WESTERN SYDNEY UNIVERSITY

illion
Formerly Dun & Bradstreet

ROZETTA
formerly CMCRC
ASSURING MARKET QUALITY

A Machine Learning Approach to Validating the use of Australian Credit Reporting Data as a Leading Indicator of Consumer Spending

Geoff Chang, Dr Oliver Obst, Dr Kathy Tannous

Motivation

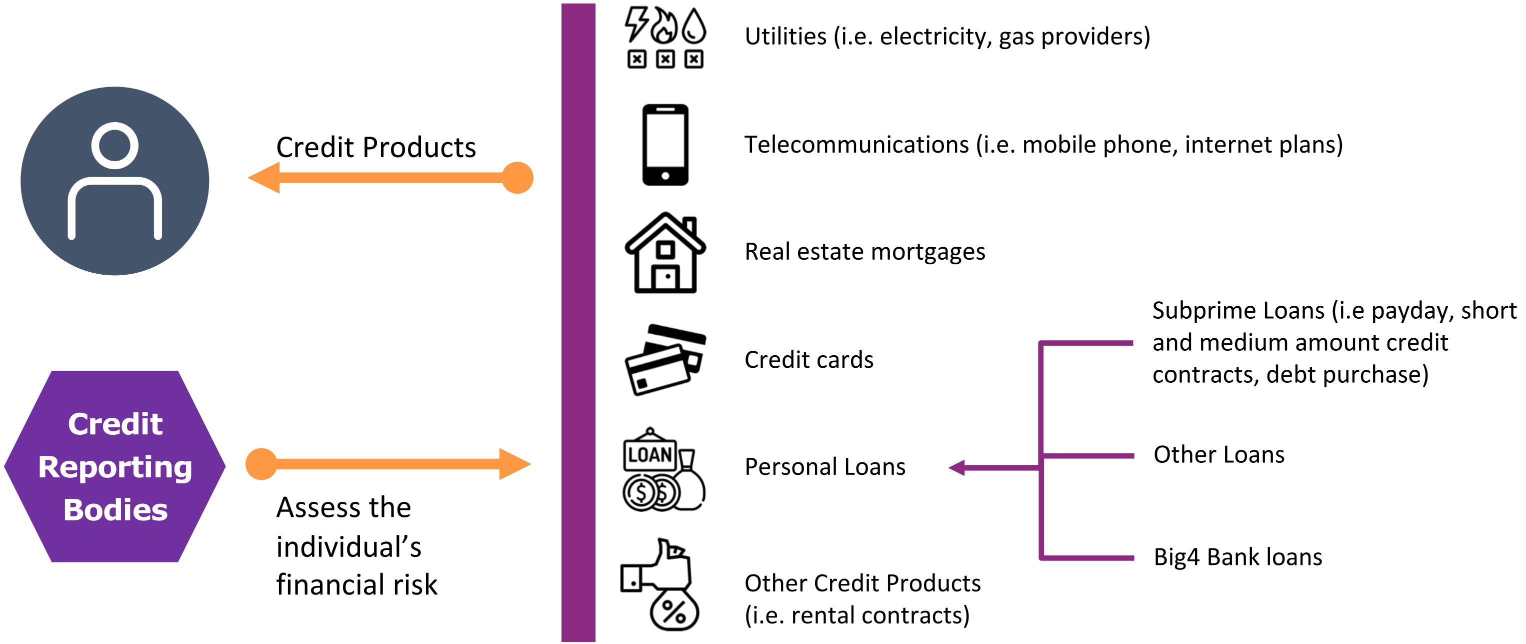
Existing economic indicators are only weakly predictive and explanative of consumer spending because of their limited sample size and surveying methodology. As our credit ecosystem is increasingly digitised, data is collected on more individuals with greater granularity. This research has the following aims:

To compare the ability of credit reporting data to explain and predict consumption with other indicators – in particular, consumer sentiment

The use of machine learning algorithms to automate the search for combinations of variables that improve the explanative and predictive power of simple models

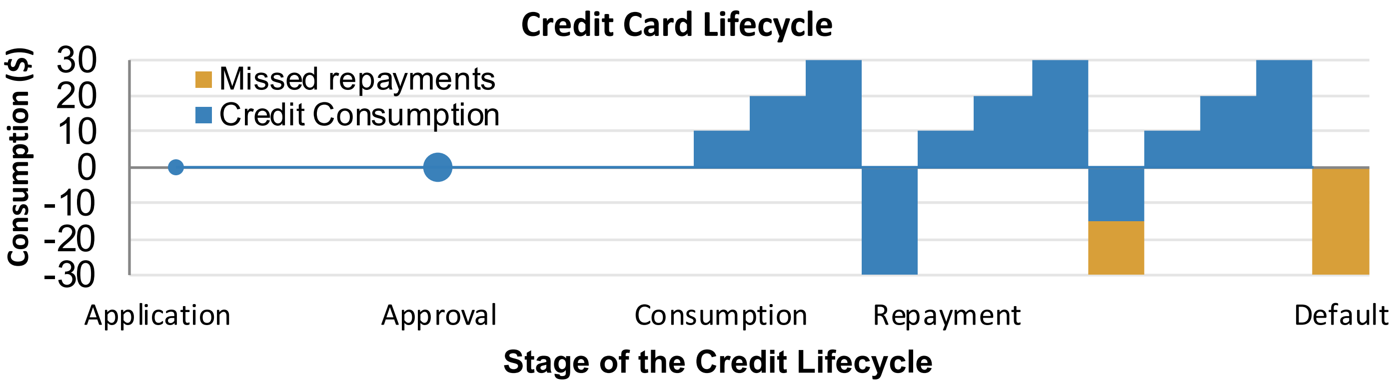
What is Credit Reporting Data?

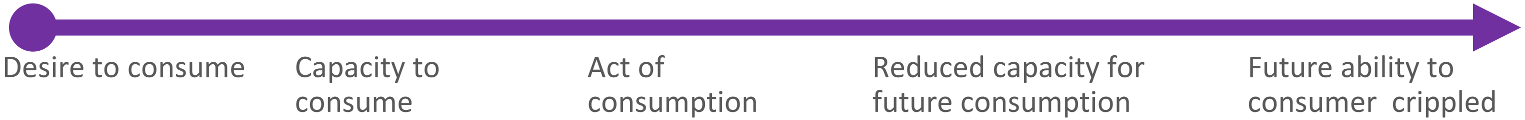
Credit reporting captures the interactions between consumers and credit providers. This study uses the monthly aggregates of 15 million Australians.



Credit reporting is analogous to sentiment

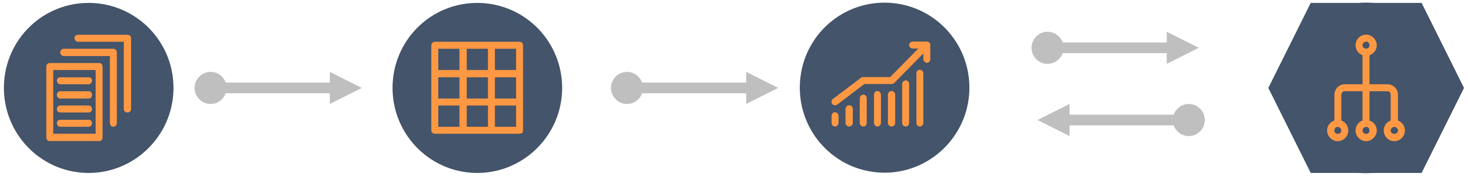
Sentiment measures the individuals capacity and willingness to spend. This study measures these same metrics at different stages of the credit lifecycle using the monthly aggregations of credit applications, account openings, repayment history, defaults and other negative events.





Methodology

Credit reporting data's ability to explain and forecast household consumption was compared using linear and non-linear models. Additionally, Tree Parzen Estimators was used to automate the search for combinations of input features that improved mode performance



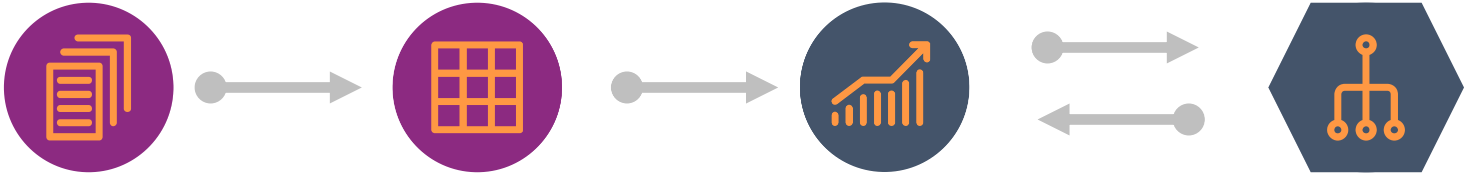
Extract data from ABS, RBA and illion

Preprocess data by using X-13-ARIMA-SEATS or first order differencing

Compare predictive and explanatory power of models

Automated search for combinations of independent variables using Tree Parzen Estimators

Data Extraction and Preprocessing



The following data is extracted between

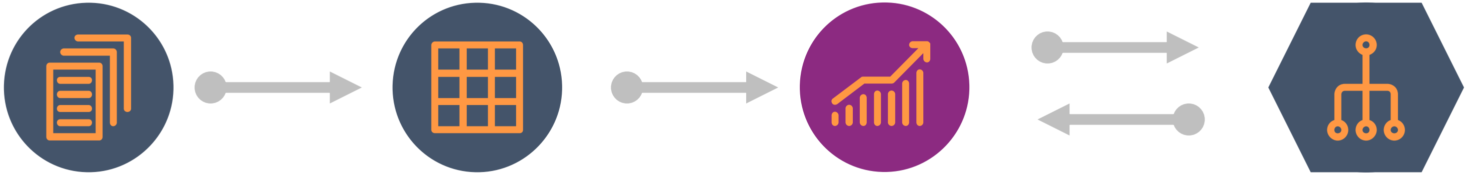
July 2014 and May 2019:

- Comprehensive credit reporting data
- All monthly and quarterly indicators from ABS and RBA
- Consumer Sentiment Surveys (WMI and ANZ-RM)
- ASX Stock aggregates

Each variable was preprocessed by:

- Decomposing seasonal and trend components using X13-ARIMA SEATS
- Where this decomposition fails due to the inability to separate seasonal and trend components, the first difference is taken
- Adjust the lags to reflect the information that would be available at each time period

Reduced form analysis to measure explanative and predictive power



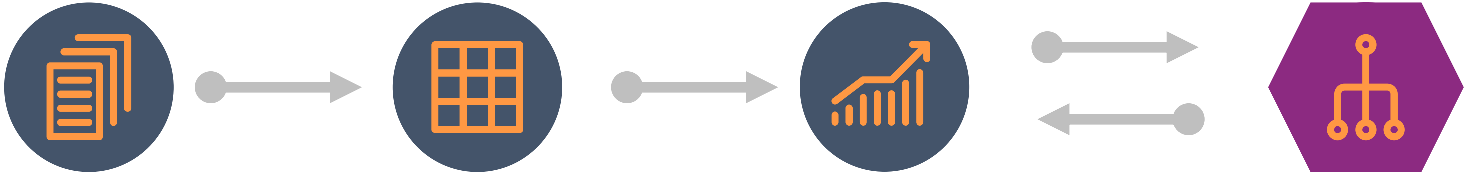
OLS method is applied to reduced form linear regression models with different combinations of independent variables

Dependent variable is the Total Retail Turnover (as a proxy for consumption).

The following independent are tested:

- Retail Turnover (autoregressive model)
- Consumer sentiment indices (ANZ-RM and W-MI)
- Credit reporting data
- Stock market wealth, housing wealth and income (CHECK!!!)

Automated Variable Search



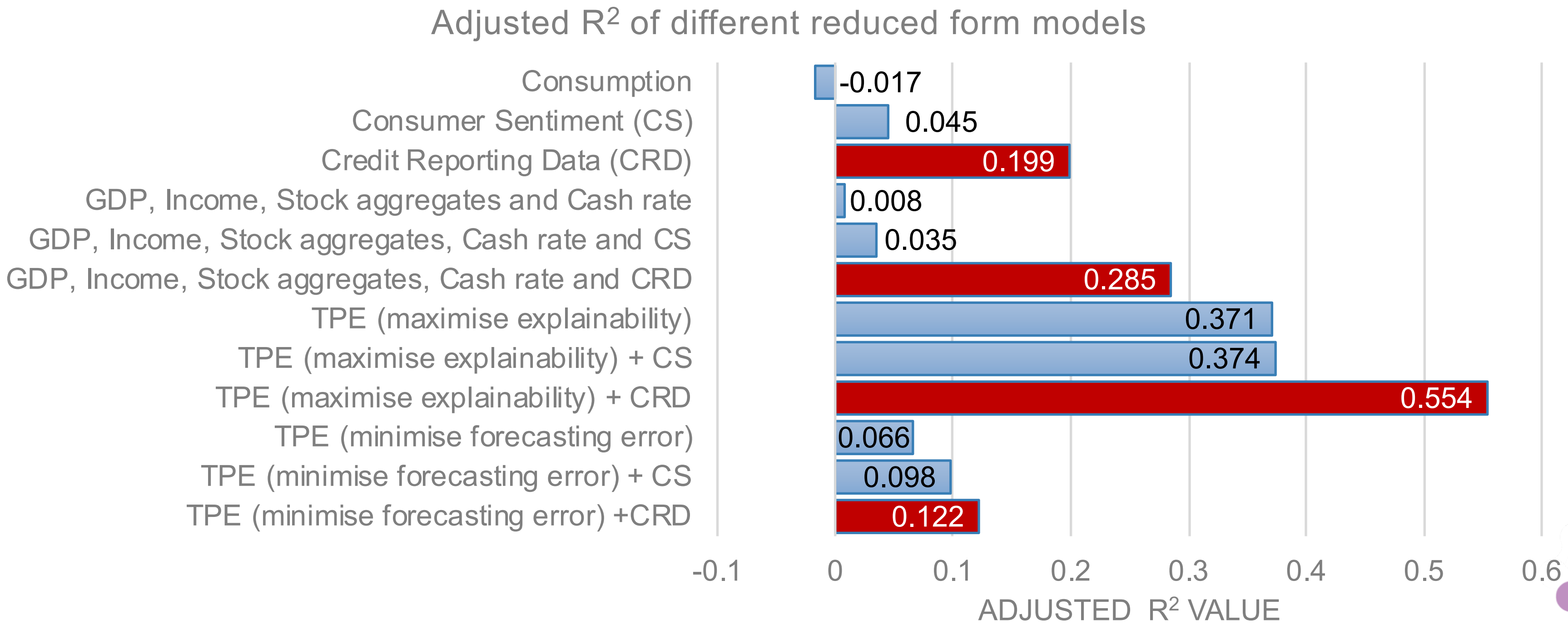
Tree Parzen Estimators is a Bayesian optimisation method that constructs tree-structured spaces and sequentially searches for combinations of variables that optimise an objective.

For explanative power, allow the algorithm to select 6 variables from the available datasets and maximise the **adjusted R^2**

For predictive power, allow the algorithm to select 6 variables and minimise the **mean absolute error.**

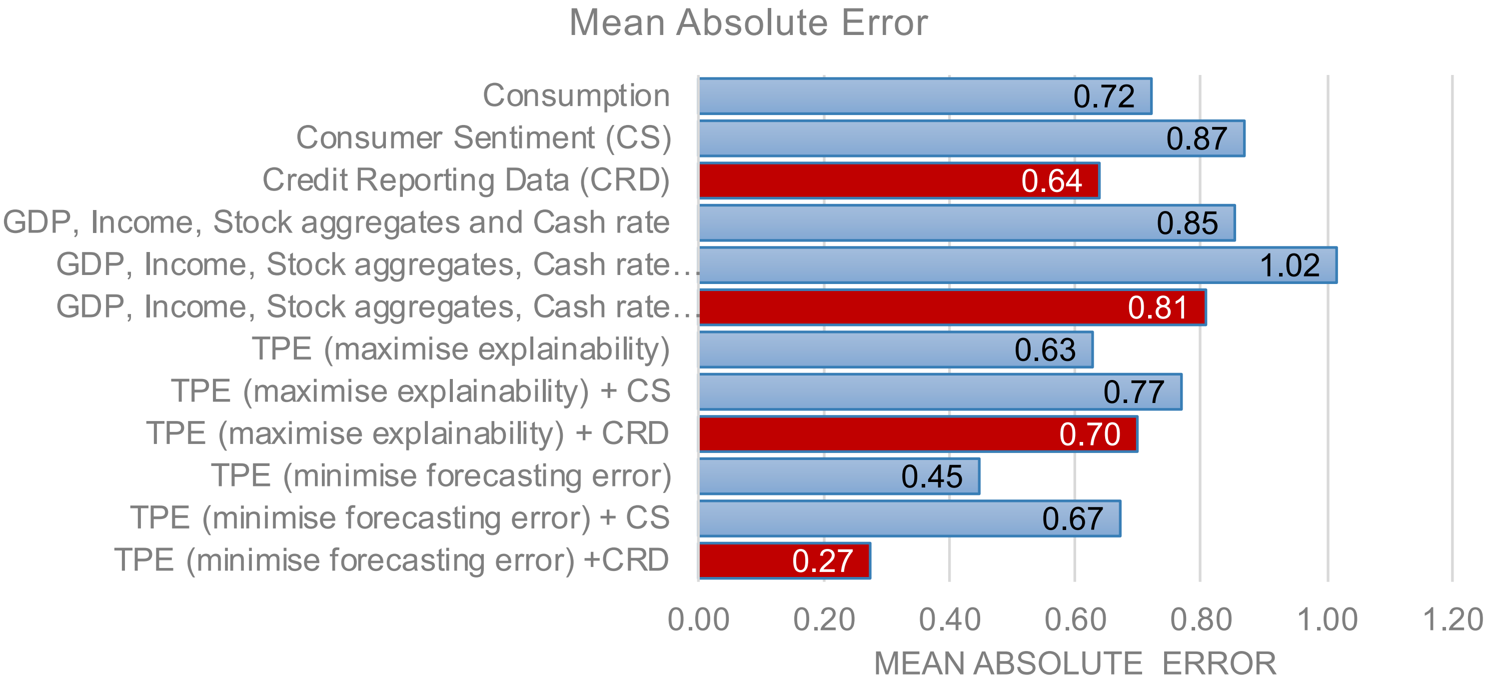
Explanative Power

The inclusion of credit reporting data in reduced form models increases the explanative power of the model, as shown by the adjusted R^2 of the OLS method.



Predictive Power

Similarly, the inclusion credit reporting data in the reduced form equations reduce the forecasting error for the out of sample datasets



Indicators Selected by TPE

Explanative Model:

- Number of commercial utility enquiries
- Number of consumer enquiries for all products
- Index numbers for the maintenance and repair of motor vehicles
- Civilian population aged 15 years and over
- Percentage change from the previous period of non-durable household products
- Number of building job values at $20 to $50 million in WA

Predictive Model

- Telco Enquiries
- Number of newly discharged bankruptcies
- Housing index percentage change from previous period
- CPIs of Adelaide
- Total value of commercial building jobs in Tasmania
- Percentage change in the price of footwear from Corresponding Quarter of Previous Year

Impact

The data and methods used in this paper strongly supplement, not replace, existing measures of consumer sentiment and the independent variable selection

- Automating searches and statistical tests is a useful tool for identifying variables for consideration in structural model
- Credit reporting data is more predictive and explanative of consumption than existing measures of consumer sentiment, if the business is able to compile and publish these indices on a regular basis, these indices would be able to supplement the media's analysis of consumer sentiment.
- Understand the dynamics between different credit products and their affect on consumption. This could help policy makers which type of products may require regulatory change (i.e. 2012 commission into payday loans)

Bringing data + analytics to life.

Geoff Chang

0424 076 890

gchang@cmcrc.com

Geoffrey.Chang@illion.com.au

WESTERN SYDNEY UNIVERSITY

illion
Formerly Dun & Bradstreet

ROZETTA
formerly CMCRC ASSURING MARKET QUALITY