

Seminar im Wintersemester 2022

Big Data Analysen in der Makroökonomie mit R

Der Lehrstuhl für Angewandte Statistik betreut im Wintersemester 2022 Seminararbeiten an der Schnittstelle zwischen Statistik und Makroökonomie aus dem Bereich der Big Data Analysen in der Makroökonomie mit der Open-Source-Programmiersprache R. Als leicht zugängliche Big Data Quellen haben sich auf gesamtwirtschaftlicher Ebene vor allem die monatliche FRED-MD und vierteljährliche FRED-QD Datenbanken McCracken und Ng (2016, 2021) etabliert. Weitere Quellen großer Daten sind Textdaten sowie Paneldaten im Rahmen der Netzwerke bzw. Spilloveranalyse. In dem Seminar werden sechs Oberthemen jeweils doppelt vergeben (mit angepasster Forschungsfrage).

Themenliste

Big Data Typen im Bereich Makroökonomie

Wirtschaftswissenschaftlicher wissen, wie wichtig Daten sind, insbesondere wenn es um die Überwachung gesamtwirtschaftlicher Bedingungen geht - die Grundlage für fundierte wirtschaftliche und politische Entscheidungen. In der Seminararbeit soll ein Überblick über die unterschiedlichen Arten von Big Data gegeben und erläutert werden, wie diese in der makroökonomischen Analyse verwendet werden können. Gehen Sie dabei auch auf mögliche Fallstricke ein. Im Rahmen einer deskriptiven statistischen Analyse wird zudem erwartet, dass Sie einen Big Data Datensatz aufbereiten und die Eigenschaften mit R beschreiben.

Ademmer et al. (2021), Bokun et al. (2022), Buono et al. (2017), Ellingsen et al. (2022) und McCracken und Ng (2016, 2021)

Hauptkomponentenanalyse

Um mit großen Datensätzen, insbesondere mit einer großen Anzahl potentieller Erklärungsvariablen bei begrenzter Anzahl an Beobachtungen, umgehen und darauf aufbauend Analysen durchführen zu können, kann die Hauptkomponentenanalyse verwendet werden. Diese reduziert die Anzahl der erklärenden Faktoren, indem Kombinationen aus diesen gebildet werden, sog. Eigenfaktoren. Zunächst soll das Konzept der Hauptkomponentenanalyse erläutert und anschließend zum Beispiel ein monatlicher Konjunkturindikator konstruiert werden. Als Beispiel dient der Chicago Fed National Activity Index (CFNAI). Zwei weitere Anwendungen wären die Verwendung der Hauptkomponente für die Prognose sowie die Kombination mit einem VAR zu einem Factor-Augmented VAR (FAVAR).

Bai und Ng (2008), Bernanke et al. (2005), Evans et al. (2002) und James et al. (2021)

Shrinkage Methoden

Um eine zu hohe Anzahl an Parametern in Regressionsmodellen zu vermeiden, können Bestrafungsterme eingeführt werden. Solche Regressionsmodelle werden als Elastisches Netz bezeichnet, da diese mit Ausnahme der Kleinst-Quadrat-Schätzung einen Straftermin für Regressoren enthalten. Die beiden anderen Spezialfälle verwenden dabei die L2 Norm (Ridge Regression) bzw. die L1 Norm (Lasso Regression). Da potentiell eine Vielzahl an Variablen einen Beitrag zur Prognose makroökonomischer Variablen wie dem BIP oder der Inflation liefern können, bieten Shrinkage Methoden als Alternative zur Dimensionsreduktion durch die Hauptkomponentenanalyse eine Methode, die Modellgüte zu verbessern und die Prognosefehler zu reduzieren und wichtige Variablen zu selektieren. Ziel der Arbeit ist es die verschiedenen Shrinkage Methoden darzustellen und anschließend anzuwenden. Anwendungsbeispiele sind die Analyse der Determinanten des langfristigen Wirtschaftswachstums sowie die Prognose makroökonomischer Variablen.

James et al. (2021), Medeiros et al. (2021) und Sala-I-Martin (1997)

Klassifikation

Das lineare Regressionsmodell nimmt an, dass die abhängige Variable quantitativ ist. Es gibt jedoch auch viele interessante Fragestellungen, in denen die abhängige Variable lediglich qualitativ/diskret ist. Die einfachste Analyse bezieht sich auf den Fall, in welchem ein Ereignis eintritt (die Wirtschaft befindet sich in einer Rezession) oder nicht eintritt (die Wirtschaft befindet sich nicht in einer Rezession). Logit/Probit Regressionen werden verwendet, um auf Basis der Regressoren die Wahrscheinlichkeit zu bestimmen, mit der eine Beobachtung zu einer bestimmten Klasse gehört. Ziel dieser Arbeit ist es bspw. mit Hilfe von Logit/Probit Regressionen und der FRED-MD/QD Datenbank Faktoren zu ermitteln, welche ex-post/in-sample die Wahrscheinlichkeit für eine Rezession erklären können. Andere mögliche Anwendungsfälle mit diskreten Ausprägungen als abhängige Variable sind Banken-, Staatsschulden- und Währungskrisen.

James et al. (2021), Laeven und Valencia (2020), Layton und Katsuura (2001), Liu und Moench (2014) und Stock und Watson (1989)

Text als Daten

Ein immer größerer Teil der menschlichen Kommunikation und Kultur wird als digitaler Text aufgezeichnet. Dementsprechend stellen bspw. digitale Zeitschriftenarchive oder Twitter eine neue Big Data Quelle für die Wirtschafts-

forschung dar. Disziplinen wie Geschichts- und Literaturwissenschaften stützen sich bereits seit langem stark auf Textdaten, analysieren diese jedoch häufig auf qualitative, oft hermeneutische Weise. Makroökonomien bevorzugen jedoch eine quantitative Analyse um Cluster oder typische Muster aufzuspüren und zu interpretieren. Bei der Analyse muss jedoch berücksichtigt werden, dass sich Texte von anderen Datenformen unterscheiden und spezielle Methoden wie das "Natural Language Processing" angewendet werden müssen und auf Encoding-Probleme geachtet werden muss. Wörterbücher (Wortlisten) können erstellt werden, um begriffliche Ereignisse zu zählen, siehe z.B. Baker et al. (2016) zur Messung der Unsicherheit über die Wirtschaftspolitik. Ziel der Arbeit ist es die Methoden zur Textanalyse darzustellen und zur Konstruktion eines Sentimentindikators oder Unsicherheitsmaßes anzuwenden.

Baker et al. (2016), Bontempi et al. (2019), Gentzkow et al. (2019) und Silge und Robinson (2017)

Netzwerk/Spilloveranalyse

Die Asienkrise Ende der 1990er Jahre sowie die globale Finanzkrise in 2008 haben die Verletztheit der internationalen Finanzmärkte gezeigt. Die Eurokrise hebt zudem die Integration der europäischen Volkswirtschaften hervor, so dass die Konjunktur in einem Land sich auf die europäischen Partner auswirkt. Die Netzwerk- bzw. Spilloveranalyse bietet einen Ansatz für die Untersuchung finanzieller und makroökonomischer Verflechtungen. Die Spillover-Indizes von Diebold und Yilmaz (2009) basieren einen solchen Index auf der Varianzzerlegung in Zusammenhang mit einem n-variablen VAR. Ziel der Arbeit ist die Darstellung der Methode zur Netzwerk/Spilloveranalyse und die empirische Anwendung auf Renditen oder Volatilität der globalen Finanzmärkte oder länder- bzw. regionenübergreifende Konjunkturübertragungen.

Bokun et al. (2022), Diebold und Yilmaz (2009, 2015) und Pfaff (2008)

Literatur

- Ademmer, M., Beckmann, J., Bode, E., Boysen-Hogrefe, J., Funke, M., Hauber, P., Heidland, T., Hinz, J., Jannsen, N., Kooths, S., Söder, M., Stamer, V. und Stolzenburg, U. (2021). *Big Data in der makroökonomischen Analyse*. Kieler Beiträge zur Wirtschaftspolitik 32. Kiel Institut für Weltwirtschaft.
- Bai, J. und Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends® in Econometrics* 3 (2), 89–163.
- Baker, S. R., Bloom, N. und Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics* 131 (4), 1593–1636.
- Bernanke, B. S., Boivin, J. und Elias, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics* 120 (1), 387–422.
- Bokun, K. O., Jackson, L. E., Kliesen, K. L. und Owyang, M. T. (2022). FRED-SD: A real-time database for state-level data with forecasting applications. *International Journal of Forecasting* (ahead of print).
- Bontempi, M. E., Frigeri, M., Golinelli, R. und Squadrani, M. (2019). *Uncertainty, perception and the internet*. Working Paper 1134. Dipartimento Scienze Economiche, Università di Bologna.
- Buono, D., Mazzi, G. L., Kapetanios, G., Marcellino, M. und Papailias, F. (2017). Big data types for macroeconomic nowcasting. *Eurostat Review on National Accounts and Macroeconomic Indicators* 7 (1), 93–145.
- Diebold, F. X. und Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal* 119 (534), 158–171.
- Diebold, F. X. und Yilmaz, K. (2015). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press.
- Ellingsen, J., Larsen, V. H. und Thorsrud, L. A. (2022). News media versus FRED-MD for macroeconomic forecasting. *Journal of Applied Econometrics* 37 (1), 63–81.
- Evans, C. L., Liu, C. T. und Pham-Kanter, G. (2002). The 2001 recession and the Chicago Fed National Index: Identifying business cycle turning points. *Economic Perspectives* 26 (3), 26–43.
- Gentzkow, M., Kelly, B. und Taddy, M. (2019). Text as data. *Journal of Economic Literature* 57 (3), 535–574.
- James, G., Witten, D., Hastie, T. und Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R*. 2. Auflage. Springer Texts in Statistics. Springer.
- Laeven, L. und Valencia, F. (2020). Systemic banking crises database II. *IMF Economic Review* 68 (2), 307–361.

- Layton, A. P. und Katsuura, M. (2001). Comparison of regime switching, probit and logit models in dating and forecasting US business cycles. *International Journal of Forecasting* 17 (3), 403–417.
- Liu, W. und Moench, E. (2014). *What predicts U.S. recessions?* Staff Report 691. Federal Reserve Bank of New York.
- McCracken, M. W. und Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34 (4), 574–589.
- McCracken, M. W. und Ng, S. (2021). FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis Review* 103 (1), 1–44.
- Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á. und Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics* 39 (1), 98–119.
- Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*. 2. Auflage. Use R! Springer.
- Sala-I-Martin, X. X. (1997). I just ran two million regressions. *American Economic Review* 87 (2), 178–183.
- Silge, J. und Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. 1. Auflage. O’Reilly.
- Stock, J. H. und Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual* 4, 351–409.