Detección y predicción de las burbujas especulativas

Konstantin Arkadievich Kholodilin

DIW Berlin y NRU HSE

Detección y predicción de las crisis económicas

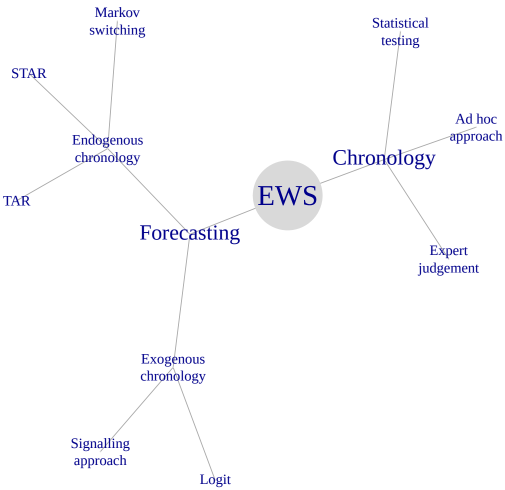
Introduction

- The path of the economy is not smooth.
- From time to time, economy experiences booms and crises.
  - Great Recession 2008–2009 worldwide;
  - Lost Decade (1991–2000 or even 1991–2010) in Japan.
- The crises are costly — they destroy jobs and wealth.
- It is important to be able to foresee the crises.
- Then, their negative consequences can be mitigated.

Early warning system

- A tool for forecasting the booms and crises is called early warning system.
- It has two main elements:
  - expansion/recession (boom/bust) dates, or chronology;
  - forecasting model.

Early warning system



Expansion/recession chronology

- The consequences of crises are very palpable
  - GDP losses,
  - increase in unemployment,
  - bankruptcies, etc.
- Nobody knows when exactly do they start and end.
- Some algorithm is needed to establish the chronology.
- The methods of establishing chronology include:
  - expert judgment;
  - *ad hoc* dating;
  - formal statistical tests.

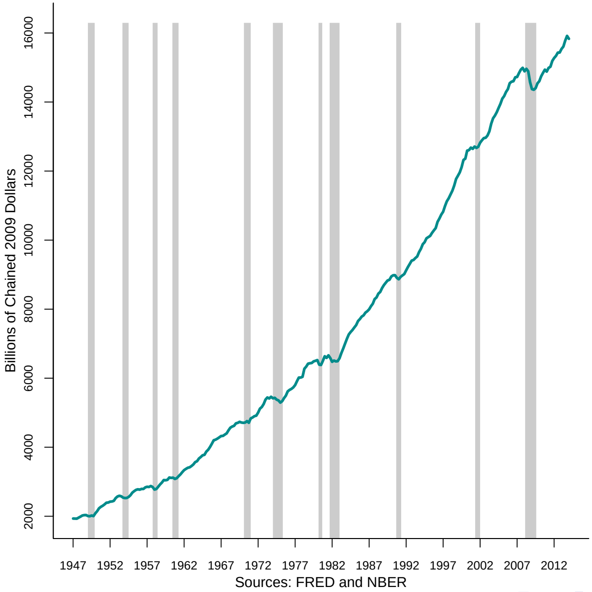
Expert judgment

- A group of experts decides on the dates of beginning and ending of booms and busts.
- The most known such a group is the NBER's Business Cycle Dating Committee: `http://www.nber.org/cycles/recessions.html`
- It maintains a chronology of the U.S. business cycle.
- The chronology comprises alternating dates of peaks and troughs in economic activity.
- *The Committee applies its judgment... and has no fixed rule to determine whether a contraction is only a short interruption of an expansion, or an expansion is only a short interruption of a contraction.*

Expert judgment

- NBER does not have a fixed definition of economic activity.
- It examines and compares the behavior of various measures of broad activity: real GDP, economy-wide employment, and real income.
- NBER also may consider indicators that do not cover the entire economy, such as real sales and the Fed's index of industrial production (IP).
- Well-defined peak or trough in real sales or IP helps to determine overall peak or trough dates,
    - particularly if the economy-wide indicators are in conflict or do not have well-defined peaks or troughs.

US real GDP vs. NBER chronology



Expert judgment

- NBER is an established authority and its chronology is a widely recognized one.
- Other countries have no such expert committee at the national level.
- Economic Cycle Research Institute (ECRI) produces business cycle chronologies for various countries:
  `https://www.businesscycle.com/`

Expert judgement

- Advantages:
  - Experts make use of different sources of information;
  - The wise judgment can be superior to a thoughtless use of sophisticated techniques.
- Disadvantages:
  - The expert's procedure is not transparent;
  - It is impossible to replicate;
  - It is a very time consuming process: it takes many months or even years to identify the cyclical phases.

Ad hoc methods

- The simplest rule of thumb is the two-quarter definition:
  - recession is defined as two consecutive quarters of decline in real GDP.
- A more sophisticated technique is the procedure developed by Bry and Boschan (1971).
- It tries to emulate the decision making process of the NBER using the monthly series of industrial production.

Ad hoc methods

- The Bry-Boschan technique does not use any formal statistical framework to do the dating.
- Instead it translates the NBER method into a set of simple decision rules.
- It basically comprises two stages:
  1. selecting the candidates for turning points and
  2. applying a censoring rule to eliminate the turning points, which don't satisfy some criteria (e.g., minimum duration).

Ad hoc methods

- A quarterly version of Bry-Boschan technique was developed by Harding and Pagan (2001).
  - Available as **R** package *BCDating*.

1. Each point in the sample is checked for being a local maximum or minimum:

$$y_t = \begin{cases} \text{peak, if } y_t = \max\left(y_{t-K}, \ldots, y_{t+L}\right) \\ \text{trough, if } y_t = \min\left(y_{t-K}, \ldots, y_{t+L}\right) \\ \text{neither peak nor trough, otherwise} \end{cases} \tag{1}$$

   Normally, one takes $K = L = 2$ for quarterly data (6 for monthly data).

2. Eliminate candidate peaks and troughs, which don't satisfy two restrictions:
   - ⋆ minimum phase duration should be 6 months (2 quarters)
   - ⋆ complete cycle must last at least 15 months (5 quarters).

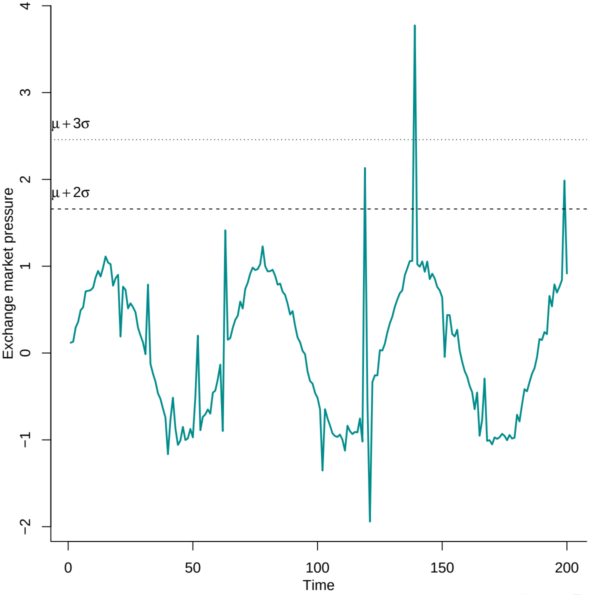
Ad hoc methods

- The most important appeal of the Bry-Boschan technique is its simplicity and transparency.
- It is also very robust in the sense that changing the sample will not affect the dates.
- Although it is sensitive to the choice of criteria and censoring rules.
- Arbitrariness is the major problem related to the *ad hoc* techniques.

Currency crises

- Currency crisis is defined as speculative pressures in foreign exchange markets.
- To identify periods of a currency crisis, constructed crisis indices need to reflect both successful and unsuccessful speculative attacks on domestic currency.
- Following Eichengreen et al. (1994) recent studies use as crisis indicator the exchange market pressure (EMP), which is a weighed average of
  - depreciation rate of nominal exchange rates,
  - percentage change in international reserves,
  - change in interest rate.
- Currency crisis is identified when EMP index exceeds a particular threshold, such as 2 or 3 standard deviations (σ) above its mean (μ).

Currency crises



Formal statistical tests

- Formal statistical tests are free from arbitrariness of both the expert judgment and *ad hoc* methods.
- Recently, they have been increasingly used to detect the asset price speculative bubbles:
  - stock prices,
  - house prices, and
  - commodity prices.
- Phillips et al. (2015) developed an explosive root test to identify multiple speculative bubbles.
  - Available as **R** package *psymonitor*.

Formal statistical tests

- The asset prices can be defined as

$$P_t = \sum_{\tau=0}^{\infty} \left(\frac{1}{1+r_f} \right)^{\tau} E_t(D_{t+\tau} + U_{t+\tau}) + B_t$$

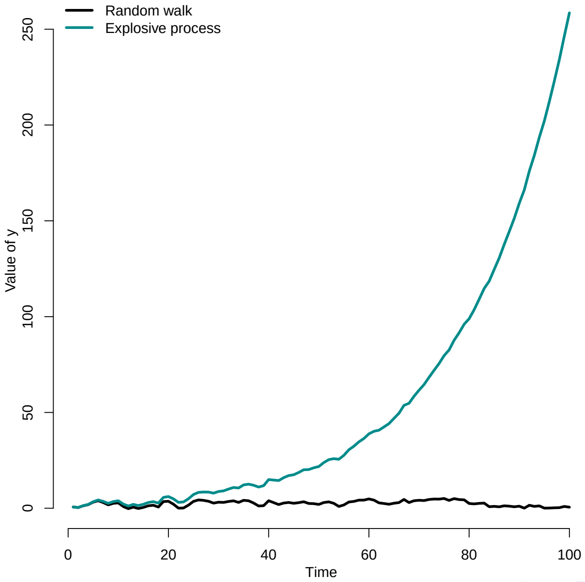
  - P_t = asset price;
  - r_f = risk-free interest rate;
  - D_t = payoff of the asset (e.g., dividends and rents);
  - U_t = unobserved fundamentals;
  - B_t = bubble component, which satisfies

$$E_t(B_{t+1}) = (1+r_f)B_t$$

Speculative bubble test

- In the absence of bubble, the degree of non-stationarity of asset price, P_t, is controlled by that of
  - its payoff and
  - unobserved fundamentals.
- If D_t is I(1) and U_t is I(0), then P_t is I(1).
- In the presence of bubble, P_t will be explosive.

Random walk vs. explosive process



Speculative bubble test

- The basic idea is to analyze the roots of an autoregressive process.
- Explosive root is tested against the alternative of a unit root (random walk).
- The latter reflects the rational expectations hypothesis.

Phillips et al. (2015) test

- The test is based on the rolling regression model:

$$\Delta y_t = \alpha_{r_1,r_2} + \beta_{r_1,r_2} y_{t-1} + \sum_{k=1}^{K} \phi_{r_1,r_2} \Delta y_{t-k} + \varepsilon_t$$

  - y_t — variable to be tested for explosive roots;
  - k — lag order;
  - α, β, ϕ — parameters to be estimated;
  - ε_t — disturbance term.
- Sample of this rolling-window regression starts from the r_1-th fraction and ends at the r_2-th fraction of the total sample (T):
  - $r_2 = r_1 + r_w$ and $r_w > 0$ — (fractional) window size of the regression.

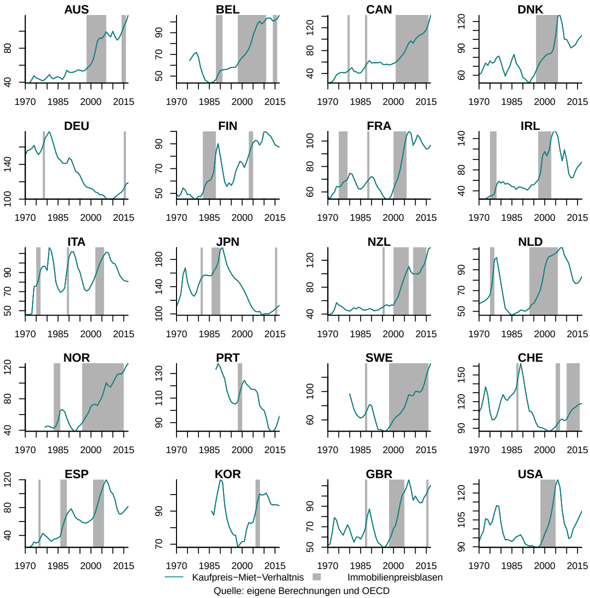
Phillips et al. (2015) test

- Based on this regression, augmented Dickey-Fuller (ADF) test is conducted on a forward expanding sample sequence.
- Phillips et al. (2015) test is the supremum value of the ADF statistic sequence:

$$SADF(r_0) = \sup_{r_2 \in [r_0, 1]} ADF_0^{r_2}$$

  - $ADF_0^{r_2}$ — ADF statistic for a sample running from 0 to r_2.

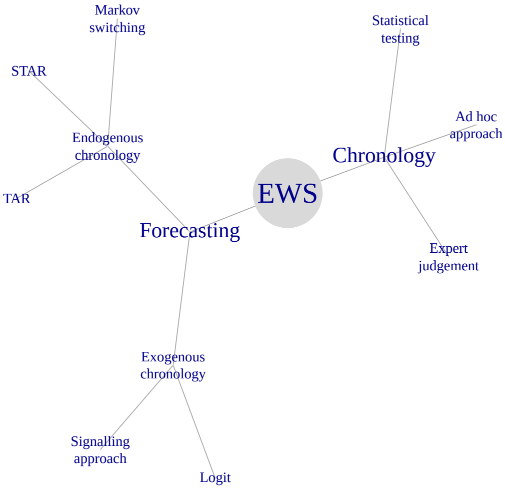
House price-to-rent ratios vs. Phillips et al. (2015) chronology



Early warning system: Exogenous chronology

- Two main elements of an early warning system are:
  - chronology and
  - forecasting model.
- Chronology can be:
  - exogenous when it is not a part of the forecasting model;
  - endogenous when it is obtained when estimating the forecasting model.
- EWS with exogenous chronology can use following forecasting approaches:
  - signalling approach;
  - discrete-choice models.

Early warning system



Signalling approach

- Signalling approach looks for indicators allowing the best prediction of crises
  - Kaminsky et al. (1998); Reinhart and Kaminsky (1999).
- Indicators are chosen s.t. their behavior deviates from "normal" prior to crises.
- Every time an indicator exceeds certain threshold value, it is interpreted as a warning signal that a crisis can occur soon.
  - e.g., threshold = mean plus two standard deviations.

Signalling approach

- Threshold values are calculated so as to strike balance between
  - the risk of having many false alarms and
  - the risk of missing the crisis altogether.
- A reasonable interval between signals and crisis must be defined:
  - any signal given within some period (window) before the beginning of the crisis is labeled a good signal;
  - any other signal sent outside that window is labeled a false alarm or noise.

Indicator performance measure

Table: Detecting crises

	Crisis	No crisis
Signal sent	A	B
Signal not sent	C	D

- Kuipers score:

$$KS = \frac{A}{A+C} - \frac{B}{B+D}$$

  - A and D correct identification;
  - B false alarm;
  - C missing signal

- KS varies between -1 (all false) and 1 (all correct).

Signalling approach

- Individual indicators can be combined to a composite indicator using their KS measures as weights.
- Signalling approach is typically used to predict the currency crises.
- Studies that used it to forecast stock and house price bubbles (Gerdesmeier et al., 2010 and Dreger and Kholodilin, 2013).

Discrete-choice models

- An alternative to signalling approach uses the crisis chronology as dependent variable.
- The chronology is defined as binary, or discrete-choice, variable:

$$C_t = \begin{cases} 1 & \text{if crisis in period } t \\ 0 & \text{otherwise} \end{cases}$$

- In such cases, the discrete-choice models are used:
  - linear probability model;
  - logit and probit models.

Linear probability model

- Probability of being in recession/crisis is defined as:

$$\begin{aligned} \Pr(C_t = 1|x_t) &= F(x_t, \beta) \\ \Pr(C_t = 0|x_t) &= 1 - F(x_t, \beta) \end{aligned} \tag{2}$$

  where
  - x_t is a vector of indicators at time t ($t = 1, \ldots, T$);
  - β is $k \times 1$ parameter vector.
- The problem is to choose a suitable model for the r.h.s. of equation.

Linear probability model

- One possibility is to retain the familiar linear regression:

$$F(x_t, \beta) = x_t'\beta$$

- In that case, we have a linear probability model, which is the multiple linear regression model when the dependent variable, y, is binary rather than continuous:

$$C_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_K x_{Kt} + \varepsilon_t$$

- where C_t is a binary crisis variable.

Linear probability model

- Because dependent variable is binary,

$$E(C_t|x_{1t}, \ldots, x_{Kt}) = \Pr(C|x_{1t}, \ldots, x_{Kt})$$

- So, for the linear probability model:

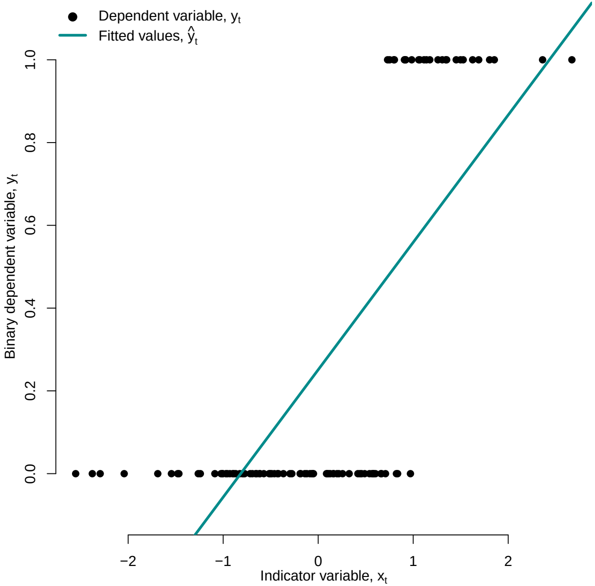
$$\Pr(C_t|x_{1t}, \ldots, x_{Kt}) = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_K x_{Kt}$$

- Regression coefficient β_k on a regressor x_{kt} is the change in the probability that $C_t = 1$ associated with a unit change in x_{kt}.

Linear probability model

- Regression coefficients of linear probability model are estimated by OLS.
- R^2 for a linear probability model makes no sense:
  - When dependent variable is continuous, it is possible to imagine a situation, in which $R^2 = 1$: All the data lie exactly on the regression line.
  - This is impossible, when the dependent variable is binary, unless the regressors are also binary.

Linear probability model



Shortcomings of LPM

- The linearity that makes the linear probability model easy to use is also its major flaw.
- Because probabilities cannot be less than 0 or exceed 1, the effect on the probability that $C_t = 1$ of a given change in x_t must be nonlinear.
- In linear probability model, the effect of a given change in income is constant, which leads to predicted probabilities that drop below 0 or exceed 1. But this is a nonsense!

Logit and probit regression

- Logit and probit regressions are nonlinear regression models specifically designed for binary dependent variables.
  - Available as **R** package *glm*.
- Because regression with binary dependent variable models the probability that $C_t = 1$, it makes sense to adopt a nonlinear form forcing predicted values to lie in $[0, 1]$.
- Because cumulative probability distribution (c.d.f.) functions produce probabilities in $[0, 1]$, they are used in logit and probit regressions.

Logit and probit regression

- The probit model uses the standard normal c.d.f.:

$$\Pr(C_t = 1|x_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_t'\beta} \exp^{-\frac{z^2}{2}} dz = \Phi(x_t'\beta) \tag{3}$$

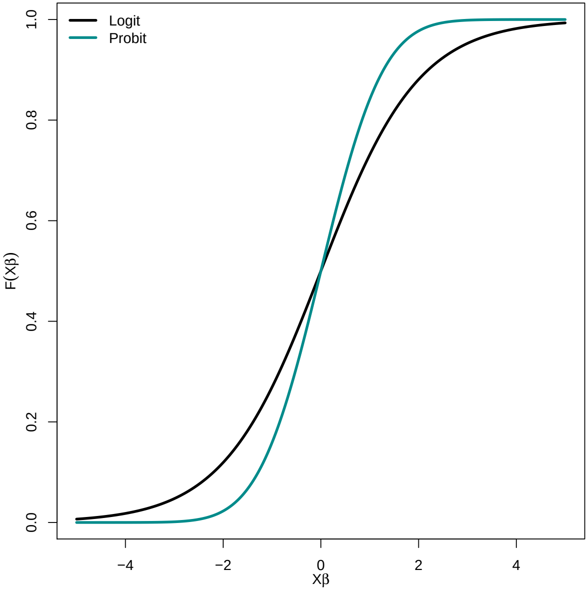
  where Φ is the standard normal distribution function.

- The logit regression uses the logistic c.d.f.:

$$\Pr(C_t = 1|x_t) = \frac{\exp(x_t'\beta)}{1 + \exp(x_t'\beta)} = \Lambda(x_t'\beta) \tag{4}$$

  where Λ is the logistic distribution function.

Logit and probit regression



Comparing alternative models

- What distribution to use?
- Logistic distribution is similar to the normal, except in tails, which are much heavier.
  - It closely resembles a Student distribution with 7 degrees of freedom.
- Hence, for intermediate values of $x_t'\beta$ (say, between -1.2 and +1.2), both distributions tend to give similar probabilities.
- Logistic distribution tends to assign larger (smaller) probabilities to $C_t = 1$, when $x_t'\beta$ is extremely small (very large) than the normal distribution.

Comparing alternative models

- We should expect different predictions from these models, if the sample contains
  1. very few responses (C_t's equal to 1) or very few nonresponses (C_t's equal to 0);
  2. very wide variation in an important independent variable, particularly if (1) is also true.
- It is virtually impossible to justify the choice of one distribution or another on theoretical grounds.

Comparing alternative models

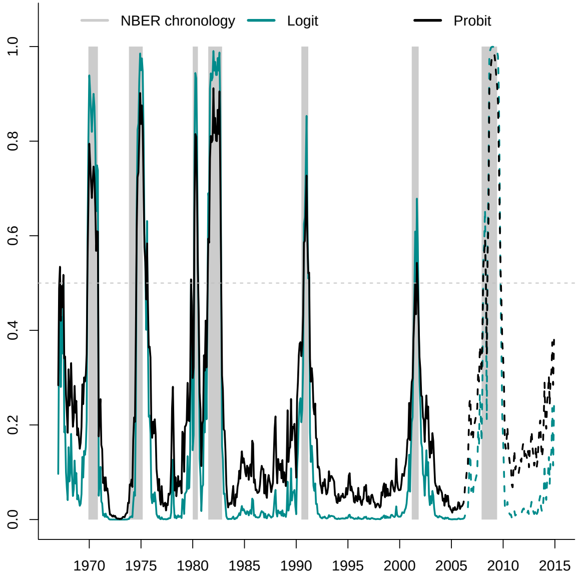
- All three models — linear probability, logit, and probit — are just approximations to the unknown population regression function $F(C_t|x_t) = \Pr(C_t = 1|x_t)$.
- Linear probability model is the easiest to use and to interpret, but it cannot capture the nonlinear nature of the true population regression function.
- Logit and probit regressions model this nonlinearity in probabilities, but their regression coefficients are more difficult to interpret.
- So which should be used in practice?
- No single right answer; different researchers use different methods.

Comparing alternative models

Table: Estimation results of LPM, logit, and probit

	Dependent variable:		
	NBER recession/expansion chronology		
	LPM	Logit	Probit
	(1)	(2)	(3)
Constant	6.146*** (0.507)	87.420*** (11.480)	41.838*** (5.432)
Build_permits	−0.0004*** (0.00004)	−0.006*** (0.001)	−0.003*** (0.0004)
Tsec_10year	−0.007 (0.009)	−0.051 (0.136)	−0.046 (0.071)
URate	−0.168*** (0.018)	−2.602*** (0.380)	−1.256*** (0.181)
TCU	−0.061*** (0.005)	−0.913*** (0.123)	−0.433*** (0.058)
Div_yield	18.258*** (2.033)	234.106*** (41.701)	119.487*** (20.002)
Observations	468	468	468
Log Likelihood	−52.974	−82.122	−85.157
Akaike Inf. Crit.	117.948	176.244	182.313
Note:			*p<0.1; **p<0.05; ***p<0.01

Comparing alternative models



Machine-learning methods (Aprendizaje automático)

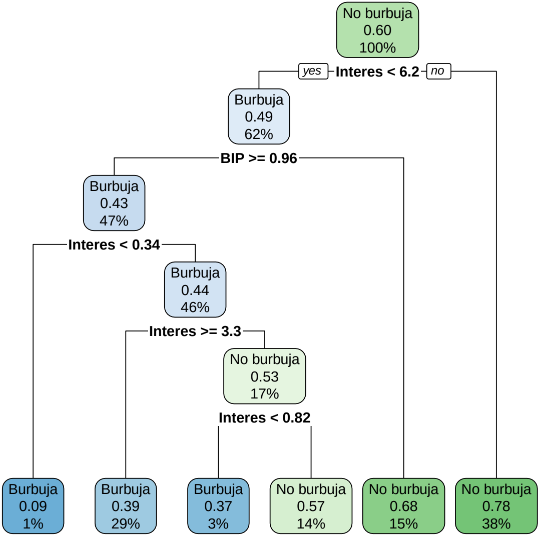
- Machine-learning methods can be useful in forecasting crises.
- Recently, large progress has been made.
- Here, the following methods are considered:

Method	Reference	**R** package
Decision tree	Breiman (2017)	*rpart*
Random forest	Breiman (2017)	*randomForest*
Support vector machine (SVM)	Cortes and Vapnik (1995)	*e1071*

Decision tree

- Decision tree asks questions and classifies period based on answers.
- For each variable, a threshold is recursively chosen.
- For the threshold, the deviations of forecasts from actual values are minimized.
- Graphically, it looks like a tree.
- The total sample is divided by thresholds in non-overlapping subsamples.

Example of decision tree



Decision tree algorithm

1. Calculate all of the Gini impurity scores:

$$GI = 1 - P(\text{crisis})^2 - P(\text{no crisis})^2$$

2. If node has the lowest impurity, then there is no point in separating the regimes anymore and it becomes a leaf node.
3. If separating data results in an improvement (lower impurity), then pick the separation with the lowest impurity value.

Decision tree algorithm: numeric data

1. Sort the indicator values.
2. Compute the averages for all adjacent values.
3. For each average, compute Gini impurity score.
4. Choose the average that hase the lowest score.
5. This value will be the cutoff.

Decision tree

- Advantages:
  - simple to calculate;
  - not so time consuming.
- Disadvantage.
  - Overfitting: in-sample performance is superior to out-of-sample performance.

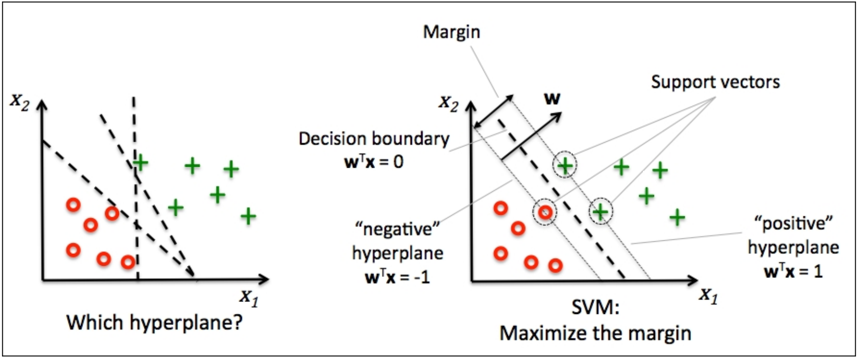
Random forest

- Random forest is closely related to decision tree.
- It allows to substantially reduce overfitting.
- The sample is divided in multiple subsamples.
  - Bootstrap with repetition.
- For each subsample, a decision tree is built.
- Random forest is a combination of all these trees.

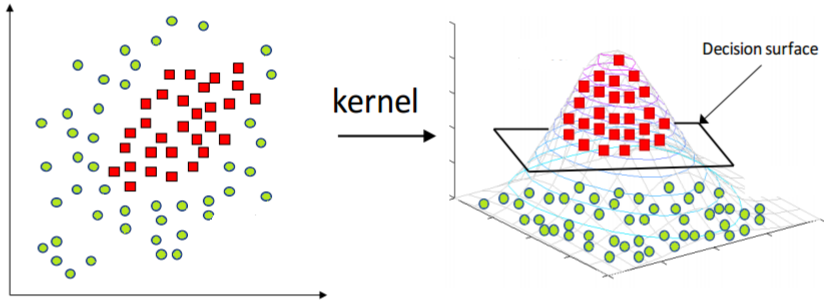
SVM

- SVM separates regimes (bubble vs. non-bubble) by a hyperplane (decision boundary).
- The best hyperplane leaves the maximum margin between each regime:
  - total distance between the extreme elements of both regimes and hyperplane is the largest.
- Two hyperparameters are used:
  - cost parameter C, to penalize misclassification, and
  - parameter γ restricts the model complexity.
- Kernel is used to transform original data and facilitate classification.

Support vector machine



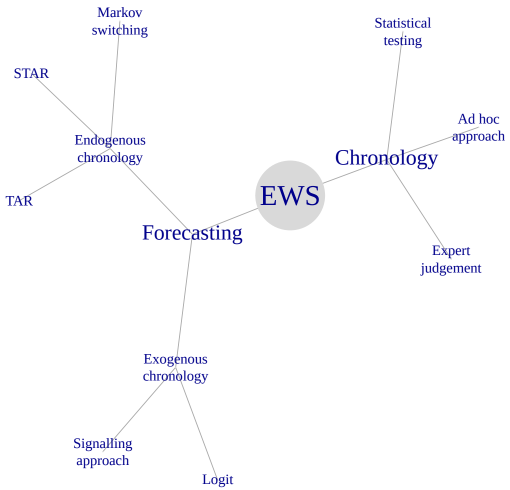
Support vector machine: kernel trick



SVM

- Advantages:
  - always finds global maximum;
  - suitable for both linear and non-linear classification (kernel trick);
  - good for high-dimensional data;
  - works well on small data sets.
- Disadvantages:
  - selection of the right kernel and parameters is ad hoc;
  - not suitable for very large samples;
  - less effective on noisier datasets with overlapping classes.

Early warning system



Early warning system: Endogenous chronology

- This class of models allows moving away from *a priori* dating of crises.
- Using these models, crises can be identified endogenously, as switches between regimes are conditioned on some
  - observable variable — TAR and STAR;
  - latent variable — Markov switching.
- Available as **R** packages:
  - TAR and STAR — *tsDyn*;
  - Markov switching — *MswM*.

TAR

- In threshold autoregressive model (TAR) of Tong (1978, 1990), regime shifts in the dependent variable, y_t, are triggered by an observable transition variable, x_t, crossing a threshold c:

$$y_t = \left(\alpha_0 + \sum_{p=1}^{P} \alpha_p y_{t-p}\right) I(x_t; c) + \left(\beta_0 + \sum_{p=1}^{P} \beta_p y_{t-p}\right) \left(1 - I(x_t; c)\right) + \varepsilon_t$$

- where
  - α's and β's are the regime-dependent parameters;
  - $I(x_t; c)$ is an indicator function;
  - ε_t is the error term, $\varepsilon_t \sim IID(0, \sigma^2)$.

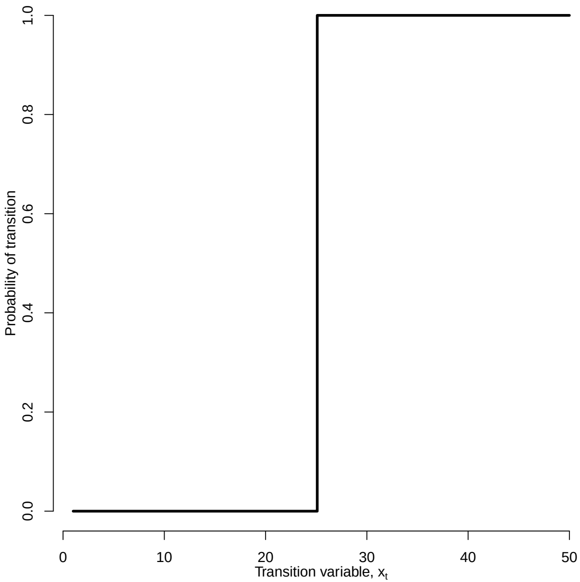
TAR

- Indicator function, $I(x_t; c)$, is defined as:

$$I(x; c) = \begin{cases} 1, & \text{if} \quad g(x_t) > c \\ 0, & \text{if} \quad g(x_t) \leq c \end{cases}$$

- If indicator function of transition variable exceeds certain threshold, then economy is in one regime (e.g., crisis),
- otherwise economy is in another regime (e.g., no crisis).

TAR indicator function



TAR

- The switch between two regimes is discontinuous:
  - an infinitely small change in the transition variable is enough to jump to another regime.
- The economy is seldom confronted with such a situation.
- There can be many different types of transition variable, x_t.
  - For $x_t = t$ a model with a structural break at time $t = c$ occurs.
  - For the analysis of currency and financial crises, other transition variables can be more useful (e.g., leading indicators).

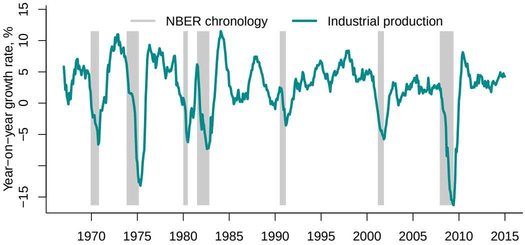
TAR

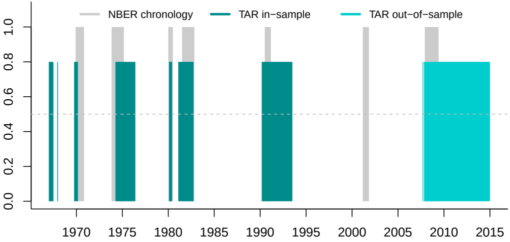
Table: Estimation results of TAR

	Dependent variable:	
	Growth rate of industrial production	
	Recession	Expansion
	(1)	(2)
Constant	-29.316^{***} (5.904)	-11.793^{***} (3.385)
D12LIP_1	0.954^{***} (0.024)	0.917^{***} (0.017)
LBuild_permits	2.853^{***} (0.545)	0.635^{**} (0.314)
Tsec_10year	-0.111^{***} (0.032)	-0.091^{***} (0.031)
URate	0.611^{***} (0.124)	0.279^{***} (0.059)
TCU	0.077 (0.059)	0.079^{***} (0.021)
Observations	190	387
R^2	0.953	0.945
Adjusted R^2	0.952	0.945
Residual Std. Error	1.228 (df = 184)	0.837 (df = 381)
F Statistic	749.851^{***} (df = 5; 184)	$1{,}317.770^{***}$ (df = 5; 381)
Note:		*p<0.1; **p<0.05; ***p<0.01

- Transition variable is a log of building permits.

TAR recession probabilities vs. NBER chronology





SETAR

- If transition variable is a lagged endogenous variable, y_{t-d}, with delay $d > 0$, self-exciting threshold autoregressive (SETAR) model results:

$$y_t = \left(\alpha_0 + \sum_{i=1}^{p} \alpha_p y_{t-i}\right) I(y_{t-d}; c) + \left(\beta_0 + \sum_{i=1}^{p} \beta_p y_{t-i}\right)\left(1 - I(y_{t-d}; c)\right) + \varepsilon_t$$

- where $\varepsilon_t \sim IID(0, \sigma^2)$.
- In SETAR model, regime-generating process is
  - not assumed to be exogenous
  - but directly linked to the lagged endogenous variable y_{t-d}.

STAR

- Smooth transition autoregressive model (STAR), suggested in Teräsvirta and Anderson (1992) and van Dijk et al. (2002), has following form:

$$y_t = \left(\alpha_0 + \sum_{i=1}^{p} \alpha_p y_{t-i}\right) F(x_t; \gamma, c) + \left(\beta_0 + \sum_{i=1}^{p} \beta_p y_{t-i}\right)\left(1 - F(x_t; \gamma, c)\right) + \varepsilon_t$$

- where
  - x_t is the transition variable, which can be
    - lagged dependent variable y_{t-d} with some delay $d > 0$;
    - exogenous observed variable, x_t;
    - (possibly nonlinear) function of lagged exogenous variable $g(x_t)$.
  - $F(x_t; \gamma, c)$ is a smooth transition function.
    - It is a continuous function bounded between 0 and 1.

STAR

- There are two main types of STAR transition function:
- Logistic function:

$$F(x_t;\gamma,c) = \frac{1}{1+\exp[-\gamma(x_t-c)]}, \quad \gamma > 0$$

- Exponential function:

$$F(x_t;\gamma,c) = 1 - \exp[-\gamma(x_t - c)^2], \quad \gamma > 0$$

- where
  - c is the threshold between the regimes;
  - γ is the smoothness parameter: the smaller γ the smoother transition.
- When $\gamma \longrightarrow \infty$, STAR transition function $F(x_t;\gamma,c)$ becomes indicator function of TAR, $I(x_t;c)$.
- When $\gamma \longrightarrow 0$, STAR model converges to a linear AR(p).

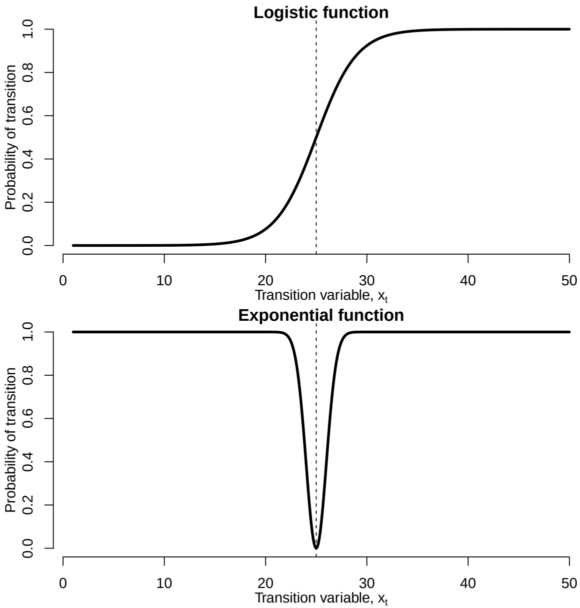
STAR

- STAR model with logistic transition function — logistic STAR (LSTAR).
- The model with exponential transition function — exponential STAR (ESTAR).
- There is no economic theory allowing to choose between LSTAR and ESTAR.
- The choice of transition function, of delay parameter d, and lag structure of the model should be based on the data.

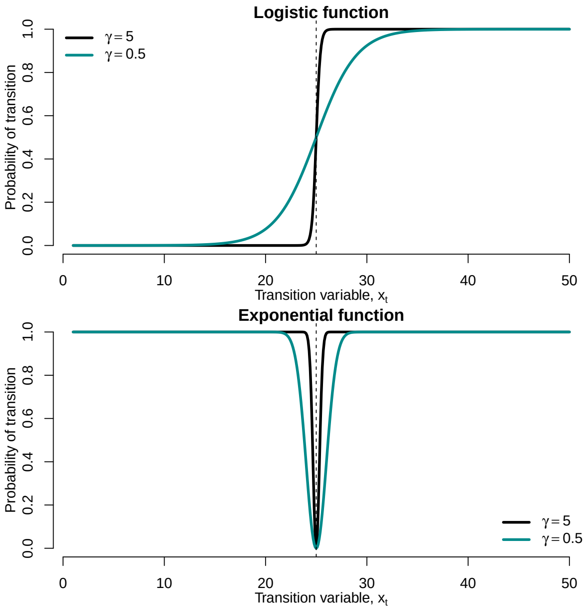
STAR

- STAR model is a regime-switching model
  - allowing for two regimes associated with the extreme values of transition function, $F(x_t; \gamma, c) = 0$ and $F(x_t; \gamma, c) = 1$,
  - where transition from one regime to the other is smooth.
- The regime that occurs at t can be determined by observed variable x_t and corresponding value of $F(x_t; \gamma, c)$.

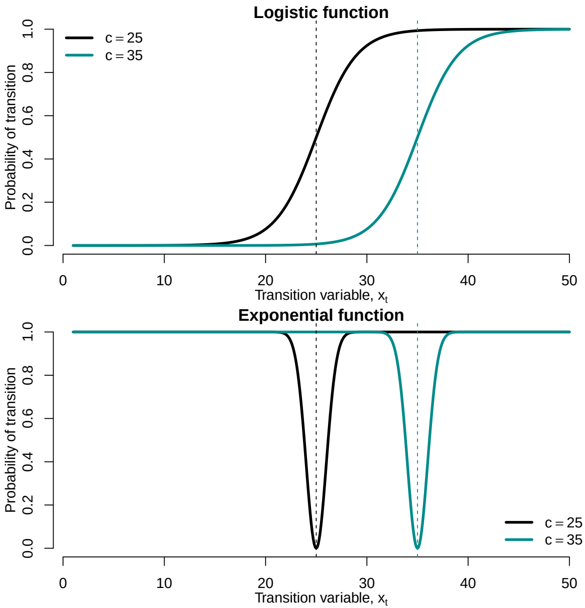
Logistic and exponential transition functions



Smoothness parameter γ



Threshold parameter c



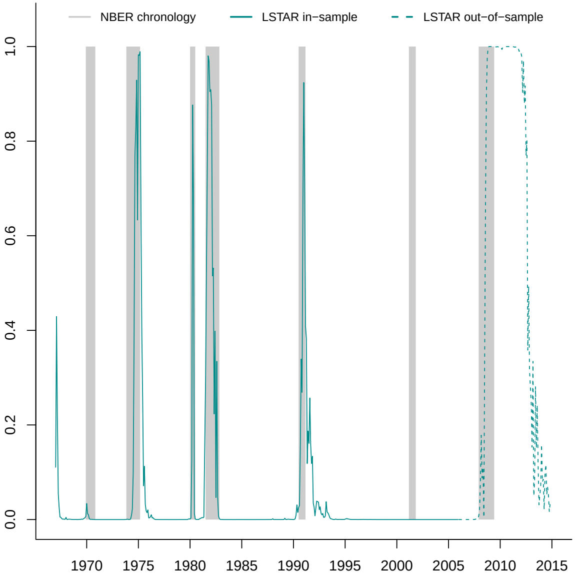
STAR

- These transition functions describe the switches between regimes differently.
- In LSTAR model, regimes are associated with small and large values of transition variable relative to threshold c:
  - when x_t is small, regime 1 dominates,
  - when x_t is large, regime 2 dominates.
- Hence, LSTAR suits well for the analysis of business cycles, where one regime corresponds to recessions, while another one — to expansions.

STAR

- In ESTAR model, regimes are associated with small and large absolute deviations of transition variable from threshold c:
  - when x_t is either too small or too high ($|x_t - c|$ is large), regime 1 occurs,
  - when x_t is close to the threshold ($|x_t - c|$ is small), regime 2 takes place.
- Hence, ESTAR can be useful in currency crisis framework:
  - no-crisis regime occurs, when exchange rate is within certain interval;
  - there is a crisis, when exchange rate violates upper/lower limit of a specified band of fluctuations.

STAR recession probabilities vs. NBER chronology

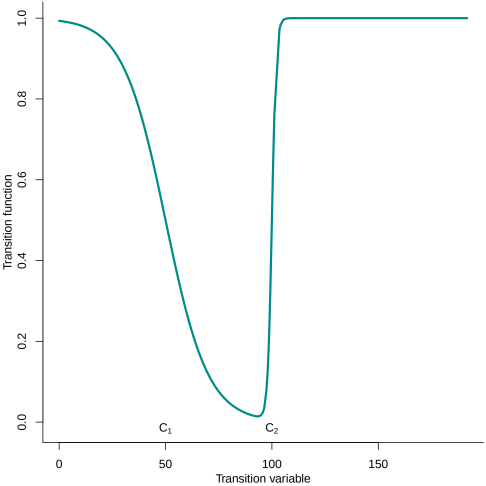


STAR extensions

- Multiple regimes:
  - More than 2 regimes can be identified.
  - Example: expansions, mild recessions, and deep recessions (depressions).
  - Studies: van Dijk and Franses (1999).
- Asymmetric transition function:
  - Switch between regimes can be different depending on the starting point.
  - Example: in recession economy declines abruptly, while in expansion the recovery can be slow.
  - Studies: Siliverstovs (2005).

Bi-parameter STAR of Siliverstovs (2005)

$$F(x_t;\gamma_1,\gamma_2,c_1,c_2)=\frac{\exp[-\gamma_1(x_t-c_1)+\exp[\gamma_2(x_t-c_2)]}{1+\exp[-\gamma_1(x_t-c_1)+\exp[\gamma_2(x_t-c_2)]},\quad \gamma_1>0,\gamma_2>0$$



Markov switching

- Markov-switching models (MS), like TAR and STAR, allow to diminish the degree of arbitrariness by estimating the crisis chronology and not relying on the *ad hoc* definitions.
- Unlike in TAR and STAR, however, the Markov switching implies the existence of an unobserved state variable, which governs switches between regimes.

Markov switching

- One of the first MS models — Hamilton (1989).
- It fostered a great interest in MS models as a tool to characterize macroeconomic fluctuations.
- Large # of extensions and refinements (Krolzig, 1997).
- Basic Markov-switching model:

$$y_t = \mu(s_t) + \sum_{p=1}^{P} \alpha_p y_{t-p} + \varepsilon_t$$

where

  - $\varepsilon_t \sim NID(0, \sigma^2)$;
  - $\mu(s_t)$ switches between two states:

$$\mu(s_t) = \begin{cases} \mu_1 < 0, & \text{if} \quad s_t = 1 \quad (\text{“expansion”}) \\ \mu_2 > 0, & \text{if} \quad s_t = 2 \quad (\text{“recession”}) \end{cases}$$

Markov switching

- In this simple model, the variance of the disturbance term, σ_ε^2, is assumed to be identical in both regimes.
- In more sophisticated models, all parameters can be regime-dependent:
  - intercept (or mean), $\mu(s_t)$;
  - autoregressive coefficients, $\alpha_p(s_t)$;
  - residual variance, $\sigma_\varepsilon^2(s_t)$.

Markov switching

- Contractions and expansions are modelled as switching regimes of unobserved process generating the growth rates of dependent variable.
- The regimes are associated with different conditional distributions of the growth rate of y_t, where, for example,
  - the intercept is positive (or bigger) in expansion and
  - negative (or lower) in recession.

Markov switching

- When the financial and currency crises are considered, two states can be defined differently.
- Stock market:
  - for stock exchange index the low values of intercept $\mu(s_t)$ represent "crashes", while high values — "booms".
- Foreign exchange market / currency crises:
  - "high volatility" is associated with crises, while "low volatility" is associated with normal regimes.
  - state-dependent residual variance $\sigma^2_\varepsilon(s_t)$ is needed.

Endogenous chronology: pros and cons

- Advantage:
  - these models reduce arbitrariness in defining crisis periods.
- Disadvantages:
  - they are computation-intensive and unstable;
  - data requirements are higher.

Forecast performance measures

- Fraction correctly predicted, or Ben-Akiva and Lerman' measure of fit, is based on how well the discrete-choice model predicts the binary dependent variable, C_t:

$$FCP = \frac{\sum_{t=1}^{T} \left(C_t \hat{F}_t + (1 - C_t)(1 - \hat{F}_t) \right)}{T}$$

  which is the average proportion of correct prediction by the forecasting model.
- $\hat{F}_t$ is predicted probability of crisis.
- The difficulty in this computation is that in unbalanced samples, the less frequent outcome will usually be predicted very badly by the standard procedure, and this measure does not pick up that point.

Forecast performance measures

- Cramer (1999) suggested an alternative measure that addresses it:

$$\lambda = E(\hat{F}_t|C_t = 1) - E(\hat{F}_t|C_t = 0)$$
$$= \frac{\sum_{t=1}^{T} C_t \hat{F}_t}{T_1} - \frac{\sum_{t=1}^{T} (1 - C_t)\hat{F}_t}{T_0}$$

where
    - T_1 is number of periods, for which $C_t = 1$,
    - T_0 is number of observations, for which $C_t = 0$.
- The first r.h.s. term is the average $\hat{F}_t$ for $C_t = 1$, while the second r.h.s. term is the average $\hat{F}_t$ for $C_t = 0$.
- Cramer's measure heavily penalizes the incorrect predictions, and since each proportion is taken within the subsample, it's not unduly influenced by large size of the group of more frequent outcomes.

Forecast performance measures

- Log probability score:

$$LPS = -\frac{1}{T}\sum_{t=1}^{T}\left[(1-\hat{F}_t)\ln(1-C_t) + \hat{F}_t \ln(C_t)\right]$$

- Quadratic probability score:

$$QPS = \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t - C_t)^2$$

  - QPS varies between 0 (perfect performance) and 1 (bad performance).

- Kuipers score:

$$KS = \frac{A}{A+C} - \frac{B}{B+D}$$

References I

Breiman, L. (2017). *Classification and regression trees.* Routledge.

Bry, G. and C. Boschan (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs.* NBER Books. National Bureau of Economic Research, Inc.

Cortes, C. and V. Vapnik (1995). Support-vector networks. *Machine learning 20*(3), 273–297.

Cramer, J. S. (1999). Predictive performance of the binary logit model in unbalanced samples. *Journal of the Royal Statistical Society: Series D (The Statistician) 48*(1), 85–94.

Dreger, C. and K. A. Kholodilin (2013). An early warning system to predict speculative house price bubbles. *Economics — The Open-Access, Open-Assessment E-Journal 7*, 1–26.

References II

Eichengreen, B., A. K. Rose, and C. Wyplosz (1994). Speculative attacks on pegged exchange rates: An empirical exploration with special reference to the European monetary system. NBER Working Papers 4898, National Bureau of Economic Research, Inc.

Gerdesmeier, D., H.-E. Reimers, and B. Roffia (2010). Asset price misalignments and the role of money and credit. *International Finance 13*(3), 377–407.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–84.

Harding, D. and A. Pagan (2001). Extracting, using and analysing cyclical information. MPRA Paper 15, University Library of Munich, Germany.

Kaminsky, G., S. Lizondo, and C. M. Reinhart (1998). Leading indicators of currency crises. *IMF Staff Papers 45*(1), 1–48.

References III

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*. Springer Verlag.

Phillips, P. C., S. Shi, and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*(4), 1043–1078.

Reinhart, C. M. and G. L. Kaminsky (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review 89*(3), 473–500.

Siliverstovs, B. (2005). The bi-parameter smooth transition autoregressive model. *Economics Bulletin 3*(23), 1–11.

Teräsvirta, T. and H. M. Anderson (1992, Suppl. De). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics 7*(S), S119–36.

References IV

Tong, H. (1978). On a threshold model. In C. Chen (Ed.), *Pattern Recognition and Signal Processing*. Sijhoff & Noordhoff.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press.

van Dijk, D. and P. H. Franses (1999). Modeling multiple regimes in the business cycle. *Macroeconomic Dynamics 3*(03), 311–340.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002). Smooth transition autoregressive models — a survey of recent developments. *Econometric Reviews 21*(1), 1–47.