Towards better banking crisis prediction: Could an automatic variable selection process improve the performance?

Xianglong (Locky) Liu

ACE 2021

Centre of Policy Studies
Victoria University

Locky.liu@vu.edu.au

Introduction

Early Warning Systems (EWS) for systemic banking crises

- Multivariate logit models
- Objective: explanatory or forecasting
- Variables usually pre-selected based on economic intuitions

Issue: variable selection & model over-fitting

- Rare systemic events
- A potentially large set of candidates for explanatory variables
  →E.g., credit: domestic & global growth, domestic & global credit gap, interactions and lags

Noval machine learning method: the Random Forest Model (Tanaka et al., 2016; Alessi and Detken, 2018)

- Mixed evidence on *out-of-sample* performance
- Black-box: difficulty to interpret

Motivation:

- To select the best set of predictors that optimises *out-of-sample* prediction results
- To keep the EWS transparent and interpretable to inform policies

Overview

This study proposes applying the Least Absolute Shrinkage and Selection Operator (LASSO) method with cross-validation to automate the variable selection process of the conventional multivariate logit framework for better predicting systemic banking crises.

LASSO:

- Variable selection method that drops variables with small coefficients and only keep the most important predictors (Tibshirani 1996)

Cross-validation:

- Resampling technique to fine-tune the LASSO method
- Ensure the selected predictors have the best out-of-sample forecasting property

Data

- 23 OECD countries
- Sample: 1970Q1-2018Q3

Evaluation:

- Recursive out-of-sample forecasting exercise (pseudo real-time)

Overview

Results:

- The LASSO method with cross-validation significantly improves the predictive performance of the multivariate logit EWS
- It identifies key early warning indicators for systemic banking crises:
  - Domestic credit growth
  - Domestic and global credit-to-GDP gaps
  - Real house price growth
  - Real effective exchange rate

Contribution:

- Highlight the importance of variable selection for systemic event forecasting
- LASSO with cross-validation
  - Extract important information from variable interactions and lags that may not be easily identified and accesses by the conventional approach
  - Improve the forecasting performance without compromising transparency

Systemic banking crisis data

Systemic banking crisis dataset:

- 23 OECD countries, 1970Q1 to 2018Q3, quarterly
- Sources: Laeven and Valencia (2018) and Detken et al. (2014)

Country	Crisis	Start	End	Start	End	Start	End	Tranquil periods	Crisis periods	Crisis share(%)
Australia	0							195	0	0.00%
Austria	1	2008q4	2012q4					178	17	8.72%
Belgium	1	2008q4	2012q4					178	17	8.72%
Canada	0							195	0	0.00%
Denmark	2	1987q1	1993q4	2008q3	2012q4			149	46	23.59%
Finland	1	1991q3	1995q4					177	18	9.23%
France	2	1994q1	1995q4	2008q1	2009q4			179	16	8.21%
Germany	2	2000q1	2003q4	2008q1	2009q4			171	24	12.31%
Greece	1	2008q1	2012q4					175	20	10.26%
Ireland	1	2008q1	2012q4					175	20	10.26%
Israel	1	1977q1	1984q2					195	0	0.00%
Italy	2	1994q1	1995q4	2008q3	2009q4			181	14	7.18%
Japan	1	1997q4	2001q4					178	17	8.72%
South Korea	1	1997q4	1998q4					190	5	2.56%
Netherlands	1	2008q1	2009q4					187	8	4.10%
New Zealand	0							195	0	0.00%
Norway	1	1991q4	1993q4					186	9	4.62%
Portugal	1	2008q3	2012q4					177	18	9.23%
Spain	2	1982q2	1985q3	2009q2	2013q2			164	31	15.90%
Sweden	2	1990q3	1993q4	2008q3	2009q4			175	20	10.26%
Switzerland	1	2008q3	2009q4					189	6	3.08%
UK	3	1973q4	1975q4	1991q1	1995q2	2007q1	2011q4	148	47	24.10%
US	2	1988q1	1988q3	2007q4	2011q4			175	20	10.26%
Total	29									

Pool of candidate explanatory variables

Candidate risk indicators - capture sources of systemic risk and macro-financial vulnerabilities.

- Based on early warning literature (Behn et al., 2013; Duca and Peltonen, 2013)
- Benchmark specification: Behn et al. (2013)

Credit	**Macro**	**External**
credit growth (A)	real GDP growth	current account/GDP
credit-to-GDP gap (B)	inflation	real effective exchange rate
	short-term interest rate	
	term spread	
Asset	**Global**	**Interaction and lags**
real house price growth	credit growth (C)	A*B
real equity price growth	credit-to-GDP gap (D)	C*D
house-price-to-income ratio	real GDP growth	A*C
house-price-to-income gap	inflation	B*D
house-price-to-rental ratio	real house price growth	Other interaction terms
house-price-to-rental gap	real equity price growth	Lags

The need for variable selection:

- Explanatory variables: a very large pool of candidates
- Dependent variable: heavily imbalanced binary class
- The importance of interaction terms and lags Davis and Karim (2008), Behn et al. (2013), Duca and Peltonen (2013)

Guideline: those can optimise out-of-sample forecasting results

LASSO with cross-validation

Least Absolute Shrinkage and Selection Operator (LASSO) Detail

- Automatically select the most important predictors by shrinking the coefficients of unimportant predictors to zero

$$L_{\lambda}^{lasso}(\beta) = L(\beta) - \lambda \sum |\beta|$$

- Penalty parameter λ
  - Decide more or less variables to be dropped

K-fold cross-validation Detail

- One of the most popular resampling techniques to evaluate model effectiveness
- Select the optimal λ that maximizes the out-of-sample prediction accuracy

Desirable properties:

- Variable selection $\rightarrow$ addressing over-fitting issues and enhancing predictive accuracy
- Inherent simplicity and transparency, controllable
- Statistically closer to the conventional modelling approach

Empirical framework

Multivariate logit model:

$$\ln \frac{Pr(Y_{i,t}=1)}{1-Pr(Y_{i,t}=1)} = domestic_{i,t-h}\beta_1 + global_{i,t-h}\beta_2 + interaction_{i,t-h}\beta_3 + c + \varepsilon_{i,t}$$

Setup of the dependent variable Detail

- $Y_{i,t}=0$ in tranquil periods
- $Y_{i,t}=1$ in the starting period of systemic events
- Drop all observations of "in-crisis" periods for each country
  - Avoid the "post-crisis bias" (Bussiere and Fratzscher, 2006)
  - Focus on the transition from tranquil time to crisis
  - Cons: nothing on crisis duration, losing observations

Country fixed effects or not?

- Fixed effects (Behn et al., 2013)
  - Selection bias: omit countries without crises
- Pooled logit models (Demirgüç-Kunt and Detragiache, 1998; Davis and Karim, 2008; Duca and Peltonen, 2013)
  - Common approach to avoid selection bias
  - Generally better for *out-of-sample* forecasting (Fuertes and Kalotychou, 2006; Dawood et al., 2017)

Model evaluation criteria

The measure of **relative usefulness** U_r, following Sarlin (2013) Detail

Key idea:

- Usefulness: the gain for policymakers to use the model compared to ignoring it
- Relative: ratio of usefulness relative to a perfect performing model

Advantage:

1. Take account of policymakers' preference between missing crises (Type I errors) and issuing false alarms (Type II errors)
2. Adjusted for the imbalanced frequency of tranquil times and crisis events.

Policy preference parameter μ

- Weighting missing crises and issuing false alarms
- Higher $\mu \rightarrow$ stronger preference to avoid missing crises
- $\mu = 0.9 \rightarrow$ the cost of missing a systemic event is far larger than issuing a false alarm

Out-of-sample forecasting exercise: 2007-08 GFC

Sample:

- The base training set: 1970Q1-2004Q4
- The evaluation set: 2005Q1-2009Q4.

Steps:

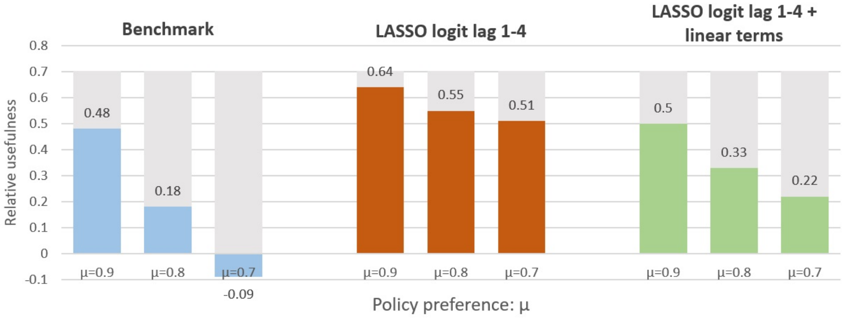
1. With the training set, use LASSO with cross-validation to select a set of predictors that jointly achieve the best out-of-sample forecasting outcomes from the pool of candidate variables.
2. Produce recursive *out-of-sample* forecasts with the selected variables over the evaluation dataset. Detail
   - pseudo real-time manner
3. Compare the forecasting performance of the LASSO logit model with the benchmark model

Selected variables

LASSO with cross-validation selected variables

	Benchmark Behn et al. (2013)	**LASSO logit** **lag 1-4**
	lag 1 credit growth (A) credit gap (B) real GDP growth inflation real house price growth real equity growth global credit growth (C) global credit gap (D) global real GDP growth A*B C*D A*C B*D	lag 1 real house price growth global credit growth $\times$ global credit gap real effective exchange rate $\times$ real house price growth credit gap $\times$ global credit gap lag 3 credit growth $\times$ credit gap credit growth $\times$ current account to GDP lag 4 current account to GDP $\times$ global equity price growth global credit gap $\times$ global inflation
Number of variables	13	8
		$\lambda = 6.021$

Main results: Forecast evaluation



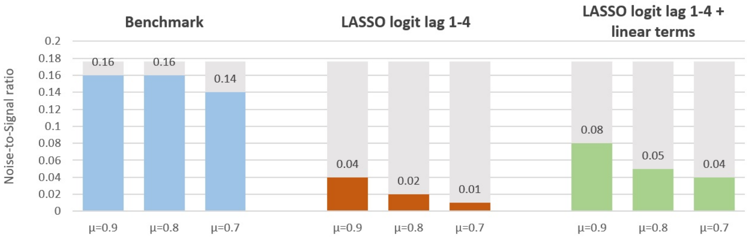
Results:

- LASSO logit models outperform the benchmark model
  → More useful, more stable and more precise predictive results
- Valuable information in the interaction terms and lags
- Policymakers benefit from more warning signals by avoiding the huge cost of crises

Main result: Noise-to-signal ratio

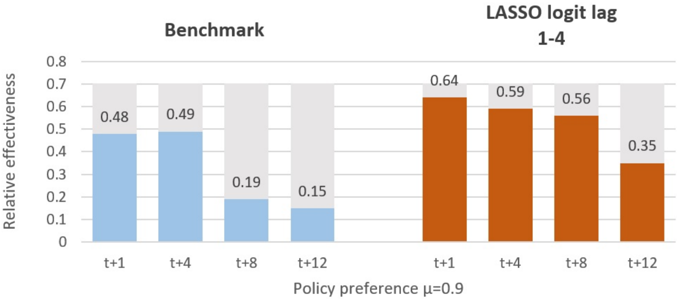
Noise-to-signal ratio:

- False alarms/Hitting crises
- Reflect forecasting accuracy, regardless of policymakers' preference
- Smaller ratio indicates more precise prediction,



Varying forecast horizon

- Policymakers may need a longer forecast horizon to allow time for policy selection and implementation
- Extend the forecast horizon and conduct h-step ahead direct forecasting exercise:



Implication for monitoring financial stability

Five key early warning indicators:

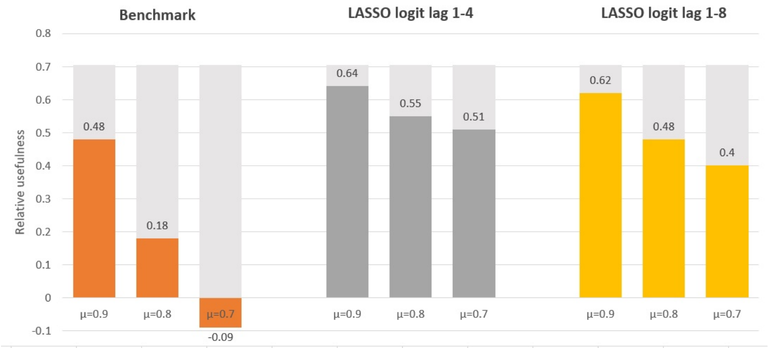
- Credit growth
- Credit gap
- Global credit gap
- Real house price growth
- Real effective exchange rate

Most important contributors of systemic banking crises:

- Frequently selected, regardless of lag orders or in interaction forms
- Reflect the most significant sources of financial vulnerability
- Require close monitoring

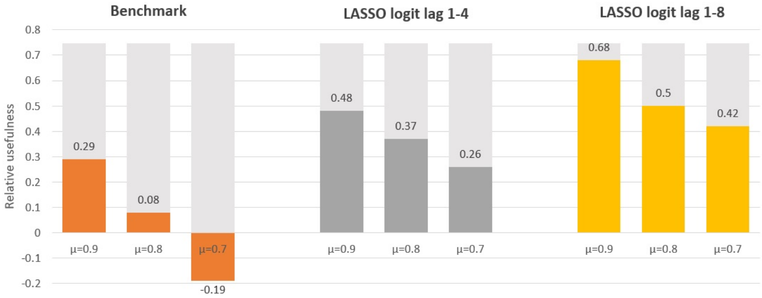
Sensitivity analysis

1. Longer lags of candidate variables
   - Further expanding the lag orders may not be necessary
   - Outperform the benchmark model



Sensitivity analysis

2. Including fixed effects
   - Australia, Canada and New Zealand are omitted from the sample
   - LASSO logit models still strictly outperform the benchmark model



Conclusion

I propose using LASSO with cross-validation to formalize and automate the variable selection process of the conventional multivariate logit EWS.

The proposed method can help build a better early warning system for systemic event forecasting

1. Extract important information that may not be easily identified and accessed by handpicking variables
2. Produce more useful, more stable and more precise forecasting outcomes
3. Keep the EWS transparent and interpretable for policy communication
4. Identify key early warning indicators for close monitoring
    - Credit growth, domestic and global credit gaps, real house price growth and real effective exchange rate.

The End

Appendix

Appendix and Backup Slides

Recursive out-of-sample exercise

The *out-of-sample* exercise is carried out recursively in the following way:

1. Starting with the base training sample 1970Q1-2004Q4, estimate the model on data that contains information only available up to that period. Then predict the probability of a systemic banking crisis that is about to happen in the next period. This is one-step ahead *out-of-sample* forecasting.
2. Compute the *in-sample* relative usefulness U_r given the policy preference μ for all thresholds from 0 to 1. Apply the threshold that yields the highest relative usefulness to the predicted probability of a crisis in the next period and generate a "real-time" warning signal of 0 or 1.
3. Expand the training set by 1 period forward and repeat steps 1-3 recursively.
4. Collect the warning signals produced over the entire evaluation period and calculate the relative usefulness based on the number of false signals and missed crises.

Return

Dependent variable setup: what to predict

EWS literature:

- Set the value of the dependent variable as 1 for a defined pre-crisis horizon and to 0 otherwise
- Use a logit model to predict the defined "pre-crisis" period

No consensus on the forecasting horizon:

- Lo Duca and Peltonen (2013): 6 quarter horizon
- Behn et al. (2013): 5-12 quarter horizon
- Detken et al. (2014): 5-16 quarter horizon

Concern:

1. Implicit assumption: all observations are independent conditional on covariates.
2. Violated due to the construction of the dependent variable: inherent serial correlation
3. Warning signals are not issued independently, but depend on whether there are signals in previous periods.
4. Serial correlation in the residual

To moderate this issue, my approach is to just predict the starting period of the crisis

- Aim to catch the point of transition from tranquil periods to systemic banking crises
- No evidence of serial correlation in errors from the linear probability model

Return

LASSO

In a generic form, the ordinary logit regression with binary response can be written as follows:

$$Pr(y_i = 1) = \pi_i = \frac{e^{y_i \beta}}{1 + e^{y_i \beta}}$$

The generic log-likelihood function can be written as in the following form accordingly:

$$L(\beta) = \sum_{i=1}^{n} \Big[y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i) \Big] = \sum_{i=1}^{n} \Big[y_i x_i \beta - \log(1 + e^{x_i} \beta) \Big]$$

In the standard logit regression, the parameter β are estimated by maximizing the log-likelihood function $L(\beta)$.

In the LASSO regression, the log-likelihood function is penalized with an additional term.

$$L_{\lambda}^{lasso}(\beta) = L(\beta) - \lambda \sum_{j=1}^{p} |\beta|$$

A higher penalty parameter indicates more coefficients of variables would shrink to zero and therefore less variables are selected to be in the model.

Limitation:

- Drop variables that

Return

K-fold cross-validation

K-fold cross-validation

1. Randomly splitting the sample into K-folds of approximately equal size
2. One fold is taken out as the *out-of-sample* **validation fold** while the other K-1 folds are used as **training folds** to estimate the model
3. The estimated parameters from the K-1 folds are used to predict the dependent variable in the one validation fold
4. This procedure is repeated for K times until every fold has served as the evaluation fold

Cross-validation in this study:

- K=5 as suggested by James et al. (2013)
- Stratified → the number of "1"s and "0"s are approximately the same across folds

Concerns of k-fold cross-validation:

- Cross-validation usually requires the data to be independent and identically distributed.
- If the data has time series structure with possible inherent serial correlation, the process of randomly reshuffling the observations will break the feature of time dependence and lead to questionable results.
- Problematic application (Hastie et al., 2009)

Justification:

- A normal K-fold cross-validation procedure is valid if the residuals are uncorrelated (Bergmeir et al., 2018)

Return

Relative usefulness

Following Kaminsky and Reinhart (1999) to construct the contingency matrix

	Crisis event (within 6 quarters)	Tranquil period
Signal	True Positive *Correct signal*	False Positive *False alarm (Type II error)*
No signal	False Negative *Missed crisis (Type I error)*	True Negative *Correct silence*

- Share of Type I error: $T_1 = \frac{FN}{TP+FN}$
  $\rightarrow$ Missed crises relative to the total number of crises
- Share of Type II error: $T_2 = \frac{FP}{FP+TN}$
  $\rightarrow$ False alarms relative to the total tranquil periods

Signals hitting wider targets:

- A signal is taken to be "correct" if it is issued within six quarters of a crisis starting period to accommodate:
  1. Some degree of ambiguity in documentation of crisis starting periods
  2. The ultimate purpose of "early warning"

Relative usefulness

The preference of policymakers takes account for trade-off between issuing false alarm and missing systemic banking crisis

- The optimal threshold of issuing signal is derived according to policymakers' preference μ
  - Set $\mu = 0.9$ to reflect the large costs of missing a crisis (lower threshold)
- Accounting for the unconditional probability of crises and tranquil periods

Following Sarlin (2013) and Alessi and Detken (2011):

1. Define a loss function that depends on the preferences of the policy marker on Type I and Type II errors and adjusts for unconditional probabilities of pre-crisis periods and tranquil periods $(P1, P2)$

$$L(\mu) = \mu T_1 P_1 + (1-\mu) T_2 P_2$$

2. Define the usefulness as the absolute gain for policymakers to use the indicator compared to ignoring it

$$U = min[\mu P_1, (1-\mu) P_2] - L(\mu)$$

3. Define the relative usefulness as the ratio of the usefulness to a perfectly performing model that achieves maximum possible usefulness

$$U_r = \frac{U_a}{\min[\mu P_1, (1-\mu) P_2]}$$

Return

Description of explanatory variables

Description of explanatory variables

Variable	Source	Description
Credit growth	BIS	The year-to-year growth rate of Credit to Private non-financial sector from all sectors at market value - US dollar - Adjusted for breaks
Credit-to-GDP gap	BIS	Calculated by applying the modified BN filter to the credit-to-GDP ratio, which is the Credit to Private non-financial sector from all sectors at market value - Percentage of GDP - Adjusted for breaks
Real GDP growth	OECD	Year-to-year growth
Inflation	OECD	Calculated from OECD CPI index
Short rate	OECD	three month Interbank rate
Long rate	OECD	mostly 10-year government bonds
Term spread	OECD	Calculated as the gap between the long rate and the short rate
Current account in percentage of GDP	OECD	
Real effective exchange rate	OECD	Index 2015=100
Equity price growth	OECD	Year-to-year growth, calculated from OECD share price index, 2010=100
Real house price growth	OECD	Year-to-year growth, calculated from OECD real house price index
House-price-to-income ratio	OECD	
House-price-to-rental ratio	OECD	
House-price-to-income gap	OECD	Calculated by applying the modified BN filter to the house-price-to-income ratio
House-price-to-rental gap	OECD	Calculated by applying the modified BN filter to the house-price-to-rental ratio

Serial correlation test

Cumby-Huizinga test for first order serial correlation

	p-value
Model 1	
Training set (1970Q1-2004Q4)	0.4310
Training set+Evaluation set(1970Q1-2009Q4)	0.0185**
Model 2	
Training set (1970Q1-2004Q4)	0.4290
Training set+Evaluation set(1970Q1-2009Q4)	0.8990

Selected variables

LASSO with cross-validation selected variables

	Benchmark Behn et al. (2013)	**Model 1** **LASSO lag 1**	**Model 2** **LASSO lag 1-4**
	lag 1 credit growth (A) credit gap (B) real GDP growth inflation real house price growth real equity growth global credit growth (C) global credit gap (D) global real GDP growth A*B C*D A*C B*D	lag 1 credit growth $\times$ credit gap real effective exchange rate $\times$ real house price growth credit growth $\times$ global credit gap credit gap $\times$ global credit gap	lag 1 real house price growth global credit growth $\times$ global credit gap real effective exchange rate $\times$ real house price growth credit gap $\times$ global credit gap lag 3 credit growth $\times$ credit gap credit growth $\times$ current account to GDP lag 4 current account to GDP $\times$ global equity price growth global credit gap $\times$ global inflation
Number of variables	13	4	8
		$\lambda = 6.867$	$\lambda = 6.021$

Main results: Forecast evaluation

	Benchmark			**Model 1** LASSO lag 1			**Model 2** LASSO lag 1-4			**Model 3** LASSO lag 1-4 + the linear terms		
Policy preference (μ)	0.9	0.8	0.7	0.9	0.8	0.7	0.9	0.8	0.7	0.9	0.8	0.7
Relative usefulness	0.48	0.18	-0.09	0.67	0.47	0.27	0.64	0.55	0.51	0.50	0.33	0.22
Noise to signal ratio	0.16	0.16	0.14	0.04	0.03	0.02	0.04	0.02	0.01	0.08	0.05	0.04

Results:

- LASSO logit models outperform the benchmark model
- Longer lags of the covariates contain valuable information
- Policymakers benefit from more warning signals by avoiding the huge cost of crises

LASSO with cross-validation is useful to extract information that may not be accessible by the standard multivariate logit models.

- More useful, more stable and more precise predictive results

Biblography

Alessi, L. and Detken, C. (2018). Identifying excessive credit growth and leverage. *Journal of Financial Stability*, 35:215–225.

Behn, M., Detken, C., Peltonen, T. A., and Schudel, W. (2013). Setting countercyclical capital buffers based on early warning models: would it work? *ECB Working Paper*, (No. 1604).

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis*, 120:70–83.

Bussiere, M. and Fratzscher, M. (2006). Towards a new early warning system of financial crises. *Journal of International Money and Finance*, 25(6):953–973.

Davis, E. P. and Karim, D. (2008). Comparing early warning systems for banking crises. *Journal of Financial Stability*, 4(2):89–120.

Dawood, M., Horsewood, N., and Strobel, F. (2017). Predicting sovereign debt crises: an early warning system approach. *Journal of Financial Stability*, 28:16–28.

Demirgüç-Kunt, A. and Detragiache, E. (1998). The determinants of banking crises in developing and developed countries. *IMF Staff Papers*, 45(1):81–109.

Detken, C., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M. M., Castro, C., Frontczak, S., Giordana, G., Giese, J., Jahn, N., et al. (2014). Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options. *European Systemic Risk Board (ESRB) Occasional Paper Series*, (No.5).

Duca, M. L. and Peltonen, T. A. (2013). Assessing systemic risks and predicting systemic events. *Journal of Banking and Finance*, 37(7):2183–2195.

Fuertes, A.-M. and Kalotychou, E. (2006). Early warning systems for sovereign debt crises: The role of heterogeneity. *Computational Statistics and Data Analysis*, 51(2):1420–1441.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data*