 or 2 using all available information:

$$Pr(I_{t-1} = 1 | \Delta s_{t-1}, \Phi_{t-1}) = ?$$

# Segmented means and volatilities

- Hamilton suggests that Bayesian inference is an unbiased and efficient way to infer regime probabilities

$$Pr(I_{t-1}=1|\Delta s_{t-1})=\frac{f(\Delta s_{t-1}|I_{t-1}=1)p_{1t-1}}{f(\Delta s_{t-1}|I_{t-1}=1)p_{1t-1}+f(\Delta s_{t-1}|I_{t-1}=2)(1-p_{1t-1})}$$

- and

$$Pr(I_{t-1}=2|\Delta s_{t-1})=\frac{f(\Delta s_{t-1}|I_{t-1}=2)(1-p_{1t-1})}{f(\Delta s_{t-1}|I_{t-1}=1)p_{1t-1}+f(\Delta s_{t-1}|I_{t-1}=2)(1-p_{1t-1})}$$

- where $p_{1t-1}$ and $p_{2t-1}=1-p_{1t-1}$ are called prior probabilities

# Segmented means and volatilities

- Using the posteriors we can calculate an expectation of the next periods regime probability as

$$p_{1t} = P\ Pr(I_{t-1} = 1|\Delta s_{t-1}) + (1-Q)\ Pr(I_{t-1} = 2|\Delta s_{t-1})$$

- or

$$p_{1t} = P\left[\frac{f_{1t-1}p_{1t-1}}{f_{1t-1}p_{1t-1} + f_{2t-1}(1-p_{1t-1})}\right] + (1-Q)\left[\frac{f_{2t-1}(1-p_{1t-1})}{f_{1t-1}p_{1t-1} + f_{2t-1}(1-p_{1t-1})}\right]$$

- and

$$p_{2t} = 1 - p_{1t}$$

# Segmented means and volatilities

- This implies that we always rely on t-1 information why the routine is termed 'recursive'!
- Because time t-1 probabilities depend on time t-2 probabilities and so on the routine incorporates all information (exchange rate returns, density parameter, probability priors) up to time t-1!

# Segmented means and volatilities

- Second step: Specification of the regime dependent distribution

$$f(\Delta s_{t-1}|I_{t-1} = 1, 2, \Phi)$$

- Simplest case:

$$\mu_{1t} = c_1;\ \mu_{2t} = c_2$$

$$h_{1t} = \sigma_1^2;\ h_{2t} = \sigma_2^2$$

- Alternatively: e.g.

$$\mu_{1t} = c_1 + \beta_1 \Delta s_{t-1};\ \ \mu_{2t} = c_2 + \beta_2 \Delta s_{t-1}$$

$$h_{1t} = b_{01} + b_{11} u_{t-1}^2 + b_{21} h_{1t-1}$$

$$h_{2t} = b_{02} + b_{12} u_{t-1}^2 + b_{22} h_{1t-1}$$

- Estimation of the model by maximizing the log-likelihood:

$$L = \sum_{t=1}^{T} log \left[ p_{1t} \frac{1}{\sqrt{2\pi h_{1t}}} exp(\Theta_1) + (1 - p_{1t}) \frac{1}{\sqrt{2\pi h_{2t}}} exp(\Theta_2) \right]$$

- and

$$\Theta_1 = \frac{-(\Delta s_t - \mu_{1t})^2}{2h_{1t}}, \Theta_2 = \frac{-(\Delta s_t - \mu_{2t})^2}{2h_{2t}}$$

# Outline

1. Modeling Nonlinearities: Markov-switching models
   - Segmented means and volatilities
   - Application to UIP

# Application to UIP

- We may suggest that UIP holds in 'normal' times, but is violated in 'non-normal' times
- The sequences of switching between the two states may help us learning what drives the UIP puzzle
- Therefore we need two separate mean equations to be estimated

$$Regime1 : \mu_{1t} = \alpha_1 + \beta_1(i_{t-1} - i^*_{t-1});\ h_{1t} = \sigma_1^2$$
$$Regime2 : \mu_{2t} = \alpha_2 + \beta_2(i_{t-1} - i^*_{t-1});\ h_{2t} = \sigma_2^2$$

- Coefficients estimated by maximizing the Log Likelihood as specified above

# Application to UIP

Application of the MSM in R

- The Estimation is done using R's package 'MSwM'
- First: Generate the standard linear model
- Second: Hand over the linear model to the MS Framework
- Third: Calibrate and estimate the Markov Switching Model
- Fourth: Check diagnostics

# Applied Econometrics of FX Markets

## 8. Modeling Nonlinearities II: STAR-Models

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Outline

1. Modeling Nonlinearities: STAR-models
   - Threshold regression models
   - Smooth transition regression models

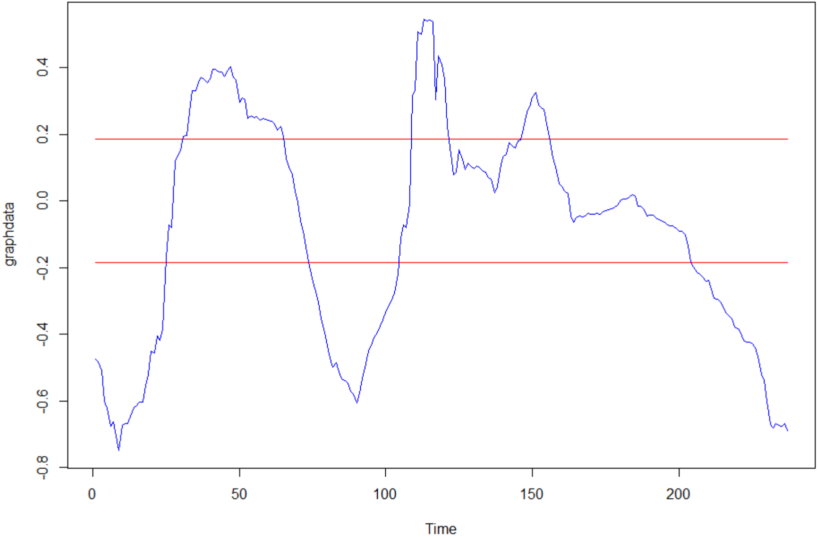
# Threshold regression models

- Literature: Enders (2010), chapter 7.4
- The Markov switching framework allowed for different regimes of the stochastic process, but it does not allow the switching to be informative (at least in the above simply version)
- It is often the case that economists have an idea what variable currently determines the regime
- For instance, the mean reversion of real exchange rate may depend on current misalignments (Kilian and Taylor, 2003)

# Threshold regression models

- Sarno et al. (2006) argue that FX trading will not react to every tiny interest differential, because:
  - transaction costs are too large
  - limits to speculation are binding (expected Sharpe ratios may be too small compared to other risky activities)
- Result: Exchange rate reaction is regime dependent
  - Large differentials: UIP works
  - Small differentials: UIP does not work

# Threshold regression model



# Threshold regression model

- The econometric model is called the Self-Exiting Threshold AutoRegressive (SETAR) model or TAR model (for short)
- Not restricted to AR models
- TAR models have been introduced by Tong (1983, 1990) and developed further by Chan (1993)

# Threshold regression model

- The econometric specification we use to demonstrate the working of TAR models is

$$\Delta s_t = \alpha + \beta^{small}(i_{t-1} - i^*_{t-1})\, I(|i_{t-1} - i^*_{t-1}| < \tau) + \beta^{large}(i_{t-1} - i^*_{t-1})\, I(|i_{t-1} - i^*_{t-1}| \geq \tau) + \varepsilon_t$$

  where $\tau$ is the threshold and $I(.) \in \{0, 1\}$ is an indicator variable.
- Hopefully, we find that at least $\beta^{large}$ is in line with UIP!

# Threshold regression model

- If the threshold is known, the model boils down to a 'piecewise' OLS regression and coefficients are consistent conditional on the threshold being being 'correct'.
- What if the threshold is unknown?
- Super-consistent estimate of $\tau$ according to Chan (1993).

# Chan procedure to estimate $\tau$

- Order the threshold variable from small to large
- Discard the smallest and largest 15% to ensure sufficient observations for either regime
- Estimate the model for every single threshold
- Pick $\tau$ from the regression producing the lowest RSS
- Check the performance of the model by AIC/BIC

# Threshold regression model

```
##############################################    TAR Model ################################
low <- 1 + as.integer(0.15*T)
high <- T - as.integer(0.15*T)
rss_min <- 10000000000000
rss <- matrix(0,T,1)
flag <- matrix(0,T,1)
thresh <- sort(abs(x))
tau <- thresh[low]

for(i in low:high){

  for(j in 1:T){
    if(abs(x[j]) < thresh[i]){flag[j] <- 0} else {flag[j] <- 1}
  }

  outer_x <- flag*x
  inner_x <- (1-flag)*x
  X <- cbind(outer_x,inner_x)
  tarmodel <- lm(y ~ X)
  rss[i] <- cumsum(tarmodel$residuals^2)[T]

  if(rss[i] < rss_min){
    rss_min <- rss[i]
    tau <- thresh[i]
  }

}
```

# Outline

1. Modeling Nonlinearities: STAR-models
   - Threshold regression models
   - Smooth transition regression models

# Applied Econometrics of FX Markets

## 9. Cross-Sectional Analysis of Currency Returns

Stefan Reitz

# Course Outline

1. Basic Econometric Concepts I
2. Basic Econometric Concepts II
3. Basic Time Series Models of Exchange Rates
4. Modeling Trends: Unit Roots in Time Series
5. Testing UIP Conditions
6. Modeling Volatility
7. Modeling Nonlinearities I: Markov-Switching
8. Modeling Nonlinearities II: STAR-Models
9. Cross-Sectional Analysis of Currency Returns

# Recap:

- Profitability of currency carry trades remains a major challenge in international finance
- Time series regressions revealed very little explanatory power of risk adjustments to uncovered interest parity for specific exchange rates

# Recap:

- Profitability of currency carry trades remains a major challenge in international finance
- Time series regressions revealed very little explanatory power of risk adjustments to uncovered interest parity for specific exchange rates
- The more recent literature asks a slightly different question: Why do some currencies earn a higher average return than others? Is there a risk-based explanation for systematic return differences across currencies?
- This brings FX market research closer to standard asset pricing in the tradition of Fama/French and Fama/MacBeth

# Theoretical groundwork

- Consumption-based asset pricing starts with a standard micro utility function
$$U(C_t, C_{t+1}) = U(C_t) + \beta E_t[U(C_{t+1})],$$
where $C_t$ – consumption, and $\beta$ – subjective discount factor (impatience)

# Theoretical groundwork

- Consumption-based asset pricing starts with a standard micro utility function
$$U(C_t, C_{t+1}) = U(C_t) + \beta E_t[U(C_{t+1})],$$
where $C_t$ – consumption, and $\beta$ – subjective discount factor (impatience)

- Representative household chooses $\xi$ in order to
$$\max_{\xi} \; U(C_t, C_{t+1}) \; s.t.$$
$$C_t = Y_t - P_t \xi$$
$$C_{t+1} = Y_{t+1} + X_{t+1}\xi,$$
where $Y_t$ – income, $P_t$ – asset price, $\xi$ – number of assets, and $X_{t+1} = P_t(1 + r_{t+1})$ – asset pay off.

# Theoretical groundwork

- First order condition is

$$P_t U'(C_t) = E_t[\beta U'(C_{t+1}) X_{t+1}]$$

# Theoretical groundwork

- First order condition is

$$\underbrace{P_t U'(C_t)}_{loss\ \ of\ \ utility} = \underbrace{E_t[\beta U'(C_{t+1}) X_{t+1}]}_{exp.\ gain\ in\ utility}$$

# Theoretical groundwork

- First order condition is

$$\underbrace{P_t U'(C_t)}_{loss\ \ of\ \ utility} = \underbrace{E_t[\beta U'(C_{t+1})X_{t+1}]}_{exp.\ gain\ \ in\ \ utility}$$

or

$$P_t = E_t[\beta \frac{U'(C_{t+1})}{U'(C_t)} X_{t+1}]$$

# Theoretical groundwork

- First order condition is

$$\underbrace{P_t U'(C_t)}_{loss\ \ of\ \ utility} = \underbrace{E_t[\beta U'(C_{t+1}) X_{t+1}]}_{exp.\ \ gain\ \ in\ \ utility}$$

or

$$P_t = E_t[\beta \frac{U'(C_{t+1})}{U'(C_t)} X_{t+1}]$$

- The asset pricing literature defines the *stochastic discount factor* as

$$M_{t+1} = \beta \frac{U'(C_{t+1})}{U'(C_t)}$$

# Theoretical groundwork

- This implies

$$P = E_t[M_{t+1}X_{t+1}],\ \textit{or}$$

$$1 = E_t[M_{t+1}R_{t+1}],\ \textit{where}\ R_{t+1} = X_{t+1}/P_t$$

# Theoretical groundwork

- This implies

$$P = E_t[M_{t+1}X_{t+1}],\ \ or$$
$$1 = E_t[M_{t+1}R_{t+1}],\ \ where\ \ R_{t+1} = X_{t+1}/P_t$$

- The risk free rate $R^f$ is known in advance so that

$$\begin{aligned} 1 &= E_t[M_{t+1}R^f] \\ &= E_t[M_{t+1}]R^f \\ R^f &= \frac{1}{E_t[M_{t+1}]} \end{aligned}$$

# Theoretical groundwork

- The empirical asset pricing literature typically works with excess returns so that

$$
\begin{aligned}
1 &= E_t[M_{t+1}R_{t+1}] \\
&= E_t[M_{t+1}(R^e_{t+1} + R^f)] \\
&= E_t[M_{t+1}R^e_{t+1}] + E_t[M_{t+1}]R^f \\
&= E_t[M_{t+1}R^e_{t+1}] + 1 \\
0 &= E_t[M_{t+1}R^e_{t+1}]
\end{aligned}
$$

# Theoretical groundwork

- The empirical asset pricing literature typically works with excess returns so that

$$
\begin{aligned}
1 &= E_t[M_{t+1}R_{t+1}] \\
&= E_t[M_{t+1}(R^e_{t+1} + R^f)] \\
&= E_t[M_{t+1}R^e_{t+1}] + E_t[M_{t+1}]R^f \\
&= E_t[M_{t+1}R^e_{t+1}] + 1 \\
0 &= E_t[M_{t+1}R^e_{t+1}]
\end{aligned}
$$

- This is the starting point for most empirical applications

# Theoretical groundwork

- Standard asset pricing starts with (skipping superscripts)

$$0 = E_t[M_{t+1}R_{t+1}]$$

- Assuming a linear factor model ($F_t$ is a vector of demeaned factors)

$$M_t = 1 - b \cdot F_t$$

we have

$$\begin{aligned} 0 &= E_t[R_{t+1}] - bE_t[R_{t+1}F_{t+1}] \\ E_t[R_{t+1}] &= bcov(R_{t+1}, F_{t+1}) + bE_t[R_{t+1}]E_t[F_{t+1}] \\ &= \frac{cov(R_{t+1}, F_{t+1})}{var(F_{t+1})} \cdot \frac{var(F_{t+1})}{b^{-1}} \end{aligned}$$

# Theoretical groundwork

- Standard asset pricing starts with (skipping superscripts)

$$0 = E_t[M_{t+1}R_{t+1}]$$

- Assuming a linear factor model ($F_t$ is a vector of demeaned factors)

$$m_t = 1 - b \cdot F_t$$

we have ($E_t[F_{t+1}] = 0$)

$$\begin{aligned} 0 &= E_t[R_{t+1}] - bE_t[R_{t+1}F_{t+1}] \\ E_t[R_{t+1}] &= bcov(R_{t+1}, F_{t+1}) + bE_t[R_{t+1}]E_t[F_{t+1}] \\ &= \underbrace{\frac{cov(R_{t+1}, F_{t+1})}{var(F_{t+1})}}_{risk\ \ exposure\ \ \beta} \cdot \underbrace{\frac{var(F_{t+1})}{b^{-1}}}_{price\ \ of\ \ risk\ \ \lambda} \end{aligned}$$

# Theoretical Groundwork

Economic Interpretation The so-called beta representation $E_t[R_{t+1}] = \lambda \cdot \beta$ of the asset pricing model has a number of useful interpretations (Cochrane, 2005)

- Coefficient $\beta$ gives the amount of risk or risk exposure
  - $\beta > 1$: Adding a unit of this asset to your portfolio increases your overall risk
  - $\beta < 1$: Adding a unit of this asset to your portfolio lowers your overall risk
  - $\beta < 0$: This asset works as a hedge instrument

# Theoretical Groundwork

Economic Interpretation The so-called beta representation $E_t[R_{t+1}] = \lambda \cdot \beta$ of the asset pricing model has a number of useful interpretations (Cochrane, 2005)

- Coefficient $\beta$ gives the amount of risk or risk exposure
  - $\beta > 1$: Adding a unit of this asset to your portfolio increases your overall risk
  - $\beta < 1$: Adding a unit of this asset to your portfolio lowers your overall risk
  - $\beta < 0$: This asset works as a hedge instrument
- Coefficient $\lambda$ gives the price of risk
  - We need $\lambda$ to be significantly estimated with a positive sign
  - This means the supposed risk factor is in fact considered by investors
  - $\lambda$ reflects risk aversion of investors as well as factor volatility, both might be time varying

# Theoretical Groundwork

Economic Interpretation The so-called beta representation $E_t[R_{t+1}] = \lambda \cdot \beta$ of the asset pricing model has a number of useful interpretations (Cochrane, 2005)

- Coefficient $\beta$ gives the amount of risk or risk exposure
  - $\beta > 1$: Adding a unit of this asset to your portfolio increases your overall risk
  - $\beta < 1$: Adding a unit of this asset to your portfolio lowers your overall risk
  - $\beta < 0$: This asset works as a hedge instrument
- Coefficient $\lambda$ gives the price of risk
  - We need $\lambda$ to be significantly estimated with a positive sign
  - This means the supposed risk factor is in fact considered by investors
  - $\lambda$ reflects risk aversion of investors as well as factor volatility, both might be time varying
- The overall model should explain a large portion of $E_t[R_{t+1}]$

# Empirical Model

- Remember: We are interested in the cross-sectional difference in excess returns
- Thus, we need to investigate K different currencies with time series of of T observations
- In order to estimate $E_t[R_{t+1}] = \lambda \cdot \beta$ the literature typically follows the standard Fama/MacBeth 2–stage routine
  1. Perform K regressions for each of the time series : $R_{i,t} = \alpha_i + \beta_i F_t + e_{i,t}$
  2. Using $E_t[R_{t+1}] \approx \overline{R}$ perform one cross-sectional regression : $\overline{R_i} = \alpha + \lambda\beta_i + u_i$
- Use Shanken errors to calculate $t$ statistics

# Empirical Model

- Instead of referring to single currencies, empirical asset pricing models typically construct (investable) portfolios according to the specific 'investment style'
- Investment style here is the well-know carry trade

# Empirical Model

- Instead of referring to single currencies, empirical asset pricing models typically construct (investable) portfolios according to the specific 'investment style'
- Investment style here is the well-know carry trade
- Test assets ($R_t^i$)
  1. Following Lustig et al. (2011) we use 48 exchange rate series against the US dollar (US investor perspective)
  2. Monthly data from Jan 1984 to Feb 2017
  3. Sort exchange rates into five buckets ranked according to their interest differential against the US dollar $i_t^i - i_t \approx f_t^i - s_t^i$
  4. By weighting exchange rates equally we calculate excess returns $R_t^i$ for five test assets

# Empirical Model

- Factor construction follows a principle component analysis of Lustig et al. (2011) revealing that the first two principle components explain more than 80% of the portfolio returns

# Empirical Model

- Factor construction follows a principle component analysis of Lustig et al. (2011) revealing that the first two principle components explain more than 80% of the portfolio returns
- Factor construction 1 ($F_t^1$)
    1. Following Lustig et al. (2011) we use the average return of all 48 exchange rate series against the US dollar similar to the market rate what is used in stock market application
    2. This gives a kind of a CAPM interpretation (DOL)

# Empirical Model

- Factor construction follows a principle component analysis of Lustig et al. (2011) revealing that the first two principle components explain more than 80% of the portfolio returns
- Factor construction 1 ($F_t^1$)
  1. Following Lustig et al. (2011) we use the average return of all 48 exchange rate series against the US dollar similar to the market rate what is used in stock market application
  2. This gives a kind of a CAPM interpretation (DOL)
- Factor construction 2 ($F_t^2$)
  1. Following Lustig et al. (2011) we use the return difference between high-yielding currencies and low-yielding currencies $F_t^2 = R_t^5 - R_t^1$
  2. This gives a kind of a High-minus-Low interpretation (HML)

# Empirical Model

- Move over to R!
- Use R-code Fama-MacBeth.R
- Use data set Fama-MacBeth.csv