

Seminar
Data Science Methoden für Finanzmärkte mit R
Wintersemester 2021/22

Als Basisliteratur verwenden wir in diesem Seminar das folgende Lehrbuch:

James, G., Witten, D., Hastie, T., und Tibshirani, R. (2013/2021). An Introduction to Statistical Learning with Applications in R. Springer.

<https://link.springer.com/book/10.1007/978-1-4614-7138-7>

Zusätzlich finden Sie bei jedem Thema weitere Forschungsartikel die als Startpunkt in die wissenschaftliche Literatur dienen. Von dort aus sollen die Studierenden eigenständig die relevante Literatur weiter erkunden und erarbeiten.

Die Studierenden werden in einem Tutorial der Universitätsbibliothek in der Literaturrecherche geschult.

In allen folgenden zehn Themen sollen die Studierenden eigenständige empirische Analysen durchführen. Dazu wird die open-source Software R verwendet. Weitere Informationen finden Sie unter:

<https://www.fernuni-hagen.de/angewandte-statistik/lehre/software.shtml>

Die Studierenden werden in Tutorials mit der Software R vertraut gemacht und zur eigenen Datenrecherche und Analyse befähigt.

Themenübersicht:

1 Dimensionsreduktion der Varianz-Kovarianz-Matrix in großen Portfolios

Institutionelle Investoren sind mit der Herausforderung konfrontiert, dass sie für ihre i.d.R. mehrere hundert Assets enthaltenen Portfolios die Varianz-Kovarianz-Matrix berechnen müssen, um adäquate Risikomanagement und Asset-Allokationsentscheidungen treffen zu können. Da die Berechnung und Anpassung für solche großen Matrizen mit erheblichem Aufwand verbunden sind, werden Verfahren benötigt, um die Dimensionen dieser Matrix zu reduzieren.

Ziel dieser Seminararbeit ist die Anwendung von Dimensionsreduktionsverfahren auf Varianz-Kovarianz-Matrizen und die Überprüfung von deren Güte.

James et al. (2013), Fabozzi et al. (2007), Ledoit und Wolf (2003), Bai und Shi (2011), Fan et al. (2008)

2 Hauptkomponentenanalyse im Faktor-Asset-Pricing

In der Asset-Pricing Literatur wurde eine Vielzahl von verschiedenen Faktoren identifiziert (z.B. Size, Value, Liquidity und Momentum), die einen potentiellen Erklärungsgehalt für Aktienrenditen liefern. Um eine hohe Anzahl von Faktoren auf weniger erklärende Faktoren zu reduzieren, kann die Hauptkomponentenanalyse verwendet werden.

Ziel dieser Seminararbeit ist die Anwendung der Hauptkomponentenanalyse auf Faktormodelle im Asset-Pricing. Aufgezeigt werden soll, wie sich die Hauptkomponentenanalyse in das Faktor-Asset-Pricing integrieren lässt und wie diese im Vergleich zu „herkömmlichen“ Modellen abschneidet.

James et al. (2013), Lettau und Pelger (2020), Feng et al. (2020), Kozak et al. (2020)

3 LASSO-Regression im Faktor-Asset-Pricing

Um die Anzahl der Asset-Pricing Faktoren auf diejenigen zu beschränken, die einen hinreichenden Erklärungsgehalt aufweisen, können LASSO-Regressionen verwendet werden. Diese „bestrafen“ das Hinzufügen neuer Faktoren und somit werden nur Faktoren in der Regression berücksichtigt, bei denen der Bestrafungsterm kleiner als der zusätzliche Nutzensgewinn ist.

Ziel dieser Seminararbeit ist die Anwendung und der Vergleich von LASSO-Regressionen mit herkömmlichen Regressionsmodellen im Bereich Asset Pricing.

James et al. (2013), Feng et al. (2020), Gu et al. (2020), Zhang et al. (2019)

4 Prognose der Richtung von Aktienmarktrenditen mit Klassifikationsmethoden

Es ist bekannt, dass die Rendite des Aktienmarkts schwierig vorherzusagen ist. Das Problem wird vereinfacht, wenn man nur die Richtung (Vorzeichen) der Rendite betrachtet. Mit der Hilfe der Klassifikationsmethoden hat man die Möglichkeit die Entwicklung des Aktienkurses der näheren Zukunft zu prognostizieren.

In dieser Arbeit wird zuerst grundsätzlich die verwendende Klassifikationsmethode erklärt, dann folgt eine empirische Studie mit den Daten des Aktienmarkts.

James et al. (2013), Nyberg (2011), Huang et al. (2005)

5 Lineare und quadratische Klassifikationsverfahren in Energiemärkten

Multivariate Klassifikationsverfahren finden breite Anwendungsmöglichkeiten im Energie- und Finanzbereich, wenn es darum geht, qualitative Antwortparameter zu prognostizieren. Schwerpunkt dieser Arbeit liegt in diesem Zusammenhang insbesondere auf der Anwendung der linearen sowie der quadratischen Diskriminanzanalyse.

Ziel dieser Seminararbeit soll es zum einen sein, lineare und quadratische Klassifikationsverfahren theoretisch fundiert zu erklären; des Weiteren sollen beide Verfahren vergleichend auf aktuelle Energiemarktdaten angewendet werden.

James et al. (2013), Rencher und Christensen (2012), Cooke und Peake (2002), Leung et al. (2000)

6 Resampling-Methoden und Bootstrapp

Computerintensive Verfahren sind in den letzten Dekaden immer wichtiger geworden. Gerade in Situationen, in denen asymptotische Ergebnisse (basierend auf unendlich großen Stichproben) wenig hilfreich sind, ist insbesondere der Bootstrap ein sehr prominentes Resampling-Verfahren.

Bei dieser Arbeit soll es darum gehen, den Bootstrap und die wichtige Variante „wild bootstrap“ (bei Heteroskedastie und zeitlich-variierenden Varianzen) darzustellen und in einer empirischen Anwendung für Aktienkursrenditen umzusetzen.

James et al. (2013), Davidson und Flachaire (2008), Godfrey (2009), Kim und Shamsuddin (2020)

7 Splines

Die Spline-Familie bietet vielfältigen Anwendungsmöglichkeiten mit nicht linearer Modellierung im Finanzbereich. Ein Spline kann der einzige Bestandteil der Modellierung sein, oder kann auch mit diversen bekannten Modellen kombiniert werden, z.B. Spline-GARCH.

Bei dieser Arbeit wird die Theorie von Splines diskutiert sowie die Anwendungsmöglichkeiten. Bei dem Anwendungsteil wird eine empirische Analyse durchgeführt mit den aktuellen Daten des Finanzmarkts.

James et al. (2013), Waggoner (1997), Engle und Rangel (2008)

8 Nichtlineares Modellieren von Rohstoffpreisen: GAMs

Die Modellierung von Rohstoffpreisen ist häufig nichtlinear, was simple Linearregressionen infolgedessen in ihrer Aussagekraft beschränkt. Generalized Additive Models (GAMs) stellen in dem Zusammenhang eine Erweiterung der Linearregression dar und ermöglichen somit die Schätzung nichtlinearer Modelle.

Ziel dieser Arbeit ist es, das Konzept der Generalized Additive Models zu erklären und am Beispiel von Rohstoffpreisen empirisch zu modellieren.

James et al. (2013), Hastie und Tibshirani (1986), Chen und Samworth (2016), Wood (2020)

9 Integration von Nichtlinearitäten in das CAPM mittels GAM

Mit Generalized Additive Models (GAM) kann man lineare Modelle einfach zu nicht linearen Modellen erweitern. Mit dieser Eigenschaft wird das Modell im Finanzbereich zusammen mit dem Capital Asset Pricing Model (CAPM) verbunden, da manche Faktoren kein lineares Verhältnis zu den Renditen aufweisen.

In dieser Arbeit wird GAM und CAPM erläutert, sowie den ausgewählten Faktoren für die empirische Anwendung. Das Ergebnis aus geschätzten nicht lineare CAPM wird auch vergleicht mit klassischem CAPM.

James et al. (2013), Neslihanoglu et al. (2017), Fama und French (2004)

10 Random Forests

Entscheidungsbäume sind nicht-parametrische Verfahren, die sich für Klassifikationsverfahren sowie für Regressionen eignen. Allerdings wird die Prognose-Performance von Entscheidungsbäumen häufig kritisiert. Random Forests sollen in dem Zusammenhang die Vorhersagegenauigkeit verbessern, indem sie verschiedene Entscheidungsbäume miteinander kombinieren.

Ziel dieser Seminararbeit soll es daher zum einen sein, die Methode der Random Forests zu erklären und in den Kontext von Entscheidungsbäumen zu setzen; zum anderen soll die Methode zur Regressionsanalyse auf aktuelle Daten des Finanzmarktes empirisch angewendet werden.

Breiman (2001), James et al. (2013), Probst und Boulesteix (2018).

Literatur

Bai, J. und Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. Columbia University Department of Economics Discussion Paper Series, No. 1112-03.

Breiman, L. (2001). Random forests. Machine Learning 45, 5-32.

- Chen, Y. und Samworth, R.J. (2016). Generalized additive and index models with shape constraints. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 78, No. 4, 729-754.
- Cooke, T., und Peake, M. (2002): The optimal classification using a linear discriminant for two point classes having known mean and covariance. *Journal of Multivariate Analysis* 82, 379-294.
- Davidson, R., und Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146, No. 1, 162-169.
- Engle, R. F., und Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies* 21, No. 3, 1187-1222.
- Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A. und Focardi, S. M. (2007). *Robust Portfolio Optimization and Management. The Frank J. Fabozzi Series, Wiley Finance, Wiley.*
- Fama, E. F., und French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives* 18, Vol. 3, 25-46.
- Fan, J., Fan, Y. und Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147, No. 1, 186-197.
- Feng, G., Giglio, S. und Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance* 75, No. 3, 1327-1370.
- Godfrey, L. (2009). *Bootstrap tests for regression models.* Springer.
- Gu, S., Kelly, B. und Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies* 33, No. 5, 2223-2273.
- Hastie, T.J. und Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science* 1, No. 3, 297-318.
- Huang, W., Nakamori, Y., und Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* 32, No. 10, 2513-2522.
- James, G., Witten, D., Hastie, T. und Tibshirani, R. (2013). *An introduction to statistical learning with applications in R.* Springer.
- Kim, J. H., und Shamsuddin, A. (2020). A bootstrap test for predictability of asset returns. *Finance Research Letters*, 35, Article 101289.
- Kozak, S., Nagel, S. und Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics* 135, No. 2, 271-292.
- Ledoit, O. und Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10, No. 5, 603-621.
- Lettau, M. und Pelger, M. (2020). Estimating latent asset-pricing factors. *Journal of Econometrics* 2018, No. 1, 1-31.

- Leung, M.T., Daouk, H., und Chen, A. (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting* 16, No. 2, 173-190.
- Neslihanoglu, S., Sogiakas, V., McColl, J. H., und Lee, D. (2017). Nonlinearities in the CAPM: Evidence from developed and emerging markets. *Journal of Forecasting* 36, No. 8, 867-897.
- Nyberg, H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting* 27, No. 2, 561-578.
- Probst, P., und Boulesteix, A. (2018). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research* 18, 1-18.
- Rencher, A.C., und Christensen, W.F. (2012). *Methods of multivariate analysis*. Wiley Series in Probability and Statistics.
- Waggoner, D. F. (1997). Spline methods for extracting interest rate curves from coupon bond prices. *Federal Reserve Bank of Atlanta Working Paper*, 97-10.
- Wood, S. (2020). Inference and computation with generalized additive models and their extensions. *TEST* 29, 307-339.
- Zhang, Y., Ma, F. und Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?. *Journal of Empirical Finance* 54, 97-117.