ESSAYS ON THE GENERAL DETERMINANTS OF CONSUMPTION AND SAVINGS

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Economics

at the University of Canterbury

by: Zhongchen Song

Supervised by: Prof. William Robert Reed

Co-Supervision by: Associate Prof. Tom Coupé

University of Canterbury

2021

Abstract

This thesis consists of 4 studies linked together by my attempts to study the determinants and behavior of consumption and savings. Chapter One provides an introduction and background for this thesis. Chapter Two replicates Fiorito and Kollintzas (2004). This paper examines the crowding-out effect between government consumption and private consumption. My replication confirms their original findings by re-creating their dataset and estimation methods using the same sources listed in Fiorito and Kollintzas' appendix. Furthermore, I concluded that their results are robust when employing more recent data.

Chapter Three investigates why savings are so high in China from the perspective of the One-Child Policy (OCP). Using data from the 2014 Gallup World Poll and Global Findex database. I compare the saving behavior of Chinese people with people from regions that do not have restrictive population policies. These regions share many cultural, demographic, and economic characteristics with China, suggesting they can be used as a counterfactual for China. The rich dataset also enables me to adopt the Blinder-Oaxaca decomposition procedure to disentangle the different channels by which the OCP could affect savings. My results suggest that there is little difference in the savings behaviour of Chinese people with their regional counterfactuals, and my estimates are generally small. Therefore, I find no evidence to support that the OCP can explain China's high saving rate. My findings also suggest that the relaxation of the OCP is unlikely to increase Chinese consumption significantly.

Chapter Four focuses on using search engine data from Baidu and Google to predict consumption-related aggregates in China. Over the last 15 years, researchers have used search intensity data like Google Trends to analyze whether the volume of internet searches can help predict consumption and consumer behavior, while limited attention has been put on economies where other search engines like Baidu dominates the market. In Chapter Four, I investigate whether Baidu and Google can help to forecast total retail sales of consumption goods in China. I estimate both the baseline models and the models augmented with Baidu/Google search term series, using both OLS and

Lasso methodologies. My results show that adding information from Baidu search intensities to the baseline model can improve the accuracy of the predictions. Furthermore, the improved performance from the Baidu data is greater than that from Google Trends or Chinese Consumer Confidence surveys.

Chapter Five investigates whether the forecasting procedures I used for Chinese consumption would also be effective in the New Zealand context. To achieve this goal, I adopt a similar estimation procedure as Chapter Four to nowcast and forecast quarterly household consumption using data from Statistics New Zealand for the period 2005 Q1 to 2020 Q4. My results indicate that models with Google Trends reduce prediction errors by 18% for nowcasting and up to 45% for forecasting over a baseline OLS model with AR terms.

Chapter Six concludes this thesis. It provides an overview of my chapters, as well as a summary of my main findings.

Acknowledgements

I would like to express my gratitude to my primary supervisor, Professor Bob Reed, and my Co-supervisor, Associated Professor Tom Coupé, for their unwavering support and mentorship. They helped me with this thesis every step along the way and provided me with countless suggestions and advice both in research and in life. I have been so incredibly lucky to be able to work with them. Their knowledge and encouragement made the past three years rewarding and exciting.

I would like to thank the tremendous help I got from both the University of Canterbury and the Economics Department. The department sponsored me with tuition and provided the possibility to come back to New Zealand after the border closure as the result of the COVID-19 pandemic. I would also like to thank Associated Professor Andrea Menclova for being so kind and loving.

I would like to thank my friends for accompanying me every step along my Ph.D. Journey. Thank you, Samuel Strickland, for your constant inquiries into the progress of my Ph.D., you are the living embodiment of an alarm clock that reminds me of my unfinished thesis, and you won’t stop until you get your way. I would also like to thank all my Ph.D. colleges, which made me feel very loved and very under-studied. Your company made the past 3 years special and unforgettable.

Finally, I would like to thank my parents Song Changbing and Liu Shuhui. They are constantly under stress for their son not being able to achieve perfection. Their stories on how well the sons and daughters of their peers are doing have been a constant source of inspiration and self-critique for me. Nonetheless, I want to thank them for their unconditional love and support. I love them, and I miss them dearly. I hope they are happy that they have inspired me along this journey.

I am genuinely going to miss my time at Christchurch and UC. It’s been such an unforgettable journey. I love you all, and thank you all for your love and company!

相逢有期，万事顺意。

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Economists have long since been interested in understanding the factors that determine economic growth and promote well-being. The most widely adopted measure of a country's income is gross domestic product (GDP). By definition, GDP is the sum of private consumption, government spending, total investment, and net exports, where government spending includes both public consumption and investment. While the relative importance of these components differs across economies – with developing countries relying more on exports and investments – consumption generally constitutes the largest component of GDP. As indicated above, an economy's total consumption is made up of two parts: private consumption and public consumption. As a result, efforts to improve living standards often involve strategies to stimulate either private consumption, public consumption, or both.

This thesis represents my attempt to better understand the determinants and behaviour of consumption, with a focus on Chinese consumption. My first step in this investigation started with an investigation of the trade-off between public and private consumption. Do government efforts to stimulate economic growth via increased public consumption serve as a substitute or a complement to private consumption? Accordingly, I replicated a well-cited paper on the "crowding-out effect."

The "crowding-out" effect between government consumption and private consumption has drawn the attention of many scholars. Increased government spending can affect private consumption via several channels. For example, financing an increase in government spending by increasing taxes or selling government bonds can reduce private consumption, implicitly increasing government consumption at the expense of decreased private consumption. Many studies (e.g., Ho, 2001; Linnemann & Schabert, 2004; Afonso & Sousa, 2009; Coenen & Straub, 2005; Ramey & Shapiro, 1998) find a "crowding-out" effect between government consumption and private consumption. In contrast, other studies report a "crowding-in" effect (e.g., Fiorito & Kollintzas, 2004; Devereux, Head & Lapham, 1996; Fatás & Mihov 2001; Blanchard & Perotti; 2002; Giordano; 2007).

I decided to investigate this subject by replicating the paper, "Public goods, merit goods, and the relation between private and government consumption," by Fiorito and Kollintzas, published in the *European Economic Review* in 2004. This paper is well-cited in the literature on "crowding out." It has received 193 Google Scholar citations as of July 2021. It explores whether government consumption crowds out private consumption by studying the determinants of private consumption in 12 European countries from 1970 to 1996.

Fiorito and Kollintzas (2004 -- henceforth, F&K) estimate that different aspects of government consumption affect private consumption differently. By splitting government spending into different categories, they identified a complementary relationship between certain parts of government spending and private consumption. They associated this complementary relationship with positive externalities. For example, increased educational expenditure by the government increases the economy's human capital, and hence its productivity, enabling the economy to support a higher level of private consumption. F&K first builds a theoretical model of household consumption and then estimates it using cross-country, time-series panel data. My replication allows me to both confirm their original findings and determine whether they are robust to more recently available data. The results of my replication are important because they contribute to our understanding of how consumption responds to government policy.

In order to replicate their results, I first adopted their estimation procedure using data collected from the same sources they listed in the appendix to their article. Like F&K, I used data from the years 1970-1996. I then extended their empirical analysis by using more recent data from the years 1996-2014. The results from my extension generally matched F&K's findings, providing further evidence in support of F&K's conclusions.

I next turned my attention to Chinese consumption and savings. During the past decades, the Chinese economy has gradually grown to become one of the world's largest, as measured by aggregate GDP. China's savings rate (national savings to GDP) is also increasing, fluctuating between 40% and 50% in the recent decades – far in excess of nearly every other economy. This contributes heavily to the country's investment-led growth. However, recently attention has focussed on "rebalancing" growth from investment to consumption. There is a sense that for China to sustain its growth, it needs to transition from a high-saving/low-consumption economy into an economy that consumes more and saves less. This transition requires a deep understanding of the determinants of consumption and savings in China.

Many theories have been put forth to explain economies' aggregate saving behaviour. One crucial determinant identified by scholars is a country's demographic composition. An increase in the old and young dependency ratio has been found to have a negative impact on savings (cf. Ando & Modigliani, 1963; Modigliani & Cao, 2004; Loayza, Schmidt-Hebble & Serven, 2000; Horioka & Terada-Hagiwara, 2012). Income is also associated with savings because wealthier people tend to save more (Loayza, Schmidt-Hebble & Serven, 2000; Horioka & Terada-Hagiwara, 2012). Likewise, productivity is associated with a positive increase in savings (Loayza, Schmidt-Hebble & Serven, 2000; Corbo & Schmidt-Hebble, 1991). The real interest rate can also affect savings because an increase in interest rates raises the cost of current consumption. As a result, people are more inclined to substitute current consumption for more savings (Hondroyiannis, 2006). Some studies link savings with economic uncertainty (Hondroyiannis, 2006; Loayza, Schmidt-Hebble & Serven, 2000).

Existing literature has also investigated why savings are so high in China. Studies have claimed links between high household saving rates and rising property prices (Wang & Wen, 2012; Li, Whally, & Zhao, 2013; Chamon & Prasad, 2010), pension reforms (Chamon, Liu & Prasad, 2013; Feng, He & Sato, 2011), sex imbalances (Wei & Zhang, 2011). Of particular interest to me is the effect of China's population control policies; namely, its One Child Policy (OCP) (Zhou, 2014; Ge, Yang & Zhang, 2018; Lugauer, Ni & Yin, 2017; Curtis, Lugauer, & Mark, 2015; Choukhmane, Coeurdacier & Jin, 2014).

One can think of the impact of the OCP on savings via two channels: an "endowment effect" and a "coefficient effect." The endowment effect refers to the impact that the number of children has on savings, holding constant the effect that any given child has. The “coefficient effect” says that public policies can affect the relationship between children and savings even if the number of children is the same. Previous research on the effect of the OCP on Chinese savings ignored this latter effect. My research directly addresses both aspects of the OCP.

My approach is to compare Chinese people with people from regions that do not have restrictive population policies (Taiwan, Hong Kong, Singapore, Malaysia, Japan, and South Korea). I adopt the Oaxaca decomposition procedure to disentangle the two channels by which the OCP could have affected savings behaviour in China, isolating the endowment and coefficient effects. My results suggest that there is little difference in the savings behaviour of Chinese people with their regional counterfactuals. This is evidence against the hypothesis that the OCP was a major contributor to China’s high saving rate. It also suggests that the recent relaxation of the OCP cannot be counted on to boost Chinese consumption.

After investigating Chinese savings, I turned my attention to Chinese consumption. I became interested in the factors that affect Chinese consumption. This led me to the forecasting literature. Internationally, many scholars have focused on using survey-based indicators like the Consumer Confidence Index and the Consumer Sentiment Index to forecast consumption. Studies generally conclude that augmenting baseline models with survey-based indicators can improve forecasting performance, though this increase is dependent on country and model specification. Associated evidence comes from the US (Carroll, Fuhrer & Wilcox, 1994; Bram & Ludvigson, 1998; Howrey, 2001), as well as other countries (Cotsomitis & Kwan, 2006; Kwan & Cotsomitis, 2007; Dees & Brinca, 2013; Lahiri, Monokroussos & Zhao, 2015; Gausden & Hasan, 2018; Juhro & Lyke, 2020).

Recently, economists have been studying how the use of internet search volume data like Google Trends or Baidu Index can improve nowcasting and forecasting. Early efforts in this field focused on predicting the unemployment rate (Ettredge, Gerdes & Karuga, 2005; Askitas & Zimmermann, 2009; D’Amuri & Marcucci, 2017; Fondeur & Karame, 2013; Naccarato et al., 2018; Mihaela, 2020). From there, scholars demonstrated that Google Trends could improve predictions of consumption and sales

(Choi & Varian, 2012; Goel et al., 2010; Vosen & Schmidt, 2011; Woo & Owen, 2018; Carrière-Swallow & Labbé, 2013).

Despite this international evidence, there are relatively few studies on using internet search data to forecast Chinese consumption. Accordingly, my third study focuses on predicting consumption aggregates in China using internet search engine data. The existing literature has focused on Google Trends and its application to western economies. Due to its being banned in 2010, Google has not been widely used in China in recent years. Instead, Baidu is the search engine of choice for most Chinese. Only very limited attention has been given to using Baidu Index for forecasting in the Chinese context. This lack of attention motivates my chapter. I make predictions for aggregated total retail sales in China using a wide range of keywords from both the Baidu Index and Google Trends. By utilizing various machine-learning algorithms, I find that Baidu Index can produce more accurate nowcasts and forecasts of retail sales in China when incorporated into a baseline model.

Finally, since I plan to remain in New Zealand for a time after completing my thesis, I wanted to investigate whether the forecasting procedures I used for Chinese consumption would also be effective in predicting New Zealand consumption. I use a similar procedure as the previous chapter to forecast quarterly household consumption using data from Statistics New Zealand for the period 2005 Q1 to 2020 Q4. I show that models with Google Trends reduce prediction errors by 18% for nowcasting and up to 45% for forecasting over a baseline OLS model with AR terms.

1.2 References

Afonso, A., & Sousa, R. M. (2011). The macroeconomic effects of fiscal policy in Portugal: a Bayesian SVAR analysis. *Portuguese Economic Journal, 10*(1), 61-82.

Ando, A., & Modigliani, F. (1963). The" life cycle" hypothesis of saving: Aggregate implications and tests. *The American economic review*, 53(1), 55-84.

Askitas, N., & Zimmermann, K. (2009). *Google Econometrics and Unemployment Forecasting* (No. 4201). *DIW Berlin Discussion Paper NO.899*.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics, 117*(4), 1329-1368.

Bram, J., & Ludvigson, S. C. (1998). Does consumer confidence forecast household expenditure? A sentiment index horse race. *Economic Policy Review*, *4*(2).

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why?. *The American Economic Review*, *84*(5), 1397-1408.

Chamon, M. D., & Prasad, E. S. (2010). Why are saving rates of urban households in China rising?. *American Economic Journal: Macroeconomics, 2*(1), 93-130.

Chamon, M., Liu, K., & Prasad, E. (2013). Income uncertainty and household savings in China. *Journal of Development Economics, 105*, 164-177.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, *88*, 2-9.

Choukhmane, T., Coeurdacier, N., & Jin, K. (2013). The one-child policy and household savings. *CEPR Discussion Paper No. DP9688*.

Coenen, G., & Straub, R. (2005). Does government spending crowd in private consumption? Theory and empirical evidence for the euro area. *International Finance, 8*(3), 435-470.

Corbo, V., & Schmidt-Hebbel, K. (1991). Public policies and saving in developing countries. *Journal of Development Economics, 36*(1), 89-115.

Cotsomitis, J. A., & Kwan, A. C. (2006). Can consumer confidence forecast household spending? Evidence from the European commission business and consumer surveys. *Southern Economic Journal*, 597-610.

Curtis, C. C., Lugauer, S., & Mark, N. C. (2015). Demographic patterns and household saving in China. *American Economic Journal: Macroeconomics, 7*(2), 58-94.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33*(4), 801-816.

Dees, S., & Brinca, P. S. (2013). Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area. *International Economics*, *134*, 1-14.

Devereux, M. B., Head, A. C., & Lapham, B. J. (1996). Monopolistic competition, increasing returns, and the effects of government spending. *Journal of Money, credit and Banking, 28*(2), 233-254.

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, *48*(11), 87-92.

Fatás, A., & Mihov, I. (2001). The effects of fiscal policy on consumption and employment: theory and evidence. *CEPR Discussion Papers No. 2760.*

Feng, J., He, L., & Sato, H. (2011). Public pension and household saving: Evidence from urban China. *Journal of Comparative Economics, 39*(4), 470-485.

Fiorito, R., & Kollintzas, T. (2004). Public goods, merit goods, and the relation between private and government consumption. *European Economic Review, 48*(6), 1367-1398.

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment?. *Economic Modelling*, *30*, 117-125.

Gausden, R., & Hasan, M. S. (2018). An assessment of the contribution of consumer confidence towards household spending decisions using UK data. *Applied Economics*, *50*(12), 1395-1411.

Ge, S., Yang, D. T., & Zhang, J. (2012). Population policies, demographic structural changes, and the Chinese household saving puzzle. *Institute for the Study of Labor (IZA) Discussion Paper*, 7026.

Giordano, R., Momigliano, S., Neri, S., & Perotti, R. (2007). The effects of fiscal policy in Italy: Evidence from a VAR model. *European Journal of Political Economy, 23*(3), 707-733.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National academy of sciences*, *107*(41), 17486-17490.

Ho, T. W. (2001). The government spending and private consumption: a panel cointegration analysis. *International Review of Economics & Finance, 10*(1), 95-108.

Hondroyiannis, G. (2006). Private saving determinants in European countries: A panel cointegration approach. *The Social Science Journal, 43*(4), 553-569.

Horioka, C. Y., & Terada-Hagiwara, A. (2012). The determinants and long-term projections of saving rates in Developing Asia. *Japan and the World Economy, 24*(2), 128-137.

Howrey, E. P. (2001). The predictive power of the index of consumer sentiment. *Brookings papers on economic activity*, *2001*(1), 175-207.

Juhro, S. M., & Iyke, B. N. (2020). Consumer confidence and consumption expenditure in Indonesia. *Economic Modelling*, *89*, 367-377.

Kwan, A. C., & Cotsomitis, J. A. (2006). The usefulness of consumer confidence in forecasting household spending in Canada: A national and regional analysis. *Economic Inquiry*, *44*(1), 185-197.

Lahiri, K., Monokroussos, G., & Zhao, Y. (2016). Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics*, *31*(7), 1254-1275.

Li, S., Whalley, J., & Zhao, X. (2013). Housing price and household savings rates: evidence from China. *Journal of Chinese Economic and Business Studies, 11*(3), 197-217.

Linnemann, L., & Schabert, A. (2004). Can fiscal spending stimulate private consumption?. *Economics Letters, 82*(2), 173-179.

Loayza, N., Schmidt-Hebbel, K., & Servén, L. (2000). What drives private saving across the world?. *Review of Economics and Statistics, 82*(2), 165-181.

Lugauer, S., Ni, J., & Yin, Z. (2017). Chinese household saving and dependent children: Theory and evidence. *China Economic Review*, *57*, p.101091.

Mihaela, S. (2020). Improving unemployment rate forecasts at regional level in romania using google trends. *Technological Forecasting & Social Change, 155*, 120-026.

Modigliani, F., & Cao, S. L. (2004). The Chinese saving puzzle and the life-cycle hypothesis. *Journal of economic literature, 42*(1), 145-170.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, *130*, 114-122.

Ramey, V. A., & Shapiro, M. D. (1998, June). Costly capital reallocation and the effects of government spending. *In Carnegie-Rochester Conference Series on Public Policy* (Vol. 48, pp. 145-194).

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of forecasting*, *30*(6), 565-578.

Wang, X., & Wen, Y. (2012). Housing prices and the high Chinese saving rate puzzle. *China Economic Review, 23*(2), 265-283.

Wei, S. J., & Zhang, X. (2011). The competitive saving motive: Evidence from rising sex ratios and savings rates in China. *Journal of Political Economy, 119*(3), 511-564.

Woo, J., & Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting*, *38*(2), 81-91.

Zhou, W. (2014). Brothers, household financial markets and savings rate in China. *Journal of Development Economics*, 111, 34-47.

Chapter 2： A replication of “Public goods, merit goods, and the relation between private and government consumption” by Riccardo Fiorito and Tryphon Kollintzas, 2004”

2.1 Overview of F&K, 2004

The first chapter of my thesis is a replication of F&K's article, "Public goods, merit goods, and the relationship between private and government consumption" (European Economic Review, 2004). This paper provided important theoretical and empirical evidence of the "Crowding out effect" between government consumption and private consumption, and investigated whether different branches of government consumption affect private consumption differently. This paper has received 37 Web of Science citations and 163 Google Scholar citations as of October 2019. It has been cited by papers in many top journals, such as the Economic Journal; Journal of Money, Credit, and Banking; American Economic Journal – Macroeconomics; Journal of Applied Econometrics; and the Journal of Economic Dynamics and Control.

F&K focused on the "Crowding out effect" between government consumption and private consumption. They split government consumption into two categories: "Public goods" and "Merit goods." "Public goods" include defense, public order, and justice, which are difficult to supply privately. "Merit goods" include health, education, and other services that can be privately provided. The authors develop a model of household spending behavior in the presence of government spending. They then estimate the model using difference GMM applied to data for 12 OECD countries.

Their study is motivated by the inconclusive evidence in the existing literature studying the response of economic aggregates to changes in government consumption composition. This response is dependent on whether government consumption substitutes private consumption or not. Theoretically, their motivation comes from the fact that different types of government consumption may affect private consumption differently. For example, expenditures in "public goods" defence, public order, and justice are typically the sectors that are difficult to be provided privately, while "merit goods," which includes health and education services, may be easy for the private sector to gain access to due to the low entry barrier. So, the important differences in the very nature of these goods may lead to different responses in private consumption.

The findings of F&K suggested a substitute relationship between "public goods" and private consumption and a complement relationship between "merit goods" and private consumption. They further concluded that since the “merit goods” consumption is about two-thirds of the government consumption, the aggregate government consumption complements private consumption.

This section proceeds by providing a brief literature review of the crowding effect. It then presents F&K’s theoretical model and explains how it is put to data. I then present F&K’s empirical findings along with my replication of their estimates. In the replication process, I first replicate their study using the same countries and time period (1970-1996) and then extend my replication to include more recent years (1996-2014).

2.2 The crowding-out effect

A fundamental question in economics is whether, and to what extent, government intervention can stimulate and grow the economy. Included in this is the role of fiscal policy.

The theory of public goods identifies a role for public sector spending to address shortcomings in the private sector’s provision of certain types of goods. This can arise, for example, when individuals do not realize or incorporate the value of benefits that extend beyond themselves (positive externalities). Examples include goods or services that have spillover benefits to many people (e.g., hospitals, parks) or when provisions are inadequate due to incomplete markets (e.g., insurance markets in the presence of a moral hazard or “lemons”) (Payne, 2009).

In addition to these reasons for government spending, public expenditures are frequently aimed at stimulating the private sector, in order to increase the provision of goods and services to the private sector. However, is the government successful in these practices? According to economic theory, when government intervenes, private sector activities in the relevant field may decrease. This theory of government expenditure canceling out private expenditure is commonly referred to as the “crowding out” effect. Empirical verification of the theoretical possibility of “crowding out” is crucial if policy makers are to know whether their stimulatory policies are to be effective.

A study of the effect of government spending on private consumption is also important because of its implication for private saving. Fast-growing economies are often accompanied by high savings rates. Prominent examples include East Asian economies such as Japan in the 1970s, Korea in the 1980s, and, of course, China in recent decades.

Overall, the literature studying the impact of government consumption on economic activities has provided extensive but mixed evidence on the effect of government consumption on private consumption and investment. Christiano & Eichenbaum (1995) and Baxter & King (1993) develop real-business cycle models that show that when private consumption falls due to the negative wealth effect, households will work more but consume less. Further, households' participation in asset markets to smooth consumption serves to amplify this effect. Devereux, Head & Lapham (1996) studied linked government spending shocks with private consumption and found that when government consumption increases, this generates a rise in aggregate productivity. An increase in real wages is then generated because of the increase in productivity, which contributes to consumption. Thus, an increase in government expenditures leads to an increase in private consumption.

Gali et al. (2007) investigate the effects of government spending on consumption using a New Keynesian model. Their model includes both Ricardian and Non-Ricardian households. They find that Ricardian households are less sensitive to government-stimulated changes in income when higher taxes are needed to finance fiscal expansion. This is because Ricardian households cut back their consumption in anticipation of future tax increases.

Private consumption is also likely to be influenced by government expenditure depending on the type of fiscal policy being implemented. For example, social welfare expenditures aimed at helping low-income households and credit-constrained agents are likely to positively affect private consumption (Furceri & Zdzienicka, 2011). Further, a fiscal policy aimed at helping the formation of human capital is likely to create long-term positive effects, increasing private investment and consumption (Easterly & Rebelo 1993).

Several studies have been conducted to assess the debate from an empirical perspective view. However, the results have been inconclusive. Results are sensitive to the countries chosen and the time span of the dataset being used. Afonso & Sousa (2011) studied the relationship between government expenditure and private consumption for 4 OECD countries (US, UK, Germany and Italy), using a VAR model. They did not find a significant relationship between the two. Later, in 2011, they conducted a similar empirical analysis of Portugal and found a negative effect.

Similarly, Mountford & Uhlig (2009) did not find a significant relationship between government spending and private consumption. In contrast, Fatás & Mihov (2001) and Blanchard & Perotti (2002), also using US data, found a crowding-in effect (positive effect) between government spending and private consumption. Giordano et al. (2007) also estimated a positive effect of government consumption on private consumption using Italian data.

2.3 F&K's Theoretical background and model.

F&K, building on Abel (1990) and Campbell & Cochrane (1999), construct a model of household consumption behaviour that incorporates habit formation, where the latter characteristic is designed to accommodate the observed effect of lagged variables on current consumption. Their model is based on the following assumptions:

1) The household's utility is dependent on current consumption as well as the household's "habit level."
2) The household's habit level is determined by the past consumption of all households in the economy.
3) The expectations of households regarding their future habit level are rational. In other words, along the steady-state equilibrium path, current consumption conditional on current habit levels must be consistent with past consumption.

The representative economic agent's preferences are characterized by the conditional expectation of lifetime utility, represented by:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, g_t, m_t, h_t^c, h_t^g, h_t^m) \tag{2.1}$$

In the above equation, E_0 is the expectation operator at time *0*, β^t is a time-homogeneous time discount factor, $u(c_t, g_t, m_t, h_t^c, h_t^g, h_t^m)$ is a utility function based on the consumption of private goods, c_t, public goods, g_t, merit goods, m_t, and the current habit levels of the three kinds of consumption goods, $h_t^c, h_t^g, \text{and } h_t^m$.

For an individual household, its aim is to maximize expected lifetime utility subject to its budget constraint. The budget constraint assumes that the representative household, by giving up one unit of consumption in any period t, receives a stochastic real, gross, after-tax return of R_{t+1} in the next period according to the Euler equation:

$$u_{c_t} = \beta E_t u_{c_{t+1}} R_{t+1} \tag{2.2}$$

The Euler equation states that people should be indifferent between consuming today and saving up and consuming in the future. The left-hand side of the equation represents the reward associated with consumption today, u_{c_t}. The right-hand side represents the reward of foregoing consumption today in order to save and enjoy consumption in the next period. If the consumer saves 1 unit today, he receives R_{t+1} consumption units in the next period, which produces $u_{c_{t+1}}$ units of utility at time period *t+1*. Because this utility comes in the future, it must be discounted by the factor β. This constitutes the right-hand side of the equation. The two sides of this equation must be equal to make sure that the consumer is indifferent between consuming today and consuming in the future.

The Euler condition may be used to illustrate the concept of substitutability/ complementarity. Specifically, according to the Euler equation above, the right-hand side is the expected discounted benefit from one unit of assets invested in the current period. We can interpret it as the opportunity cost of consumption in this period, which can be represented by p_{c_t}, so that $u_{c_t} = p_{c_t}$.

Accordingly, this equation represents the demand for current private consumption. If private consumption and public goods consumption are substitutes (complements), an increase in public goods consumption lowers (increases) the demand for private consumption at any given price, p_{c_t} . This situation is depicted in Figure 1.

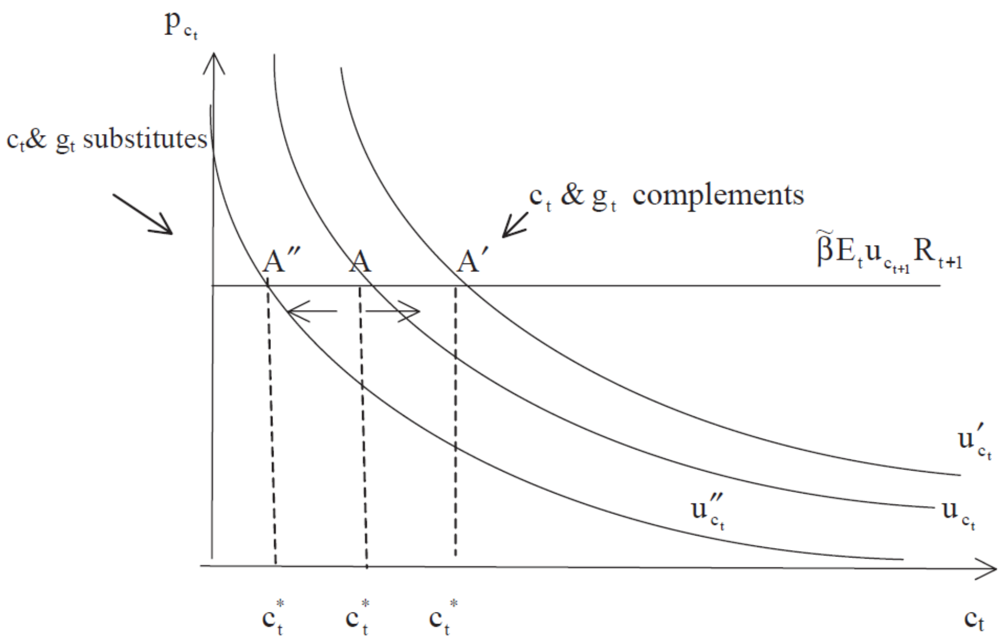


Figure 2.1

Figure 2.1 allows us to identify the relationship between private consumption and public goods and merit goods consumption. In particular, we can exploit the assumption of strict concavity of the utility function to identify these relationships as follows:

The Euler condition can be written as:

$$U = E_0[u_{c_t}(c_t, g_t, m_t, h_t^c, h_t^g, h_t^m) -$$

$$- \check{\beta} u_{c_{t+1}}(c_{t+1}, g_{t+1}, m_{t+!}, h_{t+1}^c, h_{t+1}^g, h_{t+1}^m) R_{t+1}] \tag{2.3}$$

F&K assumes that the current habit levels of households are equal to the economy-average consumption levels of the previous period. As a result, we can write:

$$U = E_0[u_{c_t}(c_t, g_t, m_t, c_{t-1}, g_{t-1}, m_{t-1}) -$$

$$- \check{\beta} u_{c_{t+1}}(c_{t+1}, g_{t+1}, m_{t+1}, c_t, g_t, m_t) R_{t+1}] \tag{2.4}$$

F&K goes on to show that the above can be expanded using a Taylor series, first-order approximation around the point *(c,g,m,R)* and that this leads to the following regression equation specification:

$$\Delta\check{c}_{t+1} = \alpha_1\Delta\check{g}_{t+1} + \alpha_2\Delta\check{m}_{t+1} + \alpha_3\left(\check{R}_{t+1} - \check{R}\right) + \alpha_4\Delta\check{c}_t + \alpha_5\Delta\check{g}_t + \alpha_6\Delta\check{m}_t + \varepsilon_{t+1}\,. \quad (2.5)$$

In the above equation, Δ stands for the difference operator, the "ˇ" symbol indicates that the natural logarithm of the variable is taken, $E(\varepsilon_{t+1}) = 0$, and the respective slope coefficients are defined as follows:

$$\alpha_1 = -u_{cg}g/u_{cc}c, \quad (2.6)$$

$$\alpha_2 = -u_{cm}m/u_{cc}c, \quad (2.7)$$

$$\alpha_3 = -\check{\beta}u_c/u_{cc}c, \quad (2.8)$$

$$\alpha_{4,}\alpha_{5,}\alpha_6 = -u_{ch}/u_{cc}\ , h = h^c, h^g, h^m; \quad (2.9)$$

where u_{cx} represents the second partial derivative of *u* with respect to *c* and *x*; and *x = (c,g,m,h)*, evaluated at *(c,g,m,h)*. Because u_{cc} must be strictly negative according to the concavity of the utility function, the signs of α_1 and α_2 depend completely on u_{cg} and u_{cm}. This leads to the following corollary.

<u>COROLLARY 1</u>:

If u is strictly concave in c; c and g(m) are substitutes, if and only if

$$\alpha_1 < 0\ (\alpha_2 < 0).$$

They are independent if

$$\alpha_1 = 0\ (\alpha_2 = 0).$$

And they are complements if

$$\alpha_1 > 0\ (\alpha_2 > 0).$$

According to this corollary, it is straightforward to identify the relationship between private consumption and the two types of government spending from the signs of the coefficients on the contemporaneous government spending variables from estimation of equation 2.5. Further, because the variables are measured in logs, the respective coefficients can be interpreted as elasticities.

The subsequent empirical analysis will focus on these two coefficients. In particular, we will use the signs of these coefficients to identify the complementarity or substitutability between private consumption and the two different types of government spending.

F&K's model also yields confirmatory checks on the underlying theory, which derives from the permanent income model with external habit formation. The following corollary covers these checks.

COROLLARY 2:

(1) If u is a strictly increasing and strictly concave in c, then $\alpha_3 > 0$.

(2) If there is external habit formation in private consumption, $\alpha_4 > 0$.

2.4 Estimation model.

F&K use dynamic panel estimation to estimate equation 2.6, which is simply equation 2.5 with the time periods backed up one period:

$$\Delta\check{c}_t = \alpha_1 \Delta\breve{g}_t + \alpha_2 \Delta\breve{m}_t + \alpha_3\left(\check{R}_t - \check{R}_{t-1}\right) + \alpha_4 \Delta\check{c}_{t-1} + \alpha_5 \Delta\breve{g}_{t-1}$$
$$+\alpha_6 \Delta\breve{m}_{t-1} + \alpha_7 \Delta d_t + \varepsilon_t \qquad (2.10)$$

After they show the estimation results for equation 2.10, they add an additional variable to control for income.

$$\Delta\check{c}_t = \alpha_1 \Delta\breve{g}_t + \alpha_2 \Delta\breve{m}_t + \alpha_3\left(\check{R}_t - \check{R}_{t-1}\right) + \alpha_4 \Delta\check{c}_{t-1} + \alpha_5 \Delta\breve{g}_{t-1}$$
$$+\alpha_6 \Delta\breve{m}_{t-1} + \alpha_7 \Delta d_t + \alpha_8 \Delta yd_t + \alpha_9 \Delta yd_t + \varepsilon_t \qquad (2.11)$$

$\Delta\check{c}$, $\Delta\breve{g}$, and $\Delta\breve{m}$ have all been previously defined. d_t is the working-age population share, $\check{R}$ is the logarithm of the after-tax, real interest rate and yd is household disposable income.

Working-age population share is a preference shifter accounting for the possibility that the relationship between government consumption and private consumption might be affected by demographic factors. F&K define the after-tax, real interest rate as:

$$r_{it} = ln \left(\frac{1 + \left((1 - tauc) * \left(\frac{irs_{it}}{100} \right) \right)}{\frac{pc_{it}}{pc_{it-1}}} \right), \qquad (2.12)$$

where *pc* is the household consumption deflator, *irs* is the short-run interest rate, and *tauc* is the effective tax rate on consumption calculated as in Fiorito & Padrini (2001).

The authors used difference GMM, with past level values of the private consumption variable and past differenced values of the right-hand side explanatory variables as instruments. They employed the Newey-West robust covariance estimator for the associated weighting matrix to address both heteroskedasticity and serial correlation.

2.5 Data.

In the original paper, the authors used a balanced panel dataset for 12 OECD countries, namely Austria, Denmark, Finland, France, Germany, Greece, Italy, Norway, Portugal, Spain, Sweden, and the United Kingdom, between 1970 and 1996. The data used for regression were drawn from the following databases:

1) National Accounts (1999) 1.1 Main aggregates (vol.1)
2) National Accounts (1999) 1.2 Detailed Tables (vol.2)
3) OECD economic outlook database.

In the first part of my replication study, I tried to recreate their main finding and estimation results by adopting the same dataset that they used. I first tried to contact the authors and asked them if they still had the original dataset. Unfortunately, they were unable to provide me with the data as they didn't have it anymore.

The authors did provide a statistics appendix at the end of their paper detailing the sources they used for data collection. I found the majority of the data at online databases, but there were many missing observations.

More than half of the countries had missing data between 1970 and 1980. These data were not available in the online database of the OECD National Account Database. I contacted the staff working for the database. However, they were unable to produce the missing data. Further, they were unable to provide any explanations for why the data were missing. On the plus side, they did mention that I might have success finding the data in extant book copies of the vintage datasets published decades ago.

Accordingly, I began searching for printed book copies of the data. After contacting several different sources, I was able to find one book copy of the database at the Western Washington University Library, and two book copies at the University of Auckland. I had these shipped to me using the University of Canterbury interloan system. These enabled me to fill in some of the gaps in my database.

Unfortunately, I was not able to reconstruct the full dataset for all countries and all years. In the end, I found data for approximately 80% of the observations. This resulted in my estimation results being similar, but still somewhat different, from the original paper, as I will show in the following paragraphs.

F&K's dataset used observations from 1970 and 1996. As part of my replication study, I extended their study by updating the time span of this data set to 2014, which is the most recently available year documented by the National Accounts database.

In categorizing the different types of government spending, I follow F&K, which, in turn, relies on the United Nations Classification of the Functions of Government (COFOG) spending categories. The COFOG classification consists of 10 spending categories:

1) General public services
2) Defence
3) Public order and safety
4) Education
5) Health
6) Social security and welfare
7) Housing and community amenities
8) Recreational, cultural and religious affairs

9) Economic services

10) Other functions

In their main specifications, F&K defined "public goods" as the sum of 1, 2, and 3; and "merit goods" as the sum of Items 4-8. Subsequent robustness checks considered modifications to these groupings.

F&K provided some descriptive statistics in Table 4 of their original paper. I updated their table using data from 2015. The results are listed in Table 1. Overall, there hadn't been major changes in the spending composition between 1995 and 2015. Public goods still take up a relatively small share of government spending, and most of the spending increase in public goods is associated with general public service. However, the increase in spending for merit goods is in general small.

It is worth noticing that some changes occurred with the data documented in the National Account database during this time.

The database gradually converted from the old 1968 System of National Accounts (SNA) to the 2008 SNA. The original, vintage data only reported data following the 1968 SNA. More recent data only reported data following the 2008 SNA. Data in the "middle" reported both. To address this problem, I collected data according to the 1968 SNA for the 1970 to 1996 dataset (matching F&K). For the more recent data (1996-2014), I used data that exclusively followed the 2008 SNA, which is the only consistent SNA available.

In addition, several countries in my dataset joined the Euro zone in the 1990s. This resulted in a shift in the currency for these countries. For example, the government spending dataset according to the COFOG classification for France are denominated in francs before 1996 according to the 1968 SNA system, but denominated in Euros according to the 2008 SNA system from 1996 onwards. These changes in SNA system resulted in a structural break in my data for many of the countries. Figure 2.1 shows the natural logarithm of private goods consumption plotted against years for the countries examined in this thesis. The graph indicates that for many of the countries, there are indeed breaks in the data around the time that countries shifted from their original currencies into Euros. I deal with this problem by keeping the two datasets separate

(1970-1996, 1996-2014), using the respective currency for each time period. The latter data are reserved for my “extension analysis”. In addition, for the countries that didn’t have a shift in currency, I still didn’t treat them as a single dataset as I believe it is better to treat the data as separate samples since there are fundamental changes in the SNA system that they are documented in.

Working-age population share and other data used to construct the after-tax, real interest rate variable are collected using the OECD Economic Outlook database. All of the deflators used in my datasets are taken from National Accounts databases.

My final dataset includes data for Austria, Denmark, Finland, France, Germany, Greece, Italy, Norway, Portugal, Spain, Sweden, and the United Kingdom for the years 1970-2014. The data for 1970-1996 are unbalanced, while the data for 1996-2014 are balanced.

With respect to replicating F&K, I was able to assemble 232 observations. While F&K does not report their total number of observations, if their data are balanced, as they say, they should have used 297 observations in their estimations. In my subsequent “extension analysis,” I use 196 observations.

Another issue I faced concerned the deflation of government expenditures. Government expenditure by function data from OECD sources is nominal data. Further, OECD sources do not report separate deflators for public goods and merit goods; only aggregate government and household consumption deflators. I followed the authors’ procedures when stated. When no guidance was available, I estimated alternative versions using different deflators.

A conceptual issue in taking F&K’s model to data was matching the COFOG functional classifications to the concepts of “public goods” and “merit goods”. I followed the authors’ approach to this problem. They defined “wider” and “smaller” variables, where only well-defined spending categories were included in the “smaller” public goods and merit goods categories. The “smaller” public goods variable included only public order and defence expenditures. The “smaller” merit goods variable included only education and health expenditures. Table 2.1 below summarizes the variables used in F&K and my corresponding replication study.

2.6 Results of replication

In recreating F&K's results, I attempted to reproduce not only their data, but also their empirical procedures. F&K reported that they used difference GMM, with differenced values from past periods serving as instruments. In the main estimation results in their original TABLE 2.5, they first estimated Equation 2.10 without household income. In their estimation results for Equation 2.11, they added an additional variable to control for household income.

F&K did not mention the specific estimation commands they used to produce their results. My first attempt to reproduce their results used Stata's built-in GMM procedure, *xtabond*. However, the *xtabond* command does not allow lagged differenced variables to be used as instruments, and F&K specifically states that they used both lagged level and lagged differenced variables.

I next used Stata's *xtabond2* procedure, which did allow me to use both lagged level and differenced variables as instruments. I obtained results that were moderately similar to F&K, but I knew something was not right because the standard *xtabond2* procedure uses all possible lagged values as instruments. In my case, that meant more than two hundred instruments, far more than F&K report using.

A major accomplishment in my replication efforts occurred when I discovered that Stata's *xtabond2* command includes a "collapse" option that limits the number of instruments used in the estimation. While F&K doesn't mention using this option, I found that when I chose it, I was able to produce results very similar to what F&K reported. Further, the associated number of instruments and over-identifying restrictions matched the numbers reported in F&K's tables.

Tables 2.2 through 2.7 below report the results of my replication analysis. The difference between the results shown in Tables 2.2 to 2.4 and the results in Tables 2.5 to 2.7 is that the regression results in Tables 2.5 to 2.7 are augmented by household disposable income *yd*. Column 1 reports F&K's estimates; Column 2 reports my replication of their results using data for the years 1970-1996 – the same year included in F&K's study; and Column 3 shows the estimation of the same specification using data for the years 1996-2014. While F&K estimates many specifications, Table 5 in their article contains their main results; and Columns 2, 3, and 5 contain the main results from that table. These are the results I focus on in my Tables 2.2 through 2.7. Note that these results excerpt estimates from the full set of results. The full set of estimates are reported in Tables 2.8 through 2.13 in the Appendix, where Table 2.8 and 2.11 is F&K's results, Table 2.9 and 2.12 is the result I get by using the data I gathered according to their specification and time period, Table 2.10 and 2.13 is the "extension" part of the analysis where I use their model specification and adopted data between 1996 and 2014. The difference between Tables 2.8 to 2.10 and Tables 2.11 to 2.13 is that the latter are augmented by household disposable income *yd*.

From the perspective of determining whether merit goods (*m*) and public goods (*g*) are substitutes or complements for private consumption, the key parameters are α_2 and α_1 in Equation 2.10, the estimates corresponding to these parameters are in bold font in Table 2.2-2.7, respectively. F&K concludes that merit goods are complements to private consumption, and public goods are substitutes. They base this conclusion on their finding that the signs of the coefficients for α_2 and α_1 are consistently positive and negative, respectively. Interestingly, they do not comment on the fact that the estimated coefficient for the public goods variable, α_1, is generally statistically insignificant.

Column 1 of Tables 2.2 through 2.7 reproduce F&K's estimates. The tables differ in that they vary the composition of the government spending categories used to create the aggregate spending variables for merit goods (*m*) and public goods (*g*). Tables 2.2 and 2.5 show the estimation results for *m1* and *g1*. The alternative variables are respectively identified as *m2* (cf. Table 2.3 and 2.6) and *g2* (cf. Table 2.4 and 2.7). I bold font the cells in the tables that correspond to the estimates of α_2 and α_1.

With respect to merit goods (*m*), F&K report estimates for α_2 of 0.498, 0.504, and 0.610 in Tables 2.2-2.4, respectively. When they add income, they obtain estimates of

0.534, 0.554, and 0.661 (cf. Tables 2.5-2.7). All estimates are significant at the 5-percent level. The coefficients can be interpreted as indicating that a one percent increase in merit good production by the public sector increases private consumption by approximately 0.5 to 0.6 percent.

Using the data I collected for the same time period as F&K (1970-1996), I obtained corresponding estimates of 0.535, 0.426, and 0.490, and estimates of 0.471, 0.388, and 0.477 when controlled for income. While my point estimates are similar to F&K's, they are generally less significant than what F&K report. None are significant at the 1 percent level when not controlling for income. Instead, the results are significant at the 10 percent level, insignificant, and significant at the 5 percent level, respectively. However, I did get more significant results after augmenting for income.

A stronger result is obtained when I replicate F&K using new data from 1996-2014. The results are shown in Column 3 in Tables 2.2-2.7. For these data, the corresponding estimates for α_2 are 0.516, 0.442, and 0.447. When I control for income, the corresponding estimates for α_2 are 0.357, 0.386 and 0.373. All estimates are significant at the 1 percent level.

Taken together, I interpret these results as providing strong evidence in confirmation of F&K's findings that merit goods and private consumption are complementary. In other words, government spending on education, health, and related goods that can in principle be supplied by the private sector, serves to stimulate private spending in the same areas.

Turning now to public goods (g), we see that F&K produced estimates of α_1 of -0.067, -0.062, and 0.055. After controlling for income, they reported estimates of -0.123, 0.012, and -0.029. These elasticities are very small. In all cases, a one percent increase in public good spending is estimated to produce less than a 0.1 percent change in absolute value in private consumption. Further, none of the estimates are significant at the 10 percent level. Given this, it is difficult to understand how F&K could have concluded that public goods are substitutes for private consumption. The results indicate, instead, that public goods and private goods are independent (cf. Corollary 1).

My replication results for both the 1970-1996 and 1996-2014 data confirm the independence of private and public goods consumption. I obtain estimates that range from -0.106 to 0.129, in all six instances before controlling for income and estimates ranging from -0.010 and 0.088 afterwards. The estimates are statistically insignificant at the 10 percent level.

In summary, my results confirm F&K's empirical estimates on public goods, though my results conflict with their interpretation of their results. While they conclude that public goods and private consumption are substitutes, my replication of their results, and indeed, their own results, indicate that these goods are independent.

Overall, I interpret my replication as confirming F&K. I find this all the more impressive in that I get results very close to what F&K obtained even when I use entirely new data from more recent years. The empirical results are consistent with the interpretation that merit goods and private consumption are complements, while public goods and private consumption are independent.

2.7 Conclusion

This study replicated F&K (2004) and extended their analysis by adopting more recent data. By gathering data using the statistical appendix they provided, I am able to match their specification and produce very similar results compared with their original study. I further confirmed their results by running their model specification using more recent data.

F&K concluded that there exists a complementary relationship between "merit good" consumption and private consumption, and a substitutionary relationship between "public good" consumption and private consumption. My replication results confirmed their empirical estimates for both "merit good" and "public good" consumption when using the same data and specification as they did in their original paper, and when using more recent data between 1996 and 2014.

However, my replication identified a discrepancy in F&K's interpretation of their own results. While they correctly interpreted the empirical results for "merit good" consumption, the interpretation of their “public good” consumption results was inconsistent with their estimates. F&K estimate, and my replication confirms, a small and statistically insignificant relationship between “public good” consumption and private consumption. This indicates that “public goods” consumption and private consumption are independent, not substitutes, as F&K concludes.

These results confirm that government spending tends to affect private consumption differently depending on the composition of government expenditure. As a result, policy-makers should account for the different responses of different kinds of government spending when making fiscal decisions.

2.8 Appendix

Figure 2. 1 private goods consumption by country

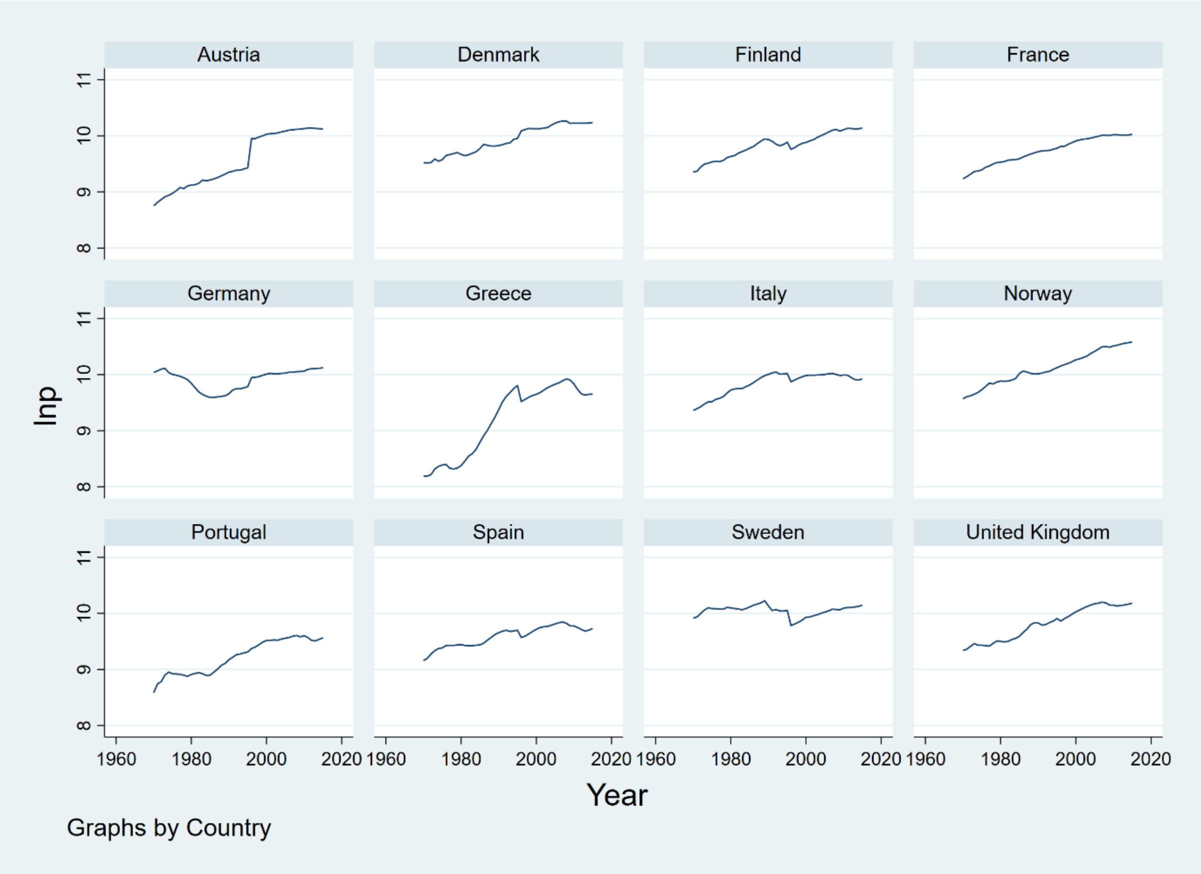


Table 2.1 Variables used by F&K and my replication study

Variable	Definition
C	Per capita household consumption in real terms
m	Merit goods in real terms (household consumption deflator)
m2	Education and health government consumption in real terms (household consumption deflator)
m3	Merit goods in real terms (government consumption deflator)
g	Public goods in real terms (household consumption deflator)
g2	Public order and defence government consumption in real terms (household consumption deflator)
g3	Public order and defence government consumption in real terms (government consumption deflator)
yd	Logged per capita household disposable income in real terms (household consumption deflator)
d	Working age population share
r	After tax real interest rate

Source: Working age population share is from the OECD economic outlook database. After tax real interest rate is calculated using data from the OECD economic outlook database. All other variables are from the OECD National Accounts Database

Table 2.2 Replication of Column 2, Table 5 from F&K

Variables	F&K	Replication (1970-1996)	Replication (1996-2014)
$\Delta \boldsymbol{c}_{i,t-1}$	0.773***	0.827***	0.843***
	(0.11)	(0.129)	(0.154)
$\Delta \boldsymbol{m}_{i,t}$	**0.498*** **	**0.535***	**0.517*** **
	(0.063)	**(0.306)**	**(0.090)**
$\Delta \boldsymbol{m}_{i,t-1}$	-0.277***	-0.437*	-0.430***
	(0.048)	(0.256)	(0.117)
$\Delta \boldsymbol{g}_{i,t}$	**-0.067**	**0.055**	**-0.106**
	(0.063)	**(0.199)**	**(0.097)**
$\Delta \boldsymbol{g}_{i,t-1}$	-0.093**	-0.049	-0.011
	(0.038)	(0.171)	(0.075)
$\Delta \boldsymbol{d}_{i,t}$	-0.375***	0.336	-0.420
	(0.134)	(0.316)	(0.410)
$\Delta \boldsymbol{r}_{i,t}$	0.082	0.011	-0.005
	(0.193)	(0.012)	(0.006)
J-test	$X^2(15) = 15.3$	$X^2(15) = 7.14$	$X^2(15) = 8.55$
	P =0.426	P =0.954	P =0.900
Observations	297	232	196

Table 2.2 reports the results from estimating equation (6). Newey-West standard errors are in parentheses. Apart from age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels. The bold font cells highlight the main variables of interest. They correspond to the estimated coefficients for merit goods (*m*) and public goods (*g*).

Table 2.3 Replication of Column 3, Table 5 from F&K

Variables	F&K	Replication (1970-1996)	Replication (1996-2014)
$\Delta c_{i,t-1}$	0.464***	0.693***	1.001***
	(0.127)	(0.171)	(0.257)
$\Delta m2_{i,t}$	**0.504*** **	**0.426**	**0.442*** **
	(0.081)	**(0.286)**	**(0.140)**
$\Delta m2_{i,t-1}$	-0.01	-0.253	-0.407***
	(0.066)	(0.231)	(0.064)
$\Delta g_{i,t}$	**-0.062**	**0.129**	**-0.085**
	(0.088)	**(0.249)**	**(0.093)**
$\Delta g_{i,t-1}$	-0.248***	-0.106	-0.104
	(0.062)	(0.233)	(0.121)
$\Delta d_{i,t}$	-0.218*	-0.112	-0.551
	(0.124)	(0.317)	(0.435)
$\Delta r_{i,t}$	0.139	0.020	-0.007
	(0.173)	(0.012)	(0.007)
J-test	$X^2(13) = 11.4$	$X^2(13) = 6.06$	$X^2(13) = 2.51$
	P =0.575	P =0.944	P =0.990
Observations	297	232	196

Table 2.3 reports the results from estimating equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels. The bold font cells highlight the main variables of interest. They correspond to the estimated coefficients for merit goods (m) and public goods (g).

Table 2.4 Replication of Column 6, Table 5 from F&K

Variables	F&K	Replication (1970-1996)	Replication (1996-2014)
$\Delta \boldsymbol{c}_{\boldsymbol{i},\boldsymbol{t-1}}$	0.543*** (0.144)	0.870*** (0.112)	0.880*** (0.190)
$\Delta \boldsymbol{m}_{\boldsymbol{i},\boldsymbol{t}}$	**0.610*** (0.082)**	**0.488** (0.246)**	**0.448*** (0.170)**
$\Delta \boldsymbol{m}_{\boldsymbol{i},\boldsymbol{t-1}}$	-0.359*** (0.103)	-0.411** (0.193)	-0.372* (0.194)
$\Delta \boldsymbol{g2}_{\boldsymbol{i},\boldsymbol{t}}$	**0.011 (0.073)**	**0.042 (0.123)**	**-0.090 (0.115)**
$\Delta \boldsymbol{g2}_{\boldsymbol{i},\boldsymbol{t-1}}$	0.055 (0.047)	-0.037 (0.105)	-0.073 (0.164)
$\Delta \boldsymbol{d}_{\boldsymbol{i},\boldsymbol{t}}$	-0.203* (0.115)	0.280 (0.388)	-0.137 (0.476)
$\Delta \boldsymbol{r}_{\boldsymbol{i},\boldsymbol{t}}$	0.428* (0.229)	0.012 (0.010)	-0.006 (0.006)
J-test	$X^2(13) = 11.7$ P =0.548	$X^2(13) = 6.55$ P =0.924	$X^2(13) = 3.98$ P =0.991
Observations	297	232	196

Table 2.4 reports the results from estimating equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels. The bold font cells highlight the main variables of interest. They correspond to the estimated coefficients for merit goods (*m*) and public goods (*g*).

Table 2.5 Replication of Column 2, Table 5 augmented by disposable income, F&K

Variables	F&K	Replication (1970-1996)	Replication (1996-2014)
$\Delta c_{i,t-1}$	0.530***	0.908***	0.743**
	(0.142)	(0.083)	(0.289)
$\Delta m_{i,t}$	**0.534*** **	**0.471*** **	**0.357** **
	(0.115)	**(0.082)**	**(0.141)**
$\Delta m_{i,t-1}$	-0.239*	-0.416***	-0.248***
	(0.146)	(0.092)	(0.052)
$\Delta g_{i,t}$	**-0.123**	**0.082**	**0.024**
	(0.095)	**(0.129)**	**(0.143)**
$\Delta g_{i,t-1}$	0.057	-0.090	-0.084
	(0.066)	(0.130)	(0.072)
$\Delta \mathrm{yd}_{\mathrm{i,t}}$	0.453**	0.562***	0.413**
	(0.210)	(0.164)	(0.207)
$\Delta \mathrm{yd}_{\mathrm{i,t-1}}$	-0.340	-0.106	-0.022
	(0.208)	(0.125)	(0.052)
$\Delta d_{i,t}$	-0.239	0.425	-0.465
	(0.156)	(0.449)	(0.421)
$\Delta r_{i,t}$	0.054	-0.006	-0.013**
	(0.292)	(0.010)	(0.006)
J-test	$X^2(11) = 11.3$	$X^2(11) = 3.84$	$X^2(11) = 3.17$
	P =0.418	P =0.974	P =0.988
Observations	297	225	185

Table 2.5 reports the results from estimating equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels. The bold font cells highlight the main variables of interest. They correspond to the estimated coefficients for merit goods (*m*) and public goods (*g*).

Table 2.6 Replication of Column 3, Table 5 augmented by disposable income, F&K

Variables	F&K	Replication (1970-1996)	Replication (1996-2014)
$\Delta c_{i,t-1}$	0.633***	0.945***	0.688***
	(0.193)	(0.143)	(0.248)
$\Delta m2_{i,t}$	**0.554*** **	**0.388*** **	**0.386*** **
	(0.105)	**(0.121)**	**(0.143)**
$\Delta m2_{i,t-1}$	-0.311***	-0.350***	-0.265***
	(0.141)	(0.087)	(0.077)
$\Delta g_{i,t}$	**0.012**	**0.088**	**-0.010**
	(0.115)	**(0.121)**	**(0.158)**
$\Delta g_{i,t-1}$	0.063	-0.105	-0.005
	(0.068)	(0.136)	(0.036)
$\Delta \mathrm{yd}_{\mathrm{i,t}}$	0.309	0.383***	0.389**
	(0.211)	(0.141)	(0.154)
$\Delta \mathrm{yd}_{\mathrm{i,t-1}}$	-0.403*	0.008	0.005
	(0.222)	(0.138)	(0.031)
$\Delta d_{i,t}$	-0.179*	0.497	-0.422
	(0.124)	(0.399)	(0.417)
$\Delta r_{i,t}$	0.374	0.001	-0.010*
	(0.284)	(0.009)	(0.006)
J-test	$X^2(13) = 9.4$	$X^2(13) = 2.54$	$X^2(13) = 3.11$
	P =0.581	P =0.999	P =0.997
Observations	297	232	196

Table 2.6 reports the results from estimating equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels. The bold font cells highlight the main variables of interest. They correspond to the estimated coefficients for merit goods (*m*) and public goods (*g*).

Table 2.7 Replication of Column 6, Table 5 augmented by disposable income, F&K

Variables	F&K	Replication (1970-1996)	Replication (1996-2014)
$\Delta c_{i,t-1}$	0.513***	0.961***	0.719***
	(0.134)	(0.077)	(0.205)
$\Delta m_{i,t}$	**0.661*** **	**0.477*** **	**0.373*** **
	(0.154)	**(0.105)**	**(0.106)**
$\Delta m_{i,t-1}$	-0.365***	-0.450***	-0.255**
	(0.130)	(0.111)	(0.122)
$\Delta g2_{i,t}$	**-0.029**	**0.040**	**0.019**
	(0.078)	**(0.095)**	**(0.150)**
$\Delta g2_{i,t-1}$	0.101	-0.064	-0.057
	(0.068)	(0.086)	(0.087)
$\Delta \mathrm{yd}_{\mathrm{i,t}}$	0.220	0.398**	0.398***
	(0.270)	(0.169)	(0.130)
$\Delta \mathrm{yd}_{\mathrm{i,t-1}}$	-0.252	-0.019	0.008
	(0.226)	(0.155)	(0.057)
$\Delta d_{i,t}$	-0.215*	0.633	-0.270
	(0.128)	(0.487)	(0.543)
$\Delta r_{i,t}$	0.199	-0.002	-0.010*
	(0.300)	(0.006)	(0.006)
J-test	$X^2(13) = 9.7$	$X^2(13) = 2.27$	$X^2(13) = 1.43$
	P =0.553	P =1	P =1
Observations	297	225	185

Table 2.7 reports the results from estimating equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels. The bold font cells highlight the main variables of interest. They correspond to the estimated coefficients for merit goods (*m*) and public goods (*g*).

Table 2.8: Original results of F&K in their Table 5

Eq.	1	2	3	4	5	6	7
$\Delta c_{i,t-1}$	0.675*** (0.175)	0.773*** (0.11)	0.464*** (0.127)	0.412*** (0.135)	0.544*** (0.139)	0.543*** (0.144)	0.498*** (0.11)
$\Delta m_{i,t}$	0.353*** (0.123)	0.498*** (0.063)			0.58*** (0.081)	0.61*** (0.082)	
$\Delta m_{i,t-1}$	-0.103 (0.084)	-0.277*** (0.048)			-0.336*** (0.082)	-0.359*** (0.103)	
$\Delta \mathrm{m2}_{\mathrm{i,t}}$			0.504*** (0.081)	0.704*** (0.136)			0.537*** (0.093)
$\Delta \mathrm{m2}_{\mathrm{i,t-1}}$			-0.01 (0.066)	-0.377*** (0.129)			-0.257*** (0.094)
$\Delta g_{i,t}$	-0.102 (0.08)	-0.067 (0.063)	-0.062 (0.088)				
$\Delta g_{i,t-1}$	-0.078 (0.065)	-0.093** (0.038)	-0.248*** (0.062)				
$\Delta g2_{i,t}$				0.028 (0.098)		0.011 (0.073)	
$\Delta g2_{i,t-1}$				0.06 (0.065)		0.055 (0.047)	
$\Delta g3_{i,t}$					0.059 (0.11)		0.06 (0.092)
$\Delta g3_{i,t-1}$					0.064 (0.057)		0.02 (0.051)

Eq.	1	2	3	4	5	6	7
$\Delta \boldsymbol{d}_{\boldsymbol{i},\boldsymbol{t}}$		-0.375*** (0.134)	-0.218* (0.124)	-0.181 (0.137)	-0.148 (0.099)	-0.203* (0.115)	-0.225** (0.102)
$\Delta \boldsymbol{r}_{\boldsymbol{i},\boldsymbol{t}}$	0.284 (0.234)	0.082 (0.193)	0.139 (0.173)	0.804*** (0.221)	0.389* (0.233)	0.428* (0.229)	0.613*** (0.19)
J-test	X^2 (11)=13.6	X^2 (15)=15.3	X^2 (13)=11.4	X^2(10)=9.2	X^2 (13)=11.5	X^2 (13)=11.7	X^2(13)=13.5
	P=0.253	P =0.426	P =0.575	P =0.513	P=0.57	P =0.548	P =0.413
Observations	297	297	297	297	297	297	297

Table 2.8 reports the results from estimating Equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 are two-sided significance levels.

Table 2.9: Replication (1970-1996) of F&K's Table 5

Eq.	1	2	3	4	5	6	7
$\Delta c_{i,t-1}$	0.990*** (0.273)	0.827*** (0.129)	0.693*** (0.171)	0.809*** (0.155)	0.926*** (0.141)	0.870*** (0.112)	0.847*** (0.183)
$\Delta m_{i,t}$	0.138 (0.285)	0.535* (0.306)			0.536** (0.270)	0.488** (0.246)	
$\Delta m_{i,t-1}$	-0.122 (0.299)	-0.437* (0.256)			-0.476** (0.234)	-0.411** (0.193)	
$\Delta m2_{i,t}$			0.426 (0.286)	0.392* (0.203)			0.502* (0.259)
$\Delta m2_{i,t-1}$			-0.253 (0.231)	-0.280* (0.148)			-0.397* (0.228)
$\Delta g_{i,t}$	0.088 (0.268)	0.055 (0.199)	0.129 (0.249)				
$\Delta g_{i,t-1}$	-0.074 (0.237)	-0.049 (0.171)	-0.106 (0.233)				
$\Delta g2_{i,t}$				0.122 (0.105)		0.042 (0.123)	
$\Delta g2_{i,t-1}$				-0.112 (0.101)		-0.037 (0.105)	
$\Delta g3_{i,t}$					-0.018 (0.150)		0.017 (0.153)
$\Delta g3_{i,t-1}$					0.001 (0.113)		-0.029 (0.115)

Eq.	1	2	3	4	5	6	7
$\Delta d_{i,t}$		0.336 (0.316)	-0.112 (0.317)	0.012 (0.318)	0.465 (0.576)	0.280 (0.388)	0.167 (0.544)
$\Delta r_{i,t}$	0.021** (0.009)	0.011 (0.012)	0.020 (0.012)	0.015 (0.014)	0.013 (0.010)	0.012 (0.010)	0.017 (0.014)
J-test	X^2 (11)=5.48	X^2(15)=7.14	X^2 (13)=6.06	X^2 (10)=4.42	X^2(13)=2.46	X^2(13)=6.55	X^2(13)=6.35
	P=0.906	P=0.954	P=0.944	P=0.926	P=0.999	P=0.924	P=0.933
Observations	232	232	232	232	232	232	232

Table 2.9 reports the results from estimating Equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 are two-sided significance levels.

Table 2.10: Replication (1996-2014) of F&K's Table 5

Eq.	1	2	3	4	5	6	7
$\Delta c_{i,t-1}$	1.038*** (0.238)	0.843*** (0.154)	1.001*** (0.257)	0.574** (0.279)	1.140*** (0.243)	0.880*** (0.190)	1.057*** (0.251)
$\Delta m_{i,t}$	0.523*** (0.182)	0.517*** (0.090)			0.061 (0.273)	0.448*** (0.170)	
$\Delta m_{i,t-1}$	-0.580*** (0.084)	-0.430*** (0.117)			-0.126 (0.261)	-0.372* (0.194)	
$\Delta m2_{i,t}$			0.442*** (0.140)	0.623** (0.291)			0.162 (0.256)
$\Delta m2_{i,t-1}$			-0.407*** (0.064)	-0.541* (0.311)			-0.192 (0.208)
$\Delta g_{i,t}$	-0.047 (0.103)	-0.106 (0.097)	-0.085 (0.093)				
$\Delta g_{i,t-1}$	0.040 (0.100)	-0.011 (0.075)	-0.104 (0.121)				
$\Delta g2_{i,t}$				0.035 (0.119)		-0.090 (0.115)	
$\Delta g2_{i,t-1}$				0.133 (0.314)		-0.073 (0.164)	
$\Delta g3_{i,t}$					-0.100 (0.076)		-0.103 (0.078)
$\Delta g3_{i,t-1}$					-0.254** (0.113)		-0.178 (0.150)

Eq.	1	2	3	4	5	6	7
$\Delta \boldsymbol{d}_{\boldsymbol{i,t}}$		-0.420 (0.410)	-0.551 (0.435)	-0.782 (0.754)	-0.396 (0.348)	-0.137 (0.476)	-0.330 (0.249)
$\Delta \boldsymbol{r}_{\boldsymbol{i,t}}$	-0.002 (0.004)	-0.005 (0.006)	-0.007 (0.007)	-0.009* (0.005)	0.003 (0.007)	-0.006 (0.006)	0.002 (0.006)
J-test	X^2 (11)=7.93	X^2 (15)=8.55	X^2 (13)=2.51	X^2 (10)=2.33	X^2 (13)=4.86	X^2 (13)=3.98	X^2 (13)=5.65
	P=0.72	P =0.900	P =0.999	P =0.993	P=0.978	P =0.991	P =0.958
Observations	196	196	196	196	196	196	196

Table 2.10 reports the results from estimating Equation (6). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 are two-sided significance levels.

Table 2.11: Original results of F&K in their Table 5, augmented by disposable income

Eq.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
$\Delta \boldsymbol{c}_{i,t-1}$	0.776*** (0.242)	0.530*** (0.142)	0.633*** (0.193)	0.477*** (0.123)	0.493*** (0.161)	0.513*** (0.134)	0.446*** (0.118)
$\Delta \boldsymbol{m}_{i,t}$	0.507** (0.226)	0.534*** (0.115)			0.649*** (0.129)	0.661*** (0.154)	
$\Delta \boldsymbol{m}_{i,t-1}$	-0.430** (0.215)	-0.239* (0.146)			-0.352*** (0.113)	-0.365*** (0.130)	
$\Delta \mathrm{m2}_{\mathrm{i,t}}$			0.554*** (0.105)	0.565*** (0.129)			0.556*** (0.121)
$\Delta \mathrm{m2}_{\mathrm{i,t-1}}$			-0.311* (0.141)	-0.269** (0.116)			-0.249** (0.100)
$\Delta \boldsymbol{g}_{i,t}$	-0.041 (0.223)	-0.123 (0.095)	0.012 (0.115)				
$\Delta \boldsymbol{g}_{i,t-1}$	-0.138 (0.009)	0.057 (0.066)	0.063 (0.068)				
$\Delta \boldsymbol{g2}_{i,t}$				0.001 (0.007)		-0.029 (0.078)	
$\Delta \boldsymbol{g2}_{i,t-1}$				0.083 (0.056)		0.101 (0.068)	
$\Delta \boldsymbol{g3}_{i,t}$					0.010 (0.081)		-0.004 (0.063)
$\Delta \boldsymbol{g3}_{i,t-1}$					0.096 (0.068)		0.086* (0.050)

Eq.	1	2	3	4	5	6	7
$\Delta yd_{i,t}$	0.114 (0.265)	0.453** (0.210)	0.309 (0.211)	0.379* (0.202)	0.166 (0.255)	0.220 (0.270)	0.392** (0.193)
$\Delta yd_{i,t-1}$	-0.207 (0.280)	-0.340 (0.208)	-0.403* (0.222)	-0.344* (0.199)	-0.191 (0.189)	-0.252 (0.226)	-0.333* (0.175)
$\Delta d_{i,t}$		-0.239 (0.156)	-0.179* (0.124)	-0.243** (0.121)	-0.165 (0.119)	-0.215* (0.128)	-0.251** (0.121)
$\Delta r_{i,t}$	0.588 (0.538)	0.054 (0.292)	0.374 (0.284)	0.270 (0.268)	0.111 (0.315)	0.199 (0.300)	0.108 (0.268)
J-test	X^2 (9)=9.7	X^2 (11)=11.3	X^2 (13)=9.4	X^2 (11)=9.2	X^2 (13)=9.2	X^2 (13)=9.7	X^2 (13)=8.4
	P=0.377	P =0.418	P =0.581	P =0.606	P=0.605	P =0.553	P =0.678
Observations	297	297	297	297	297	297	297

Table 2.11 reports the results from estimating Equation (7). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 are two-sided significance levels.

Table 2.12: Replication (1970-1996) results of F&K's Table 5, augmented by disposable income

Eq.	1	2	3	4	5	6	7
$\Delta c_{i,t-1}$	0.999*** (0.199)	0.908*** (0.083)	0.945*** (0.143)	0.924*** (0.095)	0.964*** (0.083)	0.961*** (0.077)	0.934*** (0.125)
$\Delta m_{i,t}$	0.425* (0.230)	0.471*** (0.082)			0.528*** (0.118)	0.477*** (0.105)	
$\Delta m_{i,t-1}$	-0.395 (0.277)	-0.416*** (0.092)			-0.498*** (0.121)	-0.450*** (0.111)	
$\Delta m2_{i,t}$			0.388*** (0.121)	0.469*** (0.085)			0.451*** (0.083)
$\Delta m2_{i,t-1}$			-0.350*** (0.087)	-0.418*** (0.096)			-0.415*** (0.105)
$\Delta g_{i,t}$	0.259 (0.183)	0.082 (0.129)	0.088 (0.121)				
$\Delta g_{i,t-1}$	-0.257 (0.178)	-0.090 (0.130)	-0.105 (0.136)				
$\Delta g2_{i,t}$				0.033 (0.094)		0.040 (0.095)	
$\Delta g2_{i,t-1}$				-0.052 (0.091)		-0.064 (0.086)	
$\Delta g3_{i,t}$					0.058 (0.073)		0.088 (0.085)
$\Delta g3_{i,t-1}$					-0.085 (0.064)		-0.097 (0.080)

Eq.	1	2	3	4	5	6	7
$\Delta yd_{i,t}$	0.281** (0.135)	0.562*** (0.164)	0.383*** (0.141)	0.625*** (0.191)	0.475*** (0.167)	0.398** (0.169)	0.497*** (0.153)
$\Delta yd_{i,t-1}$	-0.011 (0.142)	-0.106 (0.125)	0.008 (0.138)	-0.090 (0.172)	-0.063 (0.142)	-0.019 (0.155)	-0.050 (0.154)
$\Delta d_{i,t}$		0.425 (0.449)	0.497 (0.399)	0.372 (0.442)	0.710 (0.567)	0.633 (0.487)	0.501 (0.635)
$\Delta r_{i,t}$	0.010 (0.012)	-0.006 (0.010)	0.001 (0.009)	-0.007 (0.010)	-0.001 (0.008)	-0.002 (0.006)	0.002 (0.010)
J-test	X^2 (9)=2.25	X^2 (11)=3.84	X^2 (13)=2.54	X^2 (11)= 4	X^2 (13)=3.14	X^2 (13)=2.27	X^2 (13)=1.41
	P=0.987	P=0.974	P=0.999	P=0.970	P=0.997	P=1	P=1
Observations	225	225	225	225	225	225	225

Table 2.12 reports the results from estimating Equation (7). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels.

Table 2.13: Replication (1996-2014) results of F&K's Table 5, augmented by disposable income

Eq.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
$\Delta \boldsymbol{c}_{i,t-1}$	0.831*** (0.264)	0.743** (0.289)	0.688*** (0.248)	0.810*** (0.232)	0.738*** (0.197)	0.719*** (0.205)	0.797*** (0.194)
$\Delta \boldsymbol{m}_{i,t}$	0.327** (0.150)	0.357** (0.141)			0.256** (0.119)	0.373*** (0.106)	
$\Delta \boldsymbol{m}_{i,t-1}$	-0.264** (0.109)	-0.248*** (0.052)			-0.153 (0.125)	-0.255** (0.122)	
$\Delta \mathrm{m2}_{\mathrm{i,t}}$			0.386*** (0.143)	0.229 (0.143)			0.220* (0.127)
$\Delta \mathrm{m2}_{\mathrm{i,t-1}}$			-0.265*** (0.077)	-0.110 (0.187)			-0.135 (0.122)
$\Delta \boldsymbol{g}_{i,t}$	0.008 (0.124)	0.024 (0.143)	-0.010 (0.158)				
$\Delta \boldsymbol{g}_{i,t-1}$	-0.022 (0.085)	-0.084 (0.072)	-0.005 (0.036)				
$\Delta \boldsymbol{g2}_{i,t}$				0.009 (0.171)		0.019 (0.150)	
$\Delta \boldsymbol{g2}_{i,t-1}$				-0.176 (0.125)		-0.057 (0.087)	
$\Delta \boldsymbol{g3}_{i,t}$					0.011 (0.083)		-0.036 (0.092)
$\Delta \boldsymbol{g3}_{i,t-1}$					-0.094 (0.080)		-0.106 (0.070)

Eq.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
$\Delta \boldsymbol{yd}_{i,t}$	0.486*** (0.155)	0.413** (0.207)	0.389** (0.154)	0.395** (0.183)	0.376*** (0.138)	0.398*** (0.130)	0.375** (0.148)
$\Delta \boldsymbol{yd}_{i,t-1}$	0.005 (0.062)	-0.022 (0.052)	0.005 (0.031)	-0.026 (0.057)	-0.002 (0.042)	0.008 (0.057)	-0.032 (0.038)
$\Delta \boldsymbol{d}_{i,t}$		-0.465 (0.421)	-0.422 (0.417)	-0.212 (0.545)	-0.287 (0.304)	-0.270 (0.543)	-0.417 (0.327)
$\Delta \boldsymbol{r}_{i,t}$	-0.009** (0.004)	-0.013** (0.006)	-0.010* (0.006)	-0.010 (0.006)	-0.006 (0.006)	-0.010* (0.006)	-0.005 (0.006)
J-test	X^2 (9)=1.61	X^2 (11)=3.17	X^2 (13)=3.11	X^2 (11)=0.97	X^2 (13)=1.52	X^2 (13)=1.43	X^2 (13)=1.52
	P=0.996	P =0.988	P =0.997	P =1	P=1	P =1	P =1
Observations	185	185	185	185	185	185	185

Table 2.13 reports the results from estimating Equation (7). Newey-West standard errors are in parentheses. Apart from working age population share, all variables are logged and first-differenced. ***=0.01, **=0.05. *=0.10 denote two-sided significance levels.

Table 2.14: The structure of general government expenditure in 2015 (GDP% shares)
Update on F&K's Table 4.

	Denmark	France	Germany	Italy	Norway	Portugal	Spain	United Kingdom
Total	54.5	56.8	44.1	50.3	49.3	48.2	43.9	42.3
General Public Services	7.4	6.3	5.9	8.6	4.7	8.8	6.5	4.5
Defence	1.1	1.8	1.0	1.2	1.5	1.0	1.0	2.0
Public Order and Safety	1.0	1.6	1.5	1.9	1.1	1.9	2.0	1.9
Education	7.0	5.4	4.2	4.0	5.5	5.1	4.1	5.2
Health	8.5	8.1	7.2	7.0	8.2	6.2	6.2	7.5
Housing and Community Amenities	0.2	1.1	0.4	0.5	0.8	0.5	0.5	0.7
Recreation, Culture and Religion	1.8	1.4	1.0	0.7	1.6	0.8	1.2	0.7
Social Protection	23.5	24.3	19.1	21.3	19.5	18.5	17.2	16.0
Economic Affairs	3.6	5.7	3.2	4.3	5.4	4.9	4.4	3.1
Environment Protection	0.4	1.0	0.6	0.9	0.9	0.6	0.9	0.8

Source: OECD National Accounts Database

2.9 Reference

Abel, A. B. (1990). Asset prices under habit formation and catching up with the Joneses. *The American Economic Review, 80*(2), 38-42.

Afonso, A., & Sousa, R. M. (2011). The macroeconomic effects of fiscal policy in Portugal: a Bayesian SVAR analysis. *Portuguese Economic Journal*, *10*(1), 61-82.

Arratibel, O., Furceri, D., Martin, R., & Zdzienicka, A. (2011). The effect of nominal exchange rate volatility on real macroeconomic performance in the CEE countries. *Economic Systems*, *35*(2), 261-277.

Baxter, M., & King, R. G. (1993). Fiscal policy in general equilibrium. *The American Economic Review*, 315-334.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics, 117*(4), 1329-1368.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy, 107*(2), 205-251.

Christiano, L. J., & Eichenbaum, M. (1995). Liquidity effects, monetary policy, and the business cycle. *Journal of Money, Credit and Banking, 27*(4), 1113-1136.

Coenen, G., & Straub, R. (2005). Does government spending crowd in private consumption? Theory and empirical evidence for the euro area. *International Finance*, *8*(3), 435-470.

Devereux, M. B., Head, A. C., & Lapham, B. J. (1996). Monopolistic competition, increasing returns, and the effects of government spending. *Journal of Money, credit and Banking*, *28*(2), 233-254.

Easterly, W., & Rebelo, S. (1994). *Fiscal policy and economic growth: an empirical investigation*. *CEPR Discussion Papers. No. 855.*

Fatás, A., & Mihov, I. (2001). The effects of fiscal policy on consumption and employment: theory and evidence. *CEPR Discussion Papers No. 2760.*

Fiorito, R., & Kollintzas, T. (2004). Public goods, merit goods, and the relation between private and government consumption. *European Economic Review*, *48*(6), 1367-1398.

Furceri, D., & Sousa, R. M. (2011). Does government spending crowd out private consumption and investment? *World Economics*, *12*(4), 153-170.

Furceri, D., & Zdzienicka, A. (2011). The real effect of financial crises in the European transition economies 1. *Economics of transition, 19*(1), 1-25.

Galí, J., López-Salido, J. D., & Vallés, J. (2007). Understanding the effects of government spending on consumption. *Journal of the european economic association*, *5*(1), 227-270.

Giordano, R., Momigliano, S., Neri, S., & Perotti, R. (2007). The effects of fiscal policy in Italy: Evidence from a VAR model. *European Journal of Political Economy, 23*(3), 707-733.

Ho, T. W. (2001). The government spending and private consumption: a panel cointegration analysis. *International Review of Economics & Finance*, *10*(1), 95-108.

Lettau, M., & Ludvigson, S. C. (2004). Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption. *American Economic Review*, *94*(1), 276-299.

Linnemann, L., & Schabert, A. (2004). Can fiscal spending stimulate private consumption?. *Economics Letters*, *82*(2), 173-179.

Loayza, N., Schmidt-Hebbel, K., & Servén, L. (2000). What drives private saving across the world?. *Review of Economics and Statistics*, *82*(2), 165-181.

Ludvigson, S. C., & Steindel, C. (1998). *How important is the stock market effect on consumption?*. *FRBNY Economic Policy Review 5*(2), 29–51.

Mountford, A., & Uhlig, H. (2009). What are the effects of fiscal policy shocks?. *Journal of Applied Econometrics*, *24*(6), 960-992.

Payne, A. A. (2009). 5 Does Government Funding Change Behavior? An Empirical Analysis of Crowd-Out. *Tax Policy and the Economy*, *23*(1), 159-184.

Ramey, V. A., & Shapiro, M. D. (1998, June). Costly capital reallocation and the effects of government spending. *In Carnegie-Rochester Conference Series on Public Policy* (Vol. 48, pp. 145-194).

Chapter 3: The One-Child Policy and Chinese saving behavior

3.1 Introduction

Since the 1970s, the Chinese household savings rate, known as household savings divided by household disposable income, has dramatically increased. It currently stands at around 36%, which is about six times the OECD norm (OECD, 2019). This figure peaked at around 40% in 2010. Indeed, China's gross savings rate is among the World Bank's highest of 170 countries (World Bank, 2019). China experienced rapid economic growth in the past decades. The GDP growth rate has been fluctuating between 7% to 14% ever since the 1990s. Despite such rapid growth, economic intuition suggests that Chinese people should have little incentive to save but has a strong incentive to spend more money today. Since the main aim of saving is to smooth out future consumption. In other words, because China's households are expected to be wealthier in the future, they should improve their welfare by consuming more today. Despite this bright outlook, Chinese households are still saving up a lot of their income, thus it's curious why China's savings rate is so high.

It is important to understand why the saving rate in China is so high, especially today because many believe saving was a driving force behind China's economic growth through its impact on investment. The connection is easy to make. As in 2017, investment contributed 44.41 percent of China's GDP, compared to the world averages of 23 percentage points (CEIC, 2017). However, recently attention has focused on "rebalancing" growth from investment to consumption. There is a sense that for China to sustain its growth, it needs to transition from a high saving/low consumption economy into an economy that consumes more and saves less.

In recent years, China's economic growth has transitioned from a medium to a high speed, compared with several double-digit growth rates in the past. This is mainly due to the slowing pace of export growth as well as low consumption. Accordingly, understanding the reason behind China's high savings rate and stimulating consumption has been the subject of much recent research (Li, Whally & Zhao, 2013, 2015; Wang & Wen, 2012; Feng, He & Sato, 2011; Chu & Wen, 2017; Wei & Zhang, 2011; Chamon & Prasad, 2010; Chamon, Liu & Prasad 2013).

In this chapter, I focused on the role of China's “One-Child Policy” (OCP), which was introduced in 1979. This policy is one of the explanations that received significant attention to understanding the high Chinese savings rate. This policy limited the number of children that parents can have. China’s average birth rate has declined from roughly 3.8 to 1.6 (World Bank 2019) since the OCP was adopted. (Birth rate represents the average number of children born to women in their lives.)

The “One-Child Policy” was implemented in the 1970s, which is also the period in which Chinese households experienced a huge increase in savings. Given the concurrence in the timing of the implementation of the OCP and the increase in China's saving rate, it is tempting to assume a causal link between OCP and savings rate. Multiple studies in this field have confirmed and supported this link. Specifically, Zhou (2014); Ge, Yang & Zhang (2018); Lugauer, Ni & Yin (2017); Curtis, Lugauer, & Mark (2015); and Choukhmane, Coeurdacier & Jin (2014) all conclude that the OCP contributed partly, or even a huge amount of increase in the household sector savings rate.

The studies that focused on whether or not the OCP is a contributing factor to the increase in savings rate draw their conclusion by estimating a relationship between the number of children and savings. Their “savings function” is estimated using data during the post OCP period. By doing so, they are making the assumption that this “savings function” that they are estimating is unaffected by the OCP, in other words, the OCP doesn’t affect how each child affect people’s savings, it only has an effect on the number of children that a household have. Therefore, their counterfactual for Chinese households without the OCP is the saving behavior of Chinese families with more than one child under the OCP, while ideally, the comparison would be comparing Chinese families under the OCP with Chinese families, not under the OCP. Thus, the existing literature are all working with an imperfect counterfactual.

In this chapter of my thesis, my study differs from these previous studies in two ways. Firstly, I proposed a different set of counterfactuals compared to the existing literature, I used the saving behavior of households from other Asian regions without the OCP that has a similar culture and background to China, as my picture for Chinese households without the “One-Child Policy.” Further, in my estimation method, I allow the impact of the OCP to affect both the number of children (the "endowment effect") and the “savings function” itself, the relationship between the number of children and savings (the "coefficient effect") to coexist. I employed Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) to disentangle these effects.

The data used in this chapter of my thesis is the 2014 Gallup World Poll and the 2014 Global Findex database. By merging these two datasets together, the comparison of the saving behavior of Chinese people and the saving behavior of people in other regions without the OCP can be conducted.

The results of my empirical analysis showed that, first of all, there’s a limited difference between the saving behavior and the demographics between China and the counterfactuals used in my sample. When I adopted Blinder-Oaxaca decomposition, the results suggested that the OCP can explain a sizeable amount of the difference, but the numbers are not big throughout the specifications of my model. Overall, I interpret my results as showing little evidence to suggest that the OCP had a big influence on people’s decision to save in China.

The remainder of this chapter is structured as follows: Section 3.2 provides contextual information and a literature review on the topic. Section 3.3 discusses and briefly describes the data. Section 3.4 presents my methodology. Section 3.5 discusses my empirical results and Section 3.6 concludes.

3.2 Background

Ever since the 1970s, savings rates in China began rising at an astonishing speed, reaching levels that are the highest among other economies worldwide. They have remained at this level for the past decades. The abnormal saving behaviour of China has attracted the attention of many scholars. What makes the saving pattern so difficult to understand for China mainly comes from two perspectives. Firstly, the U-shaped age-saving profile of the Chinese households, and secondly, the exceptionally high saving even after a country has achieved sustained high growth for many years.

The first interesting aspect of Chinese savings is the U-shaped age-saving profile. According to the life cycle hypothesis (Modigliani & Brumberg, 1954), the relationship between an economy's savings rates and its age profile can be represented by a hump-shaped curve. Savings rates for the young and old members of society should be relatively low because they don't have sufficient excess income to produce savings. In contrast, the savings rates for the working-age population should be relatively high. Interestingly, research conducted on the Chinese economy by Chamon, Liu & Prasad (2010), Choukhmane et al. (2013), and Rosenzweig & Zhang (2014) all found the opposite. Younger and older members of the Chinese economy saved more compared to the majority of the working-age population.

FIGURE 3.1 shows the age saving profile of China between four different cohorts, 1990, 1995, 2000, and 2005. At the beginning of the 1990s, China had a relatively similar age-saving profile as the other economies, with saving rates increasing for the working-age population, then declining after retirement. As the savings continue to increase in the 1990s, the increase is particularly prominent for younger households (in their 20s and early 30s) and older households (in their 50s and older). (Chamon, Liu & Prasad, 2010)

This phenomenon is called the Chinese age-saving puzzle. In order to explain this puzzle, many papers have focused on the determinants of household savings. Rosenzweig & Zhang (2014) tried to explain China's U-shaped saving-age profile using data on twins and family children. In particular, they were interested in the impact of the number of children on young and middle-aged families. Choukhmane et al. (2013) studied the influence of the one-child policy and the impact of this demographic change on household savings rates. They found that households with more boys have a higher savings rate. They also hypothesized that the reason older people now save relatively more is related to the fact that children traditionally have served as old-age support for Chinese families. As the number of children has decreased as a result of the one-child policy, many older people feel the need to save to compensate for the loss in old age support from having fewer children.

Another explanation for China's U-shaped age-saving profile relates to the unreliability of China's pension system. This forces older people to save more for themselves.

An alternative explanation related to housing prices is provided by Chamon, Liu & Prasad (2010). They found that young people in China apply for mortgages at a relatively young age. In order to obtain the necessary collateral, they forgo consumption and increase their savings. Yet another explanation is provided by Song and Yang (2010). They conclude that the higher savings rate of younger people is due to the fact that the younger generation of the Chinese economy actually earns more than the middle-aged generation. This allows them to be able to save more.

Another interesting aspect of the Chinese saving puzzle is the huge amount that Chinese households are saving. FIGURE 3.2 shows that since the 1970s, China's saving rate has been increasing. A major increase took place in the six years between 1978 and 1984, during which the saving rate increased by 15%. Many economic and policy reforms happened after the 1960s, which unleashed the possibility of productivity growth for the country previously under the control of central planning by the government. Some of these reforms include the open-door policy, which opened the country to foreign direct investment. There have also been huge rounds of reforms on the agricultural front. Before the reform, households were divided into collective production units and worked as a team, which made it difficult to incentivize better performance of each household as individual performance is difficult to monitor. Under Deng Xiaoping's agricultural reform, farmable lands were privatized to allow farmers to be responsible for their own production. These means of reforms significantly boosted China's productivity and income, which in turn made people wealthier. Other reforms happened in the 1990s, which changed the original pension system, mostly paid by the government, into a 2 tiered pension system that relied on both employee contribution and government. These reforms may all have contributed to the increase in China's savings rate in the past decades. Indeed, to understand this increase, multiple hypotheses have been put forward by scholars.

A branch of economic research trying to explain the high savings rate in China focused on rising property prices. Housing prices in China have increased dramatically since China's housing reform in 1998. In 1998, China's average house price per meter was about $300. In 2009, it had more than doubled or nearly tripled. Some analysts started to question whether rising home prices could explain the high savings rates in China. Wang & Wen (2012) used simulations with Chinese time-series data on household income, housing prices, and demographics, their analysis found that rising mortgage costs increased the aggregate saving rate by at most 2 to 4 percentage points. Li, Whally, & Zhao (2013) also found no evidence linking housing prices to household saving behavior based on data from the Chinese household income project[1]. Finally, Chamon & Prasad (2010) estimated that saving behavior related to homeownership could account for only 3 percentage points of the increase in China's saving rate.

Other studies focused on pension reform as a potential reason behind this high savings rate. Chamon, Liu, & Prasad (2013) calibrate a model of savings and conclude that a significant increase in household savings can be explained by the combination of rising income uncertainty and pension reforms. They estimate that, together, these account for two-thirds of the increase in China's urban household saving rate. In contrast, Feng, He & Sato (2011) estimated a smaller effect. They concluded that pension reform in China increased household savings by 6-9 percentage points for citizens aged 25-29, and 3 percentage points for people aged 50-59.

Another strand of research investigated how sex imbalances affect the saving rate. Wei & Zhang (2011), for example, point to the competitive saving motive as a possible explanation for high Chinese saving rates. According to this motive, Chinese parents with a son increase their savings to make their child a relatively more attractive marriage partner. It follows that saving rates will increase as the sex ratio (the ratio of male to female) increases. Wei & Zhang conclude that this factor can

[1], measures and estimates the distribution of personal income and related economic factors in both rural and urban areas of China, there are four rounds of this survey, 1988, 1995, 2002 and 2007. More information on the survey can be found on http://www.ciidbnu.org/.

account for more than half of the increase in the household saving rate between 1990 and 2007.

The mixed results from these earlier studies provide ample room for alternative explanations to help us understand the high saving rate in China. The One-Child Policy (OCP) is one justification that has received a lot of attention recently. The OCP was introduced in 1979, and later changed in the 1980s to allow a second child for minority and rural parents if their first child was a daughter. The policy was terminated in 2015 and has been credited with preventing approximately 400 million births. (Whyte, Feng & Cai (2015).

An often-made argument linking fewer children to greater savings relates to retirement. Children are commonly viewed as a source of old-age support. Parents have children when they are young, in the expectation that the children will make financial transfers to them when they are old. In this sense, children and savings are viewed as financial substitutes. Parents with more children can afford to save less for retirement. Parents with fewer children need to save more to ensure adequate retirement support. Thus, faced with a restriction on the number of children they are allowed to have, parents may choose to save more (Choukhmane, Coeurdacier & Jin, 2014).

Of course, such an outcome is not certain. Rather than saving more, with fewer children, families could invest more in the one child they are allowed to have to make sure that the only child is more able to support them in the future. As proposed by Becker(1960), Becker & Lewis(1973), Becker & Tomas(1976), there may be a trade-off between the size of the family and population, and the accumulation of human capital in an economy. The problem is often referred to as the quantity-quality trade-off(Q-Q). The Q-Q trade-off theory indicates that families' decision to invest in human capital of each child is dependent on the number of children families decide to have. If families are under budget constraints, a reduction in the number of children a family has implies there will be more resources allocated to each individual child, while an increase in the number of children in a family will decrease the resources allocated to each individual child. In other words, the reduction of family size would increase human capital accumulation per child.

Most papers confirmed the trade-off between quantity and quality of children using Chinese data, which proved that the OCP has a positive influence on the accumulation of human capital in China. (Rosenzweig & Zhang, 2009; Li, Zhang & Zhu, 2008; Zhu, Whalley & Zhao, 2014; Qin, Zhuang & Yang, 2017;), others found that the OCP did not have an impact on China's human capital, (Wang & Zhang, 2018), or even a positive effect, indicating that additional siblings will increase the education outcome of the first-born. (Qian, 2009), this may be due to the fact that parents treat their children unequally, for example, by giving more attention to their firstborn and setting higher standards for them, or they pay more attention to the education of sons rather than daughters.

Instead of spending more on each child, alternatively, it is also possible for parents to consume the additional resources. If children are consumption rather than investment goods, parents could divert consumption to other outlets rather than save. In other words, the additional wealth that parents get as a result of fewer children could be used to boost consumption instead of savings.

To estimate the overall impact of the OCP on savings, one would ideally compare the observed savings of Chinese households under the OCP to what savings would be like without the OCP. The counterfactual would be an imaginary China identical in every respect to China we currently observe, except that no OCP would have been implemented. Of course, the ideal counterfactual doesn't exist. Hence researchers have had to come up with imperfect counterfactuals.

TABLE 3.1 in the appendix reports what previous studies have used, or implicitly assumed, for their counterfactuals. Since all the studies use post-OCP data to estimate the effect of children on saving, the implicit assumption is that the relationship between children and saving is the same with and without the OCP (thus, no "coefficient effect).

Ge, Yang, & Zhang (2015) link the number of children to household savings rates by different age cohorts, using provincial-level fines for violating the OCP and the share of ethnic minorities as instruments for the number of children. Across all age cohorts, they find that having one more child decreased saving rates. The estimates varied from a decrease of 3.0-4.5 percentage points for an additional child, to 12.7-18.7 percentage points. Applying these results to the OCP, they concluded that an unanticipated consequence of the OCP was to increase the savings rate. They further concluded that a relaxation of the OCP could serve to stimulate household consumption.

Lugauer, Ni, & Yin (2017) use county-level birth rates after the OCP as an instrument for the number of children. They estimate that an additional child is associated with a 5.6 percentage point decrease in household savings. Given a decrease in the number of children from 3 to 1, they calculate that the OCP was responsible for an increase of 11 percentage points in the household saving rate.

Some papers have focused on the gender dimension of children, arguing that sons are more reliable than daughters when it comes to financially supporting their parents. Focusing on the effect of male siblings, Zhou (2014) linked the number of brothers an individual has to that individual's saving rate. The idea is that brothers share the financial responsibility of supporting parents. The more brothers, the less the burden on each son to support his parents, and thus the less demand for him to save. Zhou estimated that an additional brother lowers an individual's saving rate by 5 percentage points. Based on this, he concluded that "population policies" were responsible for over one-third of the increase in the saving rate of urban households.

Banerjee, Meng & Qian (2010) examined data on Chinese households with more than one child. They estimate the difference in the effect of an additional child on saving when the first child is a daughter compared to a son. The idea is that if the first child is a daughter, an additional child reduces the need for self-saving for retirement. If the first child is a son, then retirement saving is already "accounted for," and an additional child will not have much effect on saving for retirement. Their results support the conclusion that the OCP substantially increased savings.

Choukhmane, Coeurdacier, & Jin (2014) develop a quantitative overlapping generation model and calibrate their model using micro-level Chinese data. Assuming that households under the influence of a two-child policy would behave like households with twins under the OCP, they conclude that the OCP can account for 30% to 60% of the rise in aggregate Chinese savings.

Using a similar approach, albeit with macro-level data, Curtis, Lugauer, & Mark (2015) built a simulation model to investigate the effect of the OCP on the Chinese saving rate. They conclude that a one-child reduction raised the saving rate by 5.5 percentage points in 1970 and by 4.2 percentage points in 2009.

The abovementioned studies follow two critical assumptions: (i) the decrease in the number of children in China is for the post-OCP period is because of the OCP, and (ii) How children affect savings in households is unaffected by the OCP. Both of these assumptions deserve examination.

For the first assumption, a list of studies has focused on how OCP has actually affected China's fertility rate. Zhang (2017) compared China's fertility rate with other developing countries that had high rates of fertility in the 1960s (South Korea, India, Thailand, and Mexico.) While he found some differences in fertility rates between China and other countries in the 1970s, these differences diminished over time. While the OCP initially contributed to reduced fertility in the 1970s, it didn't have lasting effects. He concluded that China would have achieved a low fertility rate even without the OCP.

Cai (2010) compared the fertility of China with 200 countries and regions and concluded that the OCP was not the dominant factor responsible for decreasing China's fertility rates. He determined that socioeconomic development and globalization played key roles in China's fertility decline.

Feng, Cai & Gu (2013) compared China's fertility rate to 16 regions that had similar birth rates pre-OCP, including South Korea, Malaysia, Thailand, and Brazil. They developed a counterfactual empirical model that predicted China's fertility rate based on the observed trend in China before the OCP, appended with fertility trends from other regions after the OCP. They concluded that China would have achieved its observed 2010 fertility rate even in the absence of the OCP. A similar conclusion was reached by Whyte, Feng, & Cai (2015) based on a counterfactual analysis using the same set of 16 regions.

Basten & Sobotka (2013) compared the fertility trajectory of a number of low fertility regions, including Singapore, South Korea, Taiwan, Hong Kong, and Japan, and confirmed that the evolution of China's fertility rate was not that different from these other regions.

FIGURE 3.3 presents a time series graph of fertility rates for China and regions that previous studies have identified as appropriate comparisons. Consistent with the previously cited research, China's post-1980 decline in fertility is similar to a number of other countries, indicating that forces beyond the OCP may have been responsible for much of China's decline in fertility rate.

FIGURE 3.4 highlights the concern with inferring causation from OCP to higher savings rate just because the increase in the household savings rate coincided with China's OCP. In this figure, the time series for Taiwan's household savings rate is superimposed on China's time series. Taiwan's increase in household savings behavior was also quite impressive. While Taiwan's rise began a little earlier and started a little higher, the two time series are strikingly similar. In 1970, Taiwan's household savings rate was 8.0%. China's was 2.0%. In 1998, Taiwan's household savings rate was 26.0%. China's was 25.9%. The two trends show many similarities. Yet Taiwan did not have an OCP.[2]

As for the second assumption. The studies identified above all estimate a variation of the following, simplified savings function:

$$savings\ rate = \alpha + \beta \cdot children + controls + \varepsilon \qquad (3.1)$$

[2] Taiwan did implement a series of family population controls in the 1960s. However, its birth rate had been dropping dramatically since at least 1950. (SOURCE: https://www.macrotrends.net/countries/TWN/taiwan/birth-rate).

The coefficient β is key to estimating the impact of the OCP. If one is willing to make an assumption about the impact of the OCP on the number of children, then the effect of the OCP on the saving rate can be estimated as $\beta \cdot \Delta children$.

The crucial assumption made here is that the OCP did not affect the savings function itself. This is a strong assumption: First, the families who currently have more than one child are unlikely to be representative of families who would have had more than one child without the OCP, as those not following the policy or those allowed not to follow the policy are unlikely to be "average" Chinese.

Second, the OCP could have changed how people spend money on children. That is, the amount of money spent on each child may be very different in an OCP environment versus a multiple child policy environment. In a multiple child policy environment, parents can be more confident that at least one of their children will support them when they are old. If forced to have only one child, such 'diversification' is not possible, possibly forcing them to invest more in their one child to ensure that he/she will be financially successful and better able to support them in their old age.

Therefore, I propose a different counterfactual in this thesis. Rather than comparing Chinese families with different numbers of children (all living in an OCP environment), I compare families in China to families in Taiwan, Hong Kong, Singapore, Malaysia, Japan, and South Korea, places that either have a sizeable Chinese population or have a similar culture to China.

I include Taiwan and Hong Kong because these are closest to China in terms of both culture and demographics. Singapore and Malaysia are included because both have sizeable Chinese populations. Japan and South Korea have similar cultures and share many common values. Both Japan and South Korea also experienced high saving rates during their economic "take-off" periods. Japan's national saving rate was greater than 40% in the 1960s and 1970s, and is still relatively high today at 20% to 30%. South Korea's experience is similar to China's in that it also experienced a dramatic increase in its saving rate in the 1980s. Today, South Korea's national saving rate is virtually identical to China's.

The advantage of my approach is that I allow the savings function to be different for China with its OCP, and for the counterfactuals, where there is no OCP. This enables us to estimate the impact of the OCP both through its effect on the number of children ("endowment effect") and its effect on how children affect savings ("coefficient effect").

Overall, my decision to use other countries as counterfactuals for China without the OCP is motivated by some strict assumptions made by the existing literature. As the previous studies assume that the post-OCP reduction in the number of children was due to the OCP. In contrast, many studies conclude that the OCP had, at best, a minor effect on Chinese fertility. Second, they assume that the estimated relationship between household savings and children observed in the post-OCP period represents the relationship that would have existed in the absence of the OCP, in other words, OCP only changed the number of children, while it doesn't change how each child affects people's savings. That means that these studies have assumed that the counterfactual of Chinese saving behavior under the OCP is represented by the post-OCP saving behavior of Chinese households that have more than one child." In other words, they estimate the counterfactual of China without the OCP using data from China under the OCP.

Finally, there is a precedent for using the experiences of other countries as counterfactuals for China. The fertility literature has frequently employed the experiences of other countries to assess the impact of the OCP on the number of children.

This highlights the value of using other countries as a point of comparison. The countries that I have chosen for comparing saving behavior (in the published version of this paper) are largely the same countries that other studies have chosen for comparing fertility behavior. Given the hypothesized connection between saving and fertility, this seems appropriate.

I readily acknowledge that my counterfactual is imperfect, as my comparison countries differ from China in important ways and I will be able to control only for some of these differences in my analysis below. Nevertheless, given that no perfect counterfactual exists, it is essential to analyze whether the use of different, imperfect counterfactuals leads to similar results or not.

3.3 Data

The 2014 Gallup World Poll database and 2014 Global Findex database are the major sources of the data used in this analysis. The Gallup World Poll database is an annual survey that collects personal data from more than 160 countries in the world. Data is obtained through private face-to-face surveys or telephone interviews. More participants are interviewed for countries with a large population while the typical survey includes around 1000 people per country, 4696 Chinese people were surveyed in the 2014 Gullup World Poll database. The database includes detailed information on an individual's family size, children's number, income, ethnicity, employment status, sexual preference, education status, job, etc, information that will be used as control variables in the empirical analysis shown in the following sections.

The Global Findex survey, in collaboration with Gallup, was conducted every three years from 2011. In accordance with the World Poll Survey, Global Findex interviews are carried out by Gallup to ensure that each person could be linked and merged between the two surveys using different IDs.

For savings, Participants of the Global Findex surveys were asked, “In the past 12 months, have you, personally, saved or set aside any money for any reason?”, The questions were asked in a multiple-choice format, the following answers are provided and the respondent has to choose an answer in the following four options, “Yes, No, Don’t know, Refuse to answer.” In this thesis, my analysis uses a binary version of this question equalling 1 if the respondent chose Yes, and 0 otherwise (including the small share of respondents that answer don’t know or refuse to answer).

While the above question allows us to analyse differences in saving behavior between China and the counterfactuals, it does not measure the saving rate, the share of income saved by the household. As a result, my main research focus is on the likelihood that a person saves as a function of children and other household characteristics, including income.

The survey does have one question that addresses the size of a person's savings. One of the follow-up questions in the Global Findex survey is, "Now, imagine that you have an emergency and you need to pay the amount of 1/20 of GNI per capita (for example, 5866 RMB, or 838 US dollars for China). How possible is it that you could come up with this amount within the next month? Is it very possible, somewhat possible, not very possible, or not at all possible? "

This is followed by the question, "What would be the main source of money that you would use to come up with1/20 of GNI per capita (5866 RMB, or 838 US dollars) within the next month?", with savings being one of the answer options. For those participants choosing savings as the source of this amount of money, this indicates that they have a meaningful amount of savings. I use these questions to create a binary variable indicating that the individual has a meaningful amount of savings.

Respondents were also asked about the motives of their savings in the Global Findex survey. Specifically, participants were being asked, "In the past 12 months, have you personally saved or set aside any money for any of the following reasons? A: To start, operate, or grow a business or farm. B: For old age. C: For education or school fees." These questions allow us to separately examine the different reasons why people save.

The questions about savings all focus on the personal aspect, in the sense that they ask participants whether or not they "personally" saved or not, and what is the reason for their "personal" savings. One consequence of asking this type of question might be that young participants are unlikely to say yes because they have personal savings, as the majority of the young people are possible still living with family or living off loans. They may be more inclined to say yes because their family has savings or their friends have savings. Although in the empirical specifications age is being controlled for, it is being controlled in a linear way, the effect of age may be very non-linear. (For example, no savings before 23 and then and linear savings later.) So as a robustness check in the empirical analysis, I restricted the sample to include only participants who are above 23 years old.

To summarize, while the questions do not allow us to estimate the share of income going to savings, the questions do allow us to measure how savings patterns in China compare to savings in my counterfactual regions across a variety of dimensions.

The merged 2014 Gallup World Poll and 2014 Global Findex databases contain 146,688 individual observations from 142 countries. Descriptive statistics are shown in TABLES 3.2 (for saving behavior) and 3.3 (for the number of children) in the appendix.

One finding immediately apparent from these tables is that China is not that different when compared to the counterfactuals. For example, 69% of the participants in China reported having saved in the past 12 months, while the average of the counterfactual respondents is 74% (cf. "All (excl. China)"). Malaysia has the highest percentage, with 83% indicating that they saved. With respect to "Meaningful savings," 46% of the Chinese sample reported having meaningful savings, compared to 53% for the counterfactuals (cf. "All (excl. China)").

This pattern is also evident in the various saving motives. Compared to the World average, Chinese people are more likely to save for old age, education, and business. However, counterfactual respondents had the same or higher percentage of people who saved. 47% of the counterfactual respondents indicated that they were saving for old age, compared to 40% of Chinese respondents. 33% of counterfactual respondents saved for education, compared to 25% of Chinese respondents. Finally, 15% of both counterfactual and Chinese respondents reported that they have saved for business.

These survey findings may seem counter-intuitive at first. However, macro-level data from the counterfactual regions indicate that they also have high national saving rates. Singapore's gross saving rate was 48% in 2017, compared to 46% for China. The corresponding rates for Taiwan, Hong Kong, Japan, Malaysia, and South Korea are 34%, 27%, 25%, 30%, and 35% (OECD National Account data, 2017).

Similarities between China and the counterfactual regions are also evident with respect to the number of children (cf. TABLE 3.3 in the appendix). The average number of children per household in China is 0.56, compared to 0.61 for the counterfactuals (cf. "All (excl. China)"). Malaysia is a noteworthy outlier with 1.49 children per household. These numbers reflect low fertility rates in the macro-level data. The average number of children born to a woman over her lifetime in China is 1.63. This compares with 1.12 for Hong Kong, 1.16 for Singapore, 1.43 for Japan, 1.89 for South Korea, and 2.02 for Malaysia.

TABLE 3.10 in the Appendix reports descriptive statistics and definitions for the other variables used in my analysis. While there are many similarities, there are also notable differences. Chinese respondents have less income, are less likely to live in an urban area, have less education, and are more likely to be married compared to the counterfactuals.

3.4 Methodology

In this analysis, I'll start my analysis by presenting a savings model consistent with the model other studies are using in their empirical analysis, where the savings decision of a family is modelled as a function of the number of children and a series of household characteristics. In order to identify how the One-Child policy could've affected savings behavior, separately evaluating the endowment effect, which is how the number of children affects saving behaviour, and the coefficient effect, which is how each child affects savings behaviour differently is important. Therefore, in the empirical analysis, I adopted the Blinder-Oaxaca decomposition procedure (Blinder, 1973; Oaxaca, 1973; Jann, 2008) to divide the total difference in saving outcome between China and the counterfactuals into these components, while making sure that the effect of other household characteristics that could've affect people's saving behaviour are being controlled for

To start with, the savings model is given by:

$$Saving\ Decision = \alpha + \beta \cdot Children + Controls + \varepsilon \qquad (3.2)$$

In the equation above, $Saving\ Decision$ is a dummy variable taking the value 1 if the respondent had saved in the past 12 months while taking the value 0 if the respondent indicated that they didn't save in the past 12 months or prefer not to respond to the question. $Children$ correspond to the number of children in the respondent's household, and $Controls$ includes a wide variety of individual and household characteristics including income, gender, age, and education.

Blinder-Oaxaca decomposition is an estimation method that decomposes the difference of the outcome variable between two groups, into a component associated with group differences in the sample characteristics of the explanatory variables, and another component associated with group differences in the coefficients of the variables.

Let the difference between the means of a given outcome variable Y for two groups *A* and *B* be represented by Δ:

$$\Delta = E(Y_A) - E(Y_B). \tag{3.3}$$

Further, let Y for each of the groups be a function of explanatory variables X according to the following linear models,

$$\begin{aligned} Y_{Ai} &= X_{Ai}{}' \beta_A + \varepsilon_{Ai},\ E(\varepsilon_A) = 0 \\ Y_{Bi} &= X_{Bi}{}' \beta_B + \varepsilon_{Bi}, E(\varepsilon_B) = 0 \end{aligned} \tag{3.4}$$

where X is a vector of variables including a constant term, β is the associated vector of coefficients, and ε is the error term.

It follows that the difference in the means of the outcome variable for the two groups can be expressed as:

$$\Delta = E(Y_A) - E(Y_B) = E(X_A)' \beta_A - E(X_B)' \beta_B. \tag{3.5}$$

By adding and subtracting terms, we represent this difference by a threefold decomposition:

$$\Delta = [E(X_A) - E(X_B)]' \beta_B + E(X_B)'(\beta_A - \beta_B) + [E(X_A) - E(X_B)]'(\beta_A - \beta_B). \tag{3.6}$$

The first term on the right-hand side, $[E(X_A) - E(X_B)]' \beta_B$ is the component due to the difference in the means of the explanatory variables evaluated at the coefficients for group *B.* This is the “endowment effect.” The second term, $E(X_B)'(\beta_A - \beta_B)$ is the component due to the difference in the coefficients across the two groups evaluated at the sample means for group *B.* This is the "coefficient effect." The third term, $[E(X_A) - E(X_B)]'(\beta_A - \beta_B)$ is an interaction term that collects the remainder of the difference as the simultaneous difference of both means and coefficients of the two groups.

The decomposition procedure shown above is the threefold decomposition using linear probability model. I will use this decomposition method to analyse the difference in saving behavior between China and its counterfactuals, focusing specifically on the effect of children.

In the Blinder-Oaxaca decomposition procedure shown above, we assumed that the relationship between children and saving outcome is linear, and therefore we are able to decompose the group difference in savings according to a linear probability model. However, my measures of people's savings are in all cases binary variables, which means a nonlinear model will probably fit the situation better.

In the empirical analysis, I will also show the decomposition results using logit model, the logit Oaxaca decomposition procedure can be easily accessed using the "logit" option of the "oaxaca" command in Stata. Obtaining a detailed decomposition for a non-linear model such as logit or probit is not as straightforward as the linear decomposition showed above, due to the fact that $E(Y_A)$ and $E(Y_B)$ can not be easily subdivided into additive components. Specifically, the contribution of a certain variable *X* depends on the value of all other covariates (Sinning, Hahn & Bauer, 2008).

So far, there's no best way of dealing with this problem. The "logit" option in "oaxaca" command in Stata calculates the contributions to the outcome variable of each independent variable in relation to their relative contributions in a decomposition at the level of the linear predictor, which is suggested by Yun(2004).

3.5 Empirical results

The empirical results are shown in Tables 3.4 to 3.15 of the appendix. Table 3.4 shows the results of estimating Equation 2 for each of the countries independently using an OLS type model. Tables 3.5 to 3.9 show the decomposition results for different measures of savings and saving motives while focusing on the children variable. Tables 3.11 to 3.15 show the complete decomposition results for different measures of savings and saving motives aggregating all the variables. All the standard errors in the tables are Newey-West standard errors.

3.5.1 Baseline regression

Table 3.4 reports the results of estimating Equation (2) using OLS for each of the countries (China, Taiwan, Hong Kong, Japan, Singapore, Malaysia, and South Korea). The dependent variable is "Saving," which For China, the corresponding estimate is -0.03, and the coefficient is significant at 1% significance level, which means that there's a negative relationship between the number of children and the families' decision to save. An interpretation of the coefficient indicates that, if the number of children in Chinese households increases by 1, the associated decline of the percentage of people who indicated that they have saved in the past 12 months is 3 percentage-point. If I were to interpret this estimate as other studies in this field, then I would get the conclusion that, because the OCP reduced the average number of children per household from 3 to 1, which is a 2 children decrease. This would imply a 6 percentage point decrease in people's probability of saving being attributed to children.

Focusing on the estimation results for the other counterfactual regions, it's clear that this negative correlation among children and the probability of people reporting that they saved for the past 12 months is not evident in the counterfactual areas. If we look at the magnitude of these estimates, these coefficients are in all cases positive or very close to zero for all the counterfactual areas. This indicates that the view that the effect of children on saving is different for China is a valid assumption to a certain degree. However, the coefficients I got for the number of children are small in absolute value for both China and the counterfactual regions. At best, the difference in the estimated effects of children on saving can only explain a small amount of the total difference in saving behavior.

3.5.2 Oaxaca decomposition results, linear probability model (LPM).

Table 3.5.A shows the decomposition for people's probability to save while focusing on the children variable. It decomposes the overall difference in the probability of saving into the three components ("Endowments," "Coefficients," and "Interaction") using a linear probability model. The first two rows report the unconditional difference in the probability of saving for a Chinese respondent compared to a respondent from the respective counterfactual region. The third row reports the difference, with positive (negative) numbers indicating that the probability is larger (smaller) for Chinese respondents.

The next four rows show the decomposition result when applying the Blinder-Oaxaca decomposition to the overall difference of the outcome variable. "TOTAL" estimates the total difference in saving behavior between Chinese respondents and their counterfactuals that is attributed to children. The three rows above break the difference into the individual "Endowment", "Coefficient", and "Interaction" components, where the sum of these is equal to "TOTAL". In all cases, the reference group is China.

The "Endowment" component identifies how much counterfactuals' probabilities of saving would change if counterfactual respondents had the same number of children, on average, as Chinese respondents. For example, the "Endowment" component for Hong Kong is 0.003. This indicates that the percentage of Hong Kong respondents who save would increase by 0.3 percentage points if Hong Kong households had the same number of children, on average, as Chinese households. This is a very small effect. I find similar small effects for the "Endowment" component when using other regions as counterfactuals.

The "Coefficient" component quantifies the change in the predicted percentage of counterfactuals who would save if their relationship between children and saving was the same as China's, as represented by the estimated coefficient on the number of children in China's savings function.

Overall, the coefficient effect is larger in absolute value than the endowment effect. The largest coefficient effect I estimated (in absolute value) is -0.048 for Malaysia. This indicates that the percentage of Malaysia's respondents who save would be 4.8 percentage points less if Malaysians saving behavior was governed by China's savings function. When I estimate a pooled saving function for all the counterfactual regions, the associated children's coefficient leads to a coefficient effect of only -0.26 percentage points.

Across all counterfactual regions, the estimated coefficient effects I get range from -0.5 percentage points (South Korea) to -4.8 percentage points (Malaysia). This indicates that children have a stronger, negative effect on savings in China. This helps explain why counterfactuals tend to have a higher share of people who were able to save in the past 12 months, though the effect is small.

Finally, the "Interaction" component measures the simultaneous effect of differences in endowments and coefficients. Similar to the endowment and coefficient effects, I find the interaction effect for the number of children variable to be small whatever counterfactual used.

By comparing the "TOTAL" and "DIFFERENCE" rows in TABLE 3.5, we see that the children variable can explain a sizeable portion of the overall difference between China and the counterfactuals in the percentage of people that save. For example, when pooling all the counterfactuals (cf. "All (excl. China)"), about half of the difference between this pooled counterfactual and China can be explained by the number of children (compare -0.025 with -0.054). More precisely, since the interaction and endowment effects are very small, most of the overall effect of children can be explained by the higher negative impact of children on savings in China (that is, the coefficient effect).

Although the children variable can explain a sizeable share of the difference in the percentage of people that save between China and the counterfactuals, the overall impact is still small. Indeed, the combined effect I get for children ranges from -1.2 percentage points for South Korea to -4.1 percentage points for Japan. This, combined with the overall small difference in saving behavior between China and the counterfactual regions, suggests that the OCP has not had a substantial impact on Chinese saving behavior.

As a robustness check, in Table 3.5.B I restricted the sample to participants whose age is above 23, the reason for this is explained in section 3.3, the Data part of this thesis. By comparing the “DIFFERENCE” rows between table 3.5.A and 3.5.B, we can see that after restricting my sample, the difference in people’s decision to save between China and the counterfactuals reduced. For example, the difference in my dependent variable for Taiwan is 5.7 percentage points. This number is reduced to 4.4 percentage points when using the restricted sample. This indicates that people’s saving decision varies greatly among young people in different jurisdictions.

Overall, the results are consistent with the results I got when using the full sample. The biggest endowment effect I get is for Hong Kong, which is 0.9 percentage points. The small endowment effects indicated that the amount of people who save does not change that much if the counterfactuals have China’s number of children. Consistent with the previous results using the entire sample, once again, the coefficient effects are bigger in absolute value and more significant compared to the endowment effect. The estimates I get range between -0.7 percentage points for South Korea and 6.3 percentage points for Malaysia. This suggests that the difference in the saving function between China and the counterfactuals can explain a sizeable amount of the difference.

It’s worth noting that after restricting the sample size to only include individuals whose above 23 years old, the coefficient effects are bigger compared to the decomposition results when I was using the full sample, although this is not a huge increase. Overall, the numbers are still small in absolute size. By comparing the “TOTAL” rows in Table 3.5.A and 3.5.B, we see that the amount of difference children variable can explain of the overall difference between China and the counterfactuals in the percentage of people that save have increased. For example, the pooled result using all the counterfactuals (cf. “All (excl. China)”), indicated that children can explain 2.5 percentage points of the difference when using the full sample, it now can explain 3.1 percentage points of the difference. Although consistent with the earlier results, most of the overall effect of children is attributed to the relatively bigger coefficient effects.

Overall, the combined effect for children ranges from -1.8 percentage points for South Korea to -4.3 percentage points for Japan. These are still small effects especially considering that the overall difference in saving behavior between China and the counterfactual regions is also reduced when I set restrictions to the sample.

So far, I have focused on the decision to save or not. Next, I will focus on having “meaningful savings,” as this is likely to be more closely associated with the saving rate. Table 3.6.A shows the decomposition result for this outcome variable. Based on the estimates from the respective saving functions, I find that the endowment, coefficient, and interaction effects are all small. Overall, children seem to play an even smaller role in explaining differences in “meaningful savings” than for savings in general. Table 3.6.B shows the decomposition result of the restricted sample, after deleting all observations below the age of 23. Overall, the magnitude of the estimates didn’t change that much compared to when using the full sample. The endowment, coefficient, and the interaction effects are still small, and the combined effect cannot explain a huge difference in the percentage of people who have “meaningful savings”.

Next, I focus on the specific reasons to save. Tables 3.7-3.9 show the decomposition results for people’s decision to save for old age, education, and business, respectively.

Table 3.7.A and 3.7.B focuses on the decision to save for old age. Children are sometimes viewed as a source of old-age support, as one possible motive to raise children is to provide parents financial support when they get older. The OCP could push people to save more for their old age because they have fewer children to rely on. My decomposition results offer little evidence for this.

The endowment effects estimated are close to zero across all counterfactual regions, and this result is robust in the robustness check in Table 3.7.B. For example, the percentage of people who save for old age in the pooled counterfactuals sample ("All (excl. China)") would decrease by only 0.1 percentage points if respondents in the counterfactual regions had the same number of children, on average, as Chinese respondents according to the estimation of the full sample.

Similarly, I find that the coefficient effect is mostly negative. This suggests that the percentage of counterfactual respondents saving for old age would be lower if their saving behavior was governed by China's savings function. While the coefficient effect can explain a sizeable part of the (relatively small) difference in the percentage of people saving for old age, the effect is again small in absolute value.

Table 3.8.A and 3.8.B reports decomposition results for people's decision to save for education. Education is one of the largest expenditures in raising a child. Further, with fewer children, Chinese families may invest more in education to ensure a better education outcome, so that their one child will be better able to support them in their old age. As a result, one could expect the OCP to have a substantial impact on respondents' decision to save for education.

Once again, however, comparisons of the saving behavior of the Chinese and counterfactual respondents do not identify major differences (except for Malaysia). When decomposing the overall difference associated with the effect of children, the absolute sizes of the overall differences that can be explained by children are insubstantial, ranging from -0.5 to 1.6 percentage points when using the full sample, and ranging from -0.9 to 2.3 when using the restricted sample. Even in Malaysia, where the overall difference is relatively large, the total effect of children is small.

Table 3.9.A and 3.9.B report my final decomposition results, for the decision to save for business. I expect that the decision to save for business would be largely unaffected by the number of children and the results are consistent with this. All three of the components are close to 0 for each of the counterfactual regions.

3.5.3 Oaxaca decomposition results, logit model

So far, the empirical results are all Oaxaca decomposition using the linear probability model. Table C of 3.5-3.9 shows the decomposition results using logit models.

Table 3.5.C shows the logit decomposition results of the contribution of children to people's probability to save. Overall, the results are not that different compared to the results I got using a linear probability model. The largest endowment effects are for Singapore and Japan, which is 0.15 and 0.12 percentage points respectively, the rest of the endowment effects are close to 0 percentage points, regardless of the counterfactual used, which is consistent with the results in Table 3.5.A using linear probability model. In terms of the coefficient effects, once again, they are bigger in absolute value compared to the endowment effects. The biggest coefficient effect is 3.8 percentage points for Japan, this number drops to 1.9 percentage points when using the pooled data using all the counterfactuals. When we compare the combined explaining ability of children between the decomposition results of the LPM and logit model, it's clear that there's a small drop in magnitude when using the logit model. The combined effect of children varies between 0 and 2 percentage points, which indicates that children cannot explain a huge amount of the difference in the percentage of people who save.

Moving on to my proxy for people's ability to save for a certain amount of money. Table 3.6.C shows the logit decomposition results of having a "meaningful amount of savings". The results I got confirm the decomposition result of the LPM. The endowment effects are all small in magnitude and insignificant, while the coefficient effects are bigger and more significant. The combined effects of children range between 1.1 to 4.6 percentage points, which indicates that children can explain a sizeable amount of the difference, although in general, the numbers are still quite small.

I next focus on the logit decomposition results for different saving motives in Table C of 3.7 to 3.9, in all three cases, the results are consistent with the results I got when using LPM. With the endowment effects in most cases close to 0 and insignificant, while the coefficient effects are slightly bigger in absolute value and more significant. The combined effect indicates that the explaining power of children can contribute to a sizeable part of the difference in people's decision to save for old age and education, and I did not find a huge effect of children on people's decision to save for business.

3.5.3 Empirical Summary

In conclusion, my results indicate that the OCP is not a likely candidate to explain China's high saving rate, and this result is robust regardless of the sample used (full sample or only participants above 23 years old) or the model used (LPM or logit). At the micro-level, respondents' saving behavior in China is not very different from their behavior in related regions that do not have restrictive population policies. Further, when I focus specifically on the effect of the number of children on saving, I find that the sizes of the associated effects are small.

3.6 Conclusion

To economists, China's high savings rate has been a long-standing mystery. In this thesis, a series of possible explanations for the increase of savings rate are presented, then I focused on one of the major explanations relating the Chinese saving puzzle to the One-Child Policy (OCP), a population control policy implemented ever since the 1970s.

The One-Child policy is hypothesized to influence savings rate by reducing the number of children a household have. With fewer children to help and support them when they're old, parents are expected to respond by raising their savings to fund retirement. This explanation has a lot of appeal, particularly since the increase in the saving rate in China occurred at about the same time as the OCP implementation. Accordingly, several studies, consistent with the OCP hypothesis, indicate a substantial negative relationship between savings and boys.

A closer examination, however, raises doubts. Not all studies found significant effects on savings as a result of the decrease in the number of children. During the heady reform era of the early 1980s, many other changes took place in China apart from the OCP. As a result, the increase in savings in China could be attributed to many different reasons. In addition, previous studies finding a negative correlation between children and saving during the time the OCP was in place based their study on household data collected after the implementation of the OCP. This means that, these studies implicitly assume that if there's no OCP, families will behave exactly like households whose family has more than one child living under OCP. In other words, these studies are assuming a “counterfactual” for China’s OCP that relies on the saving behavior of households living within an OCP environment.

While testing the hypothesis that OCP is responsible for the increase in the savings rate, my study makes two methodological advances. First, I used another counterfactual to explain the OCP's influence. I compared the Chinese people's saving behaviour with the saving behaviour of people from other Asian regions sharing similar cultural and demographic characteristics (Taiwan, Hong Kong, Japan, Singapore, Malaysia, and South Korea). Through integrating information from two datasets: the 2014 Gallup World Poll and the 2014 Global Findex survey. I was able to use other regions’ current situation as a counterfactual for China without the OCP. The combined micro-datasets also allow me to match with saving behaviour with a large number of personal characteristics.

Second, the Blinder-Oaxaca decomposition method is used to separately evaluate the “endowment” and the “coefficient” effect of how children affect saving. This is particularly important because it is possible for OCP to affect saving through two different channels. Either directly by setting a restriction on the number of children a household can have (the “endowment effect”), or indirectly, by altering the nature of how each child empirically affects savings. (the “coefficient effect”)

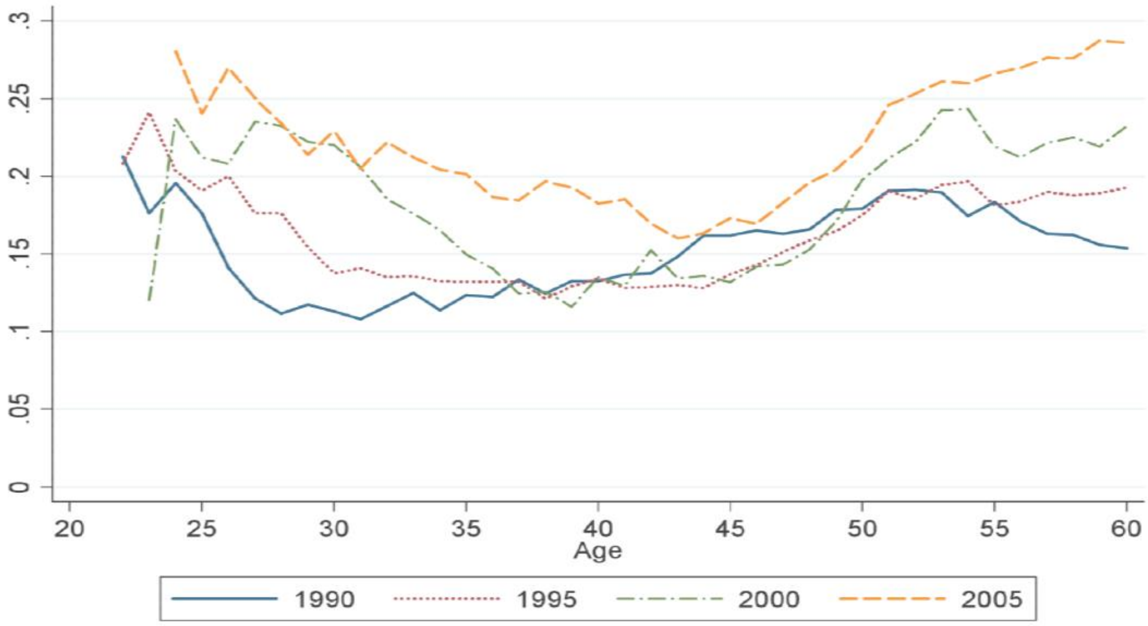
Previous studies have been forced to assume that with and without OCP, how each child affect savings decision is the same. This assumption had to be made because the only data available to them was for households living under the influence of the OCP. In contrast, my data allows the empirical relationship between children and saving to differ between China and counterfactuals that do not have restrictive population policies.

My main empirical finding is that there is little substantial difference between Chinese people's saving behavior and to those families living in the "counterfactual" areas. This indicates that the OCP is not a compelling explanation for the high savings rate in China. However, while the Blinder-Oaxaca decomposition result suggests that children may explain some of the variations in saving behavour between Chinese people and counterfactual people, the results are too small to support a major role for children as a saving behaviour determinant.

My findings are important for understanding current discussions on how China could switch from an investment economy to a consumer economy. My results do not support the statement that the relaxation of China's OCP can boost consumption of Chinese households.

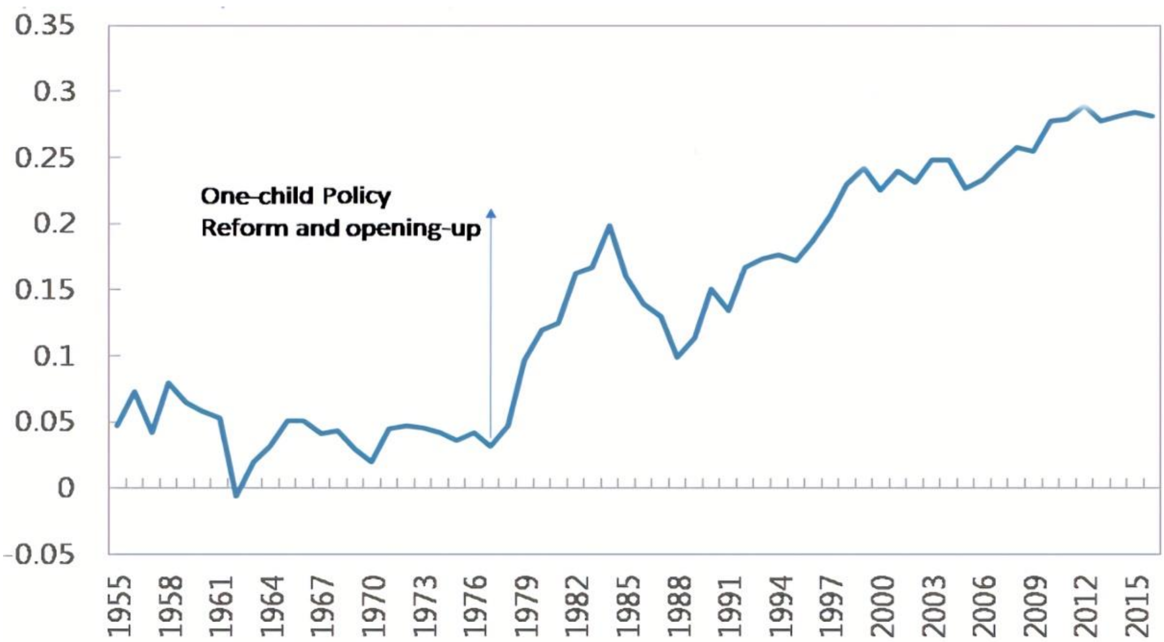
3.7 Appendix

Figure 3. 1 China's age saving profile



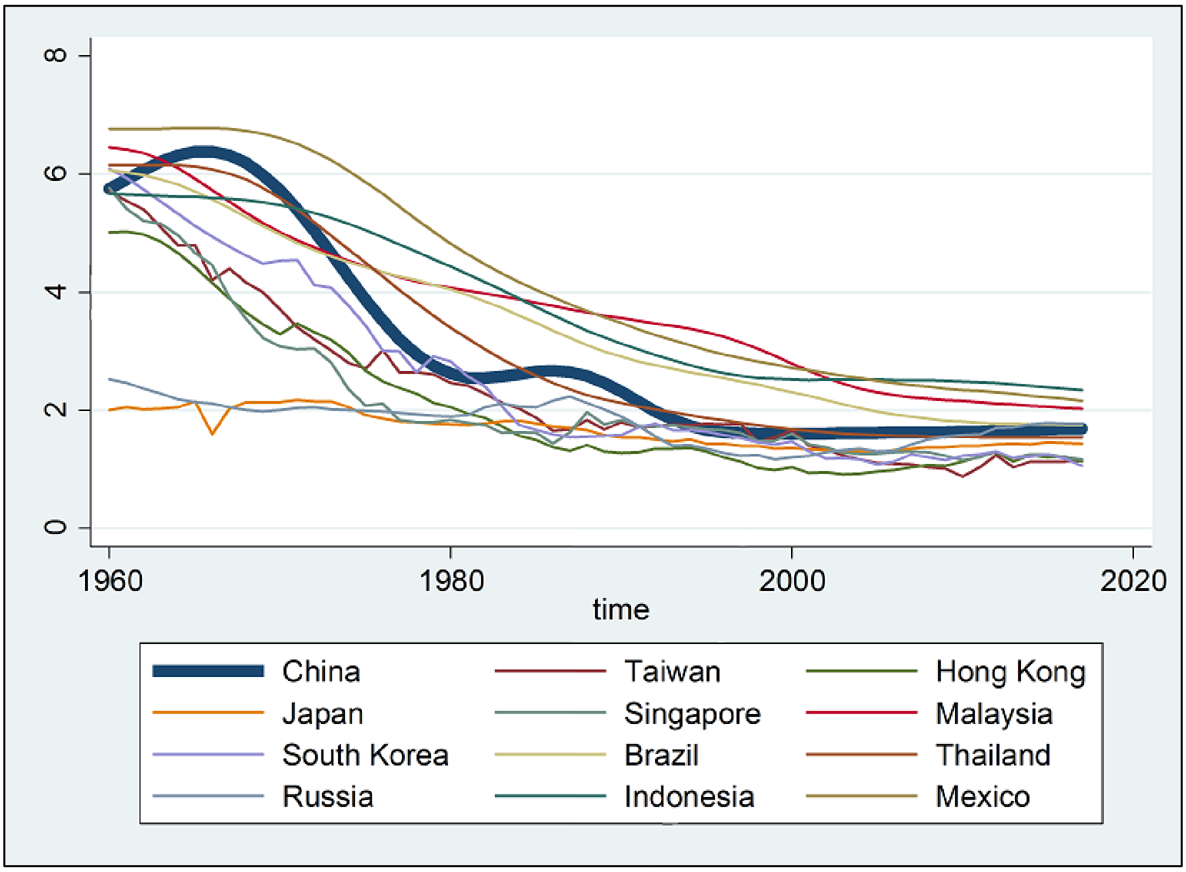
(Source: Chamon, Liu & Prasad, 2010)

Figure 3. 2 China's household saving rate, 1955-2015



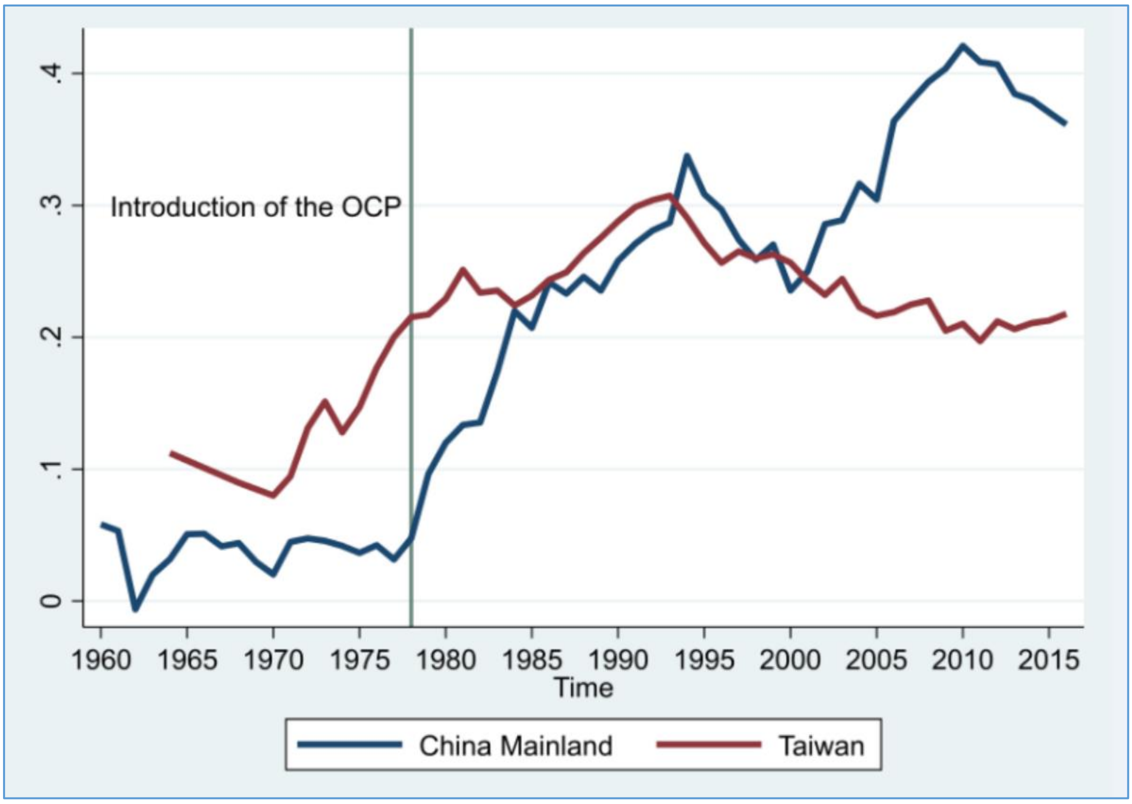
(Source: IMF, 2017)

Figure 3. 3 Fertility rate of China and other regions: 1960-2017



(Source: Source: The World Bank, 2017)

Figure 3. 4 Household savings rate, 1955-2015: China VS. Taiwan



(Source: Source: The World Bank, 2017)

Table 3.1 What have other studies used as counterfactuals for assessing the effect of the One-Child Policy?

Dependent Variable	Children Variable	Authors	Endowment Effect?	Coefficient Effect?	Counterfactual
Saving rate	No. of children	Ge, Yang & Zhang (2015)	Yes	No	Different age cohorts of Chinese household with more than one child
Saving rate	No. of children	Lugauer, Ni, & Yin (2017)	Yes	No	Chinese households with more than one child
Saving rate	No. of brothers or siblings	Zhou (2014)	Yes	No	Chinese households with male siblings
Saving rate	No. of children	Banerjee, Meng, & Qian (2010)	Yes	No	Chinese households with more than one child and the first child is a daughter
Saving rate	Dummy for twins	Choukhmane, Coeurdacier, & Jin (2014)	Yes	No	Chinese households with twins
Saving rate	No. of children	Curtis, Lugauer, & Mark (2015)	Yes	No	Chinese households with more than one child

Table 3.2 Descriptive statistics for saving behavior by country

Saving Variable	Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
"Save"	Obs.	4184	1000	1007	1006	1000	1000	1000	6013	146688
	Mean	0.69	0.74	0.67	0.73	0.73	0.83	0.74	0.74	0.54
	Std. Dev.	0.46	0.44	0.47	0.44	0.44	0.38	0.44	0.44	0.5
	Min	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1
"Meaningful savings"	Obs.	4184	1000	1007	1006	1000	1000	1000	6013	146688
	Mean	0.46	0.5	0.59	0.77	0.52	0.33	0.45	0.53	0.27
	Std. Dev.	0.5	0.5	0.49	0.41	0.5	0.47	0.5	0.5	0.44
	Min	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1
"Saving for old age"	Obs.	4184	1000	1007	1006	1000	1000	1000	6013	146688
	Mean	0.4	0.46	0.39	0.46	0.48	0.58	0.45	0.47	0.2
	Std. Dev.	0.49	0.5	0.49	0.5	0.5	0.49	0.5	0.5	0.4
	Min	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1
	Obs.	4184	1000	1007	1006	1000	1000	1000	6013	146688

Saving Variable	Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
"Saving for education"	Mean	0.25	0.35	0.27	0.25	0.32	0.51	0.3	0.33	0.19
	Std. Dev.	0.43	0.48	0.44	0.43	0.47	0.5	0.46	0.47	0.39
	Min	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1
"Saving for business"	Obs.	4184	1000	1007	1006	1000	1000	1000	6013	146688
	Mean	0.15	0.19	0.09	0.04	0.11	0.2	0.27	0.15	0.12
	Std. Dev.	0.35	0.39	0.29	0.2	0.31	0.39	0.44	0.36	0.39
	Min	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1

SOURCE: Authors' calculations based on the Gallup World Poll and Global Findex.

Table 3.3 Descriptive statistics for number of children in the household by country

Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
Obs.	4184	1000	1007	1006	1000	1000	1000	6013	142504
Mean	0.56	0.55	0.44	0.41	0.44	1.49	0.31	0.61	1.2
Std. Dev.	0.84	0.93	0.8	0.83	0.79	1.9	0.7	1.14	1.69
Min	0	0	0	0	0	0	0	0	0
Max	8	6	6	4	5	30	5	30	34

Table 3.4 Determinants of "Save": China and counterfactuals

Variable	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
childnum	-0.030*** (0.009)	0.013 (0.015)	0.023 (0.018)	0.056*** (0.017)	0.040** (0.018)	0.002 (0.006)	-0.013 (0.018)	0.008* (0.005)
businessown	0.077*** (0.019)	0.122*** (0.035)	0.004 (0.045)	0.002 (0.044)	0.173*** (0.042)	0.042 (0.027)	0.045 (0.031)	0.062*** (0.015)
income	0.000* (0.000)	0.000 (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000* (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000** (0.000)
gender	0.064*** (0.014)	-0.002 (0.026)	0.033 (0.028)	-0.040 (0.030)	0.026 (0.028)	-0.046* (0.023)	0.006 (0.027)	-0.003 (0.011)
age	0.003 (0.002)	-0.007 (0.005)	-0.002 (0.005)	0.008 (0.006)	0.002 (0.005)	0.001 (0.004)	0.012** (0.005)	0.000 (0.002)
agesq	-0.000*** (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000*** (0.000)	-0.000** (0.000)
help	0.163*** (0.019)	0.140*** (0.026)	0.120*** (0.029)	0.036 (0.039)	0.075** (0.037)	0.101*** (0.023)	0.088*** (0.025)	0.105*** (0.012)
urban	0.083*** (0.018)	-0.058** (0.027)	0.036 (0.055)	0.017 (0.028)	----	0.028 (0.024)	-0.023 (0.041)	-0.008 (0.014)
sector	0.006 (0.014)	0.056** (0.026)	0.004 (0.028)	-0.012 (0.037)	0.048* (0.029)	-0.011 (0.026)	-0.023 (0.025)	0.014 (0.011)
food	0.015 (0.028)	0.169*** (0.043)	0.175*** (0.061)	0.183** (0.072)	0.327*** (0.090)	0.079** (0.033)	0.128*** (0.037)	0.151*** (0.019)

Variable	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
shelter	0.069***	-0.058	0.013	0.035	-0.107	0.019	0.063**	0.022
	(0.022)	(0.045)	(0.049)	(0.071)	(0.112)	(0.029)	(0.031)	(0.017)
receivemoney	-0.002	-0.053	0.020	0.076	0.007	-0.008	0.004	0.003
	(0.022)	(0.085)	(0.041)	(0.067)	(0.044)	(0.030)	(0.038)	(0.018)
secondary	0.102***	0.165***	0.138***	0.177***	0.129***	0.118***	0.127**	0.158***
	(0.017)	(0.048)	(0.048)	(0.050)	(0.044)	(0.042)	(0.052)	(0.019)
college	0.112***	0.270***	0.273***	0.239***	0.107**	0.161***	0.179***	0.225***
	(0.028)	(0.050)	(0.052)	(0.054)	(0.052)	(0.044)	(0.058)	(0.021)
married	0.102***	0.021	0.007	0.109***	-0.007	0.009	0.103***	0.040***
	(0.021)	(0.040)	(0.036)	(0.033)	(0.036)	(0.031)	(0.038)	(0.014)
borrowed	0.021	0.006	0.023	0.023	0.113***	0.100***	0.057**	0.052***
	(0.015)	(0.028)	(0.030)	(0.034)	(0.031)	(0.026)	(0.025)	(0.012)
welfare	0.031	0.031	0.022	-0.002	0.030	0.018	0.011	0.019
	(0.020)	(0.042)	(0.039)	(0.040)	(0.031)	(0.025)	(0.034)	(0.014)
_cons	0.412***	0.638***	0.433***	0.074	0.313**	0.542***	0.116	0.397***
	(0.062)	(0.106)	(0.136)	(0.166)	(0.125)	(0.093)	(0.119)	(0.047)
Obs.	4184	1000	1007	1006	1000	1000	1000	6013
R-squared	0.085	0.166	0.192	0.120	0.095	0.109	0.220	0.124
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Country Dummy	NO	NO	NO	NO	NO	NO	NO	YES

NOTE: Standard errors are in parentheses. *** p<.01, ** p<.05, * p<.1

Table 3.5.A Decomposition results: "Save/Children" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVE"			
China	0.686	0.686	0.686	0.686	0.686	0.686	0.686
Counterfactual	0.743	0.670	0.732	0.730	0.828	0.743	0.741
DIFFERENCE	***-0.057***	***0.016***	***-0.045***	***-0.044***	***-0.142***	***-0.057***	***-0.054***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVE"			
Endowments	0.000	0.003	0.008***	0.005*	-0.002	-0.003	-0.001*
Coefficients	-0.024**	-0.024***	-0.036***	-0.031***	-0.048***	-0.005	-0.026***
Interaction	0.000	-0.007**	-0.013***	-0.008***	0.030**	-0.004	0.002**
TOTAL	***-0.024***	***-0.028***	***-0.041***	***-0.034***	***-0.020***	***-0.012***	***-0.025***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1   Complete decomposition result shown in TABLE 11 in the Appendix.

Table 3.5.B Decomposition results: “Save/Children” sample above 23 years old, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – “SAVE”			
China	0.692	0.692	0.692	0.692	0.692	0.692	0.692
Counterfactual	0.735	0.645	0.742	0.725	0.815	0.750	0.735
DIFFERENCE	***-0.044***	***0.05***	***-0.051***	***-0.033***	***-0.124***	***-0.059***	***-0.043***
				DECOMPOSITION – EFFECT OF CHILDREN ON “SAVE”			
Endowments	0.000	0.003	0.009***	0.005*	-0.002	-0.005	-0.000
Coefficients	-0.034***	-0.030***	-0.036***	-0.038***	-0.063***	-0.007	-0.032***
Interaction	-0.001	-0.008**	-0.016***	-0.009***	0.038**	-0.006	0.001
TOTAL	***-0.035***	***-0.035***	***-0.043***	***-0.042***	***-0.027***	***-0.018***	***-0.031***
				Observations			
China	3831	3831	3831	3831	3831	3831	3831
Counterfactual	858	865	975	839	801	869	5207

*** p<.01, ** p<.05, * p<.1

Table 3.5.C Decomposition results: “Save/Children” full sample, logit model

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – “SAVE”			
China	0.686	0.686	0.686	0.686	0.686	0.686	0.686
Counterfactual	0.743	0.670	0.732	0.730	0.828	0.743	0.741
DIFFERENCE	***-0.057***	***0.016***	***-0.045***	***-0.044***	***-0.142***	***-0.057***	***-0.054***
				DECOMPOSITION – EFFECT OF CHILDREN ON “SAVE”			
Endowments	0.000	0.003	0.012***	0.015*	-0.003	-0.002	-0.001*
Coefficients	-0.017*	-0.015**	-0.038**	-0.030***	-0.027**	-0.006	-0.019***
Interaction	0.000	0.001	-0.020	-0.007**	0.014	-0.005	0.001*
TOTAL	***-0.017***	***-0.011***	***-0.046***	***-0.022***	***-0.016***	***-0.013***	***-0.025***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1

Table 3.6.A Decomposition results: "Meaningful Savings/Children" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "MEANINGFUL SAVINGS"			
China	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Counterfactual	0.505	0.59	0.772	0.523	0.33	0.451	0.529
DIFFERENCE	***-0.044***	***0.13***	***-0.312***	***-0.063***	***0.13***	***-0.009***	***-0.069***
				DECOMPOSITION – EFFECT OF CHILDREN ON "MEANINGFUL SAVINGS"			
Endowments	0.000	0.001	0.003	0.005*	0.003	-0.002	-0.000
Coefficients	-0.023**	-0.013	-0.016**	-0.027***	-0.021	-0.003	-0.007
Interaction	-0.001	-0.036	-0.006*	-0.007**	0.013	-0.003	0.000
TOTAL	***-0.024***	***-0.048***	***-0.019***	***-0.029***	***-0.005***	***-0.008***	***-0.007***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1   Complete decomposition result shown in TABLE 12 in the Appendix.

Table 3.6.B Decomposition results: "Meaningful Savings/Children" sample above 23 years old, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
			OVERALL DIFFERENCE – "MEANINGFUL SAVINGS"				
China	0.474	0.474	0.474	0.474	0.474	0.474	0.474
Counterfactual	0.521	0.617	0.792	0.561	0.347	0.480	0.561
DIFFERENCE	***-0.047***	***-0.144***	***-0.318***	***-0.088***	***0.127***	***-0.006***	***-0.087***
			DECOMPOSITION – EFFECT OF CHILDREN ON "MEANINGFUL SAVINGS"				
Endowments	0.001	0.000	0.002	0.004	0.002	-0.003	-0.000
Coefficients	-0.033**	-0.013	-0.016**	-0.029***	-0.038	-0.005	-0.010
Interaction	-0.001	-0.003	-0.007*	-0.007**	0.023	-0.005	0.000
TOTAL	***-0.033***	***-0.016***	***-0.021***	***-0.032***	***-0.013***	***-0.013***	***-0.010***
				Observations			
China	3831	3831	3831	3831	3831	3831	3831
Counterfactual	858	865	975	839	801	869	5207

*** p<.01, ** p<.05, * p<.1

Table 3.6.C Decomposition results: "Meaningful Savings/Children" full sample, logit model

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "MEANINGFUL SAVINGS"			
China	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Counterfactual	0.505	0.59	0.772	0.523	0.33	0.451	0.529
DIFFERENCE	***-0.044***	***0.13***	***-0.312***	***-0.063***	***0.13***	***-0.009***	***-0.069***
				DECOMPOSITION – EFFECT OF CHILDREN ON "MEANINGFUL SAVINGS"			
Endowments	0.000	0.002	0.003	-0.088	0.005	-0.003	0.000
Coefficients	-0.027**	-0.014*	-0.015	-0.032***	-0.019	-0.004	-0.010
Interaction	-0.001	-0.004	-0.005	0.004	0.013	-0.003	0.001
TOTAL	***-0.028***	***-0.016***	***-0.017***	***-0.116***	***-0.001***	***-0.010***	***-0.009***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1

Table 3.7.A Decomposition results: "Saving for Old Age/Children" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR OLD AGE"			
China	0.397	0.397	0.397	0.397	0.397	0.397	0.397
Counterfactual	0.457	0.387	0.462	0.485	0.58	0.449	0.47
DIFFERENCE	***-0.059***	***0.01***	***-0.065***	***-0.087***	***-0.182***	***-0.051***	***-0.072***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR OLD AGE"			
Endowments	0.000	0.001	-0.003	0.002	-0.006	0.001*	-0.001*
Coefficients	-0.032***	-0.02**	-0.007	-0.025**	-0.067***	-0.024**	-0.034***
Interaction	-0.001	-0.006*	-0.002	-0.007**	0.041***	-0.02***	0.002**
TOTAL	***-0.033***	***-0.025***	***-0.012***	***-0.03***	***-0.032***	***-0.043***	***-0.033***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1   Complete decomposition result shown in TABLE 13 in the Appendix.

Table 3.7.B Decomposition results: "Saving for Old Age/Children" sample above 23 years old, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR OLD AGE"			
China	0.426	0.426	0.426	0.426	0.426	0.426	0.426
Counterfactual	0.510	0.426	0.475	0.528	0.597	0.495	0.503
DIFFERENCE	***-0.084***	***-0.000***	***-0.049***	***-0.102***	***-0.170***	***-0.069***	***-0.077***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR OLD AGE"			
Endowments	0.000	0.001	-0.004	0.001	0.005	0.008	-0.000
Coefficients	-0.032***	-0.024**	-0.007	-0.025**	-0.054***	-0.023**	-0.029***
Interaction	-0.001	-0.006*	-0.003	-0.006*	0.033**	-0.020***	0.001**
TOTAL	***-0.033***	***-0.029***	***-0.014***	***-0.030***	***-0.016***	***-0.035***	***-0.028***
				Observations			
China	3831	3831	3831	3831	3831	3831	3831
Counterfactual	858	865	975	839	801	869	5207

*** p<.01, ** p<.05, * p<.1   Complete decomposition result shown in TABLE 13 in the Appendix.

Table 3.7.C Decomposition results: "Saving for Old Age/Children" full sample, logit model

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR OLD AGE"			
China	0.397	0.397	0.397	0.397	0.397	0.397	0.397
Counterfactual	0.457	0.387	0.462	0.485	0.58	0.449	0.47
DIFFERENCE	***-0.059***	***0.01***	***-0.065***	***-0.087***	***-0.182***	***-0.051***	***-0.072***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR OLD AGE"			
Endowments	0.000	0.001	-0.003	0.005	-0.006	0.012	-0.001*
Coefficients	-0.016	-0.016*	-0.009	-0.023**	-0.060***	-0.027**	-0.027***
Interaction	-0.001	-0.003	-0.003	-0.005	0.037***	-0.015	0.002**
TOTAL	***-0.017***	***-0.018***	***-0.015***	***-0.023***	***-0.029***	***-0.030***	***-0.028***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1

Table 3.8.A Decomposition results: "Saving for Education/Children" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
			OVERALL DIFFERENCE – "SAVING FOR EDUCATION"				
China	0.253	0.253	0.253	0.253	0.253	0.253	0.253
Counterfactual	0.354	0.268	0.247	0.323	0.506	0.299	0.332
DIFFERENCE	***-0.101***	***-0.015***	***0.007***	***-0.07***	***-0.253***	***-0.046***	***-0.08***
			DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR EDUCATION"				
Endowments	0.000	0.008***	0.026***	0.017***	-0.040***	0.028***	-0.003**
Coefficients	0.016	-0.004	-0.05***	-0.038***	0.016	-0.018**	-0.009
Interaction	0.000	-0.001	-0.018***	-0.011***	-0.001	-0.015**	0.001
TOTAL	***0.016***	***0.003***	***-0.042***	***-0.032***	***-0.025***	***-0.005***	***-0.011***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1   Complete decomposition result shown in TABLE 14 in the Appendix.

Table 3.8.B Decomposition results: "Saving for Education/Children" sample above 23 years old, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
			OVERALL DIFFERENCE – "SAVING FOR EDUCATION"				
China	0.254	0.254	0.254	0.254	0.254	0.254	0.254
Counterfactual	0.344	0.234	0.249	0.298	0.483	0.297	0.314
DIFFERENCE	***-0.090***	***0.020***	***0.005***	***-0.044***	***-0.230***	***-0.043***	***-0.060***
			DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR EDUCATION"				
Endowments	0.001	0.007***	0.029***	0.018***	-0.049***	0.033***	-0.001
Coefficients	0.006	-0.008	-0.048***	-0.050***	-0.012	-0.023***	-0.021***
Interaction	0.000	-0.002	-0.021***	-0.013***	0.007	-0.020***	0.001
TOTAL	***0.007***	***-0.003***	***-0.040***	***-0.045***	***-0.054***	***-0.007***	***-0.021***
			Observations				
China	3831	3831	3831	3831	3831	3831	3831
Counterfactual	858	865	975	839	801	869	5207

*** p<.01, ** p<.05, * p<.1

Table 3.8.C Decomposition results: “Saving for Education/Children” full sample, logit model

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
			OVERALL DIFFERENCE – “SAVING FOR EDUCATION”				
China	0.253	0.253	0.253	0.253	0.253	0.253	0.253
Counterfactual	0.354	0.268	0.247	0.323	0.506	0.299	0.332
DIFFERENCE	***-0.101***	***-0.015***	***0.007***	***-0.07***	***-0.253***	***-0.046***	***-0.08***
			DECOMPOSITION – EFFECT OF CHILDREN ON “SAVING FOR EDUCATION”				
Endowments	0.000	0.007***	0.003	0.012***	-0.050***	0.027	-0.003**
Coefficients	0.016*	-0.016	-0.024***	-0.042	0.006	-0.005	-0.004
Interaction	0.000	-0.003	-0.008***	-0.010**	-0.000	-0.006	0.001
TOTAL	***0.016***	***-0.012***	***-0.029***	***-0.040***	***-0.044***	***-0.016***	***-0.006***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1

Table 3.9.A Decomposition results: "Saving for Business/Children" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR BUSINESS"			
China	0.147	0.147	0.147	0.147	0.147	0.147	0.147
Counterfactual	0.192	0.093	0.043	0.113	0.196	0.266	0.15
DIFFERENCE	***-0.045***	***0.054***	***0.104***	***0.034***	***-0.049***	***-0.119***	***-0.003***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR BUSINESS"			
Endowments	0.001	0.001	0.005***	0.001	0.001	-0.000	-0.000
Coefficients	-0.028***	-0.005	-0.018***	-0.008	-0.01	-0.002	-0.006
Interaction	-0.001	-0.002	-0.006***	-0.002	0.006	-0.002	0.000
TOTAL	***-0.028***	***-0.006***	***-0.019***	***-0.009***	***-0.003***	***-0.004***	***-0.006***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1   Complete decomposition result shown in TABLE 15 in the Appendix.

Table 3.9.B Decomposition results: "Saving for Business/Children" sample above 23 years old, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR BUSINESS"			
China	0.146	0.146	0.146	0.146	0.146	0.146	0.146
Counterfactual	0.203	0.096	0.044	0.118	0.199	0.284	0.155
DIFFERENCE	***-0.057***	***0.050***	***0.102***	***0.028***	***-0.053***	***-0.139***	***-0.009***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR BUSINESS"			
Endowments	0.001	0.000	0.006***	0.001	0.001	0.001	-0.000
Coefficients	-0.032***	-0.008	-0.019***	-0.012*	-0.021	-0.005	-0.009*
Interaction	-0.001	-0.002	-0.008***	-0.003	0.012	-0.005	0.000
TOTAL	***-0.032***	***-0.010***	***-0.021***	***-0.014***	***-0.008***	***-0.009***	***-0.009***
				Observations			
China	3831	3831	3831	3831	3831	3831	3831
Counterfactual	858	865	975	839	801	869	5207

*** p<.01, ** p<.05, * p<.1 Complete decomposition result shown in TABLE 15 in the Appendix.

Table 3.9.C Decomposition results: "Saving for Business/Children" full sample, logit model

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR BUSINESS"			
China	0.147	0.147	0.147	0.147	0.147	0.147	0.147
Counterfactual	0.192	0.093	0.043	0.113	0.196	0.266	0.15
DIFFERENCE	***-0.045***	***0.054***	***0.104***	***0.034***	***-0.049***	***-0.119***	***-0.003***
				DECOMPOSITION – EFFECT OF CHILDREN ON "SAVING FOR BUSINESS"			
Endowments	0.00	0.000	0.062	0.001	0.001	-0.004	-0.000
Coefficients	-0.051	-0.000	-0.012***	0.000	-0.015	-0.001	-0.002
Interaction	-0.001	-0.000	-0.012	-0.002	0.006	-0.001	0.000
TOTAL	***-0.052***	***-0.000***	***0.038***	***-0.001***	***-0.008***	***-0.006***	***-0.002***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.1

Table 3.5 Definitions and descriptive statistics for control variables by country

Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
***businessown**: If the person is a business owner: Yes = 1, No = 0*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.69	0.74	0.67	0.73	0.73	0.83	0.74	0.74	0.54
Std. Dev.	0.46	0.44	0.47	0.44	0.44	0.38	0.44	0.44	0.5
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***income**: Per capita annual income in international dollars*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	6,610	15,724	28,679	18,639	14,699	10,565	15,151	17,257	8,513
Std. Dev.	16,623	16,618	82182	26,007	21,352	23719	12,767	38,952	192,653
Min	0	0	0	532	0	0	398	0	0
Max	482,814	223,496	1,958,872	532,085	376,945	376,324	132,525	1,958,872	72,900,000
***gender**: Participant's gender: Male = 1, Female = 0*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.46	0.48	0.45	0.43	0.50	0.57	0.52	0.49	0.47
Std. Dev.	0.5	0.5	0.5	0.5	0.5	0.49	0.5	0.5	0.5
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***age**: Participant's age*									

Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	47.00	43.84	45.30	56.96	41.80	36.64	50.61	45.81	41.74
Std. Dev.	16.95	17.12	17.94	15.87	17.03	14.02	19.71	18.13	17.88
Min	15	15	15	15	15	15	15	15	15
Max	92	95	90	91	90	82	92	95	99
***help*: If the household sent financial help to others last year: Yes = 1, No = 0**									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.15	0.41	0.29	0.14	0.16	0.30	0.35	0.28	0.22
Std. Dev.	0.35	0.49	0.46	0.35	0.36	0.46	0.48	0.45	0.41
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***urban*: If the participant is from a large city: Yes = 1, No = 0**									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.29	0.64	0.93	0.39	1	0.53	0.84	0.72	0.39
Std. Dev.	0.45	0.48	0.25	0.49	0	0.50	0.37	0.44	0.49
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***sector*: If the participant works in a public sector: Yes = 1, No = 0**									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.38	0.47	0.47	0.16	0.34	0.34	0.47	0.37	0.38

Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
Std. Dev.	0.49	0.50	0.50	0.37	0.48	0.47	0.50	0.48	0.48
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***food**: If the household had enough money for food last 12 months: Yes = 1, No = 0*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.92	0.84	0.93	0.93	0.92	0.75	0.78	0.86	0.68
Std. Dev.	0.27	0.37	0.25	0.25	0.27	0.43	0.42	0.35	0.47
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***shelter**: If the household had enough money for shelter last 12 months: Yes = 1, No = 0*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.85	0.89	0.91	0.94	0.95	0.71	0.73	0.85	0.77
Std. Dev.	0.36	0.32	0.29	0.24	0.22	0.45	0.44	0.35	0.42
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***receivemoney**: If the household received help in the form of money or food in the past 12 months: Yes = 1 No = 0*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.14	0.03	0.13	0.03	0.11	0.18	0.14	0.10	0.21
Std. Dev.	0.34	0.17	0.34	0.17	0.32	0.38	0.35	0.30	0.41
Min	0	0	0	0	0	0	0	0	0

Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
Max	1	1	1	1	1	1	1	1	1
<u>secondary</u>: If the participant completed secondary education (9-15 years of education): Yes = 1, No = 0									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.27	0.46	0.62	0.61	0.59	0.54	0.48	0.55	0.50
Std. Dev.	0.44	0.50	0.49	0.49	0.49	0.50	0.50	0.50	0.50
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
<u>college</u>: If the participant has a college degree: Yes = 1, No = 0									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.07	0.40	0.23	0.26	0.21	0.31	0.38	0.30	0.16
Std. Dev.	0.26	0.49	0.42	0.44	0.4	0.46	0.49	0.46	0.37
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
<u>married</u>: If the participant is married: Yes = 1, No = 0									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.8	0.64	0.56	0.68	0.6	0.58	0.62	0.61	0.52
Std. Dev.	0.4	0.48	0.5	0.47	0.49	0.49	0.48	0.49	0.5
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
<u>borrowed</u>: If the participant borrowed any money during the past 12 months: Yes = 1, No = 0									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688

Statistic	China	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)	World
Mean	0.33	0.29	0.26	0.20	0.20	0.60	0.39	0.32	0.43
Std. Dev.	0.47	0.45	0.44	0.40	0.40	0.49	0.49	0.47	0.49
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1
***welfare**: If the participant received any transfers from the government in the past 12 months: Yes = 1, No = 0*									
Obs.	4,184	1,000	1,007	1,006	1,000	1,000	1,000	6,013	146,688
Mean	0.16	0.11	0.17	0.11	0.28	0.31	0.20	0.20	0.15
Std. Dev.	0.37	0.31	0.37	0.32	0.45	0.46	0.40	0.40	0.35
Min	0	0	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1	1	1

SOURCE: Authors' calculations based on the Gallup World Poll and Global Findex.

Table 3.6 Total decomposition results: "Save" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVE"			
China	0.686	0.686	0.686	0.686	0.686	0.686	0.686
Counterfactual	0.743	0.670	0.732	0.730	0.828	0.743	0.741
DIFFERENCE	***-0.057***	***0.016***	***-0.045***	***-0.044***	***-0.142***	***-0.057***	***-0.054***
				TOTAL DECOMPOSITION			
Endowments	-0.145***	-0.145***	-0.097***	-0.028	-0.120***	-0.074**	-0.087***
Coefficients	0.074***	0.153***	-0.011	0.072***	-0.050**	0.060***	0.050***
Interaction	0.015	0.008	0.063**	-0.088***	0.029	-0.420	-0.017
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.10.

Table 3.7 Total decomposition results: "Meaningful Savings" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "MEANINGFUL SAVINGS"			
China	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Counterfactual	0.505	0.59	0.772	0.523	0.33	0.451	0.529
DIFFERENCE	***-0.044***	***0.13***	***-0.312***	***-0.063***	***0.13***	***-0.009***	***-0.069***
				TOTAL DECOMPOSITION			
Endowments	-0.206***	-0.226***	-0.199***	-0.011	-0.07***	-0.077*	-0.133***
Coefficients	-0.028	-0.097***	-0.306***	-0.045**	0.129***	-0.002	-0.05***
Interaction	0.134***	0.193***	0.193***	-0.005	0.071**	0.088**	0.114***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.10.

Table 3.8 Total decomposition results: “Saving for Old Age” full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – “SAVING FOR OLD AGE”			
China	0.397	0.397	0.397	0.397	0.397	0.397	0.397
Counterfactual	0.457	0.387	0.462	0.485	0.58	0.449	0.47
DIFFERENCE	***-0.059***	***0.01***	***-0.065***	***-0.087***	***-0.182***	***-0.051***	***-0.072***
				TOTAL DECOMPOSITION			
Endowments	-0.068**	-0.062	-0.07**	-0.007	-0.09***	-0.047	-0.05***
Coefficients	-0.01	0.083***	0.028	-0.067***	-0.269***	0.011	-0.037***
Interaction	0.019	0.011	-0.022	-0.013	0.176***	-0.016	0.015
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.10.

Table 3.9 Total decomposition results: "Saving for Education" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
			OVERALL DIFFERENCE – "SAVING FOR EDUCATION"				
China	0.253	0.253	0.253	0.253	0.253	0.253	0.253
Counterfactual	0.354	0.268	0.247	0.323	0.506	0.299	0.332
DIFFERENCE	***-0.101***	***-0.015***	***0.007***	***-0.07***	***-0.253***	***-0.046***	***-0.08***
			TOTAL DECOMPOSITION				
Endowments	-0.047*	-0.038	0.048*	-0.051**	-0.14***	0.009	-0.044***
Coefficients	-0.038**	0.027	-0.07***	-0.036*	-0.107***	-0.046**	-0.045** *
Interaction	-0.015	-0.004	0.028	0.017	-0.005	-0.009	0.009
			Observations				
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.10.

Table 3.10 Total decomposition results: "Saving for Business" full sample, LPM

	Taiwan	Hong Kong	Japan	Singapore	Malaysia	South Korea	All (excl. China)
				OVERALL DIFFERENCE – "SAVING FOR BUSINESS"			
China	0.147	0.147	0.147	0.147	0.147	0.147	0.147
Counterfactual	0.192	0.093	0.043	0.113	0.196	0.266	0.15
DIFFERENCE	***-0.045***	***0.054***	***0.104***	***0.034***	***-0.049***	***-0.119***	***-0.003***
				TOTAL DECOMPOSITION			
Endowments	-0.052**	-0.035	0.003	-0.013	-0.11***	-0.019	-0.051***
Coefficients	-0.023	0.049***	0.066***	0.001	0.02	-0.12***	-0.001
Interaction	0.03	0.039	0.035**	0.046**	0.04*	0.012	0.049***
				Observations			
China	4184	4184	4184	4184	4184	4184	4184
Counterfactual	1000	1007	1006	1000	1000	1000	6013

*** p<.01, ** p<.05, * p<.10.

3.8 References

Ando, A., & Modigliani, F. (1963). The" life cycle" hypothesis of saving: Aggregate implications and tests. *The American Economic Review, 53*(1), 55-84.

2008 年国务院常务会议汇总 (2008). Summary of the 2008 executive meeting of the state council. *Chinese Government Website.* Retrieved from http://www.gov.cn/jrzg/2008-12/09/content_1192649.html

Banerjee, A., Meng, X., & Qian, N. (2010). The life cycle model and household savings: Micro evidence from urban China. *Yale University Working Paper*.

Basten, S., Sobotka, T., Zeman, K., Abassi-Shavazi, M. J., Adsera, A., Van Bavel, J., ... & Zhao, Z. (2014). Future fertility in low fertility countries. *VID Working Paper* No. 05/2013.

Becker, G. S. (1960). An Economic Analysis of Fertility, Demographic and economic change in developed countries: a conference of the Universities. *National Bureau Commitee for Economic Research, 209*.

Becker, G. S., & Lewis, H. G. (1973). On the Interaction between the Quantity and Quality of Children. *Journal of political Economy, 81(2, Part 2)*, S279-S288.

Becker, G. S., & Tomes, N. (1976). Child endowments and the quantity and quality of children. *Journal of political Economy, 84(4, Part 2),* S143-S162.

Binkaia, C., & Rudaib, Y. (2013). Land Supply, Housing Price and Household Saving in Urban China: Evidence from Urban Household Survey. *Economic Research Journal*, *1*.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 8(4), 436-455.

Cai, Y. (2010). China's below-replacement fertility: Government policy or socioeconomic development?. *Population and Development Review, 36*(3), 419-440.

CEIC Data. (2019). China Investment: % of GDP [1952 - 2019] [Data & Charts]. Retrieved from https://www.ceicdata.com/en/indicator/china/investment--nominal-gdp

Chamon, M. D., & Prasad, E. S. (2010). Why are saving rates of urban households in China rising?. *American Economic Journal: Macroeconomics*, *2*(1), 93-130.

Chamon, M., & Prasad, E. (2005). Determinants of household saving in China. *Photocopy. IMF working paper,* 08/145.

Chamon, M., Liu, K., & Prasad, E. (2013). Income uncertainty and household savings in China. *Journal of Development Economics*, *105*, 164-177.

Choukhmane, T., Coeurdacier, N., & Jin, K. (2013). The one-child policy and household savings. *CEPR Discussion Paper No. DP9688.*

Curtis, C. C., Lugauer, S., & Mark, N. C. (2015). Demographic patterns and household saving in China. *American Economic Journal: Macroeconomics*, *7*(2), 58-94.

Feng, J., He, L., & Sato, H. (2011). Public pension and household saving: Evidence from urban China. *Journal of Comparative Economics*, *39*(4), 470-485.

Feng, W., Cai, Y., & Gu, B. (2013). Population, policy, and politics: how will history judge China's one-child policy?. *Population and Development Review, 38*, 115-129.

Ge, S., Yang, D. T., & Zhang, J. (2012). Population policies, demographic structural changes, and the Chinese household saving puzzle. *Institute for the Study of Labor (IZA) Discussion Paper,* 7026.

International Monetary Fund. (2016). The People's Republic of China: Selected Issues. *IMF Staff Country Reports*, *16*(271), 1. doi: 10.5089/9781475524383.002

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.

Lardy, N. R. (2016). China: Toward a consumption-driven growth path. In *Seeking Changes: The Economic Development in Contemporary China* (pp. 85-111).

Li, H., Zhang, J., & Zhu, Y. (2008). The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*, 45(1), 223-243.

Li, S., Whalley, J., & Zhao, X. (2013). Housing price and household savings rates: evidence from China. *Journal of Chinese Economic and Business Studies*, *11*(3), 197-217.

Lugauer, S., Ni, J. and Yin, Z., 2019. Chinese household saving and dependent children: Theory and evidence. *China Economic Review*, *57*, p.101091.

Modigliani, F., & Cao, S. L. (2004). The Chinese saving puzzle and the life-cycle hypothesis. *Journal of economic literature, 42*(1), 145-170.

Modigliani, F. and Brumberg, R.H. (1954) Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. In: Kurihara, K.K., (Ed.), *Post-Keynesian Economics*, 388-436.

National Bureau of Statistics of China. (1955-2015). Household savings rate. *The Chinese Statistical Yearbook.* China Statistics Press.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 693-709.

OECD Data. (2019) Household accounts - Household savings. Retrieved from https://data.oecd.org/hha/household-savings.html

OECD Data. (2019) National income - Saving rate. Retrieved from https://data.oecd.org/natincome/saving-rate.htm

Qian, N. (2009). Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China. *National Bureau of Economic Research Working Paper*. No. w14973.

Qin, X., Zhuang, C. C., & Yang, R. (2017). Does the one-child policy improve children's human capital in urban China? A regression discontinuity design. *Journal of Comparative Economics,* 45(2), 287-303.

Rosenzweig, M. R., & Zhang, J. (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's "one-child" policy. *The Review of Economic Studies,* 76(3), 1149-1174.

Rosenzweig, M., & Zhang, J. (2014). Co-residence, life-cycle savings and inter-generational support in urban China. *National Bureau of Economic Research Working Paper*. No. w20057.

Sinning, M., Hahn, M., & Bauer, T. K. (2008). The Blinder–Oaxaca decomposition for nonlinear regression models. *The Stata Journal,* 8(4), 480-492.

Song, M., & Yang, D. T. (2010). Life cycle earnings and the household saving puzzle in a fast-growing economy. *Chinese University of Hong Kong Working Paper*.

Wang, X., & Wen, Y. (2012). Housing prices and the high Chinese saving rate puzzle. *China Economic Review*, *23*(2), 265-283.

Wang, X., & Zhang, J. (2018). Beyond the Quantity–Quality tradeoff: Population control policy and human capital investment. *Journal of Development Economics, 135*, 222-234.

Wei, S. J., & Zhang, X. (2011). The competitive saving motive: Evidence from rising sex ratios and savings rates in China. *Journal of Political Economy*, *119*(3), 511-564.

Whyte, M. K., Feng, W., & Cai, Y. (2015). Challenging myths about China's one-child policy. *The China Journal, (74),* 144-159

Wolf, M. (2018). Consumption to replace investment as key to China growth | *Financial Times.* Retrieved from https://www.ft.com/content/627ab75c-4256-11e8-97ce-ea0c2bf34a0b

World Bank, World Development Indicators. (2019). Fertility rate, total (births per woman) | Data. Retrieved from https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=CN

World Bank, World Development Indicators. (2019). Gross savings (% of GDP) | Data. Retrieved from https://data.worldbank.org/indicator/NY.GNS.ICTR.ZS

Yun, M. S. (2004). Decomposing differences in the first moment. *Economics letters*, 82(2), 275-280.

Zhang, J. (2017). The evolution of China's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives, 31*(1), 141-60.

Zhou, W. (2014). Brothers, household financial markets and savings rate in China. *Journal of Development Economics*, *111*, 34-47.

Zhu, X., Whalley, J., & Zhao, X. (2014). Intergenerational transfer, human capital and long-term growth in China under the one child policy. *Economic Modelling*, 40, 275

Chapter 4: Forecasting Chinese consumption series with internet search volume data from Baidu and Google

4.1 Introduction

Since Google launched their keyword research and keyword search volume service Google Trends, researchers have tried to utilize this data in analyzing and predicting consumer choice and consumption economic aggregates around the world. For example, Ettredge, Gerdes & Karuga, 2005; Choi & Varian, 2012; Vosen & Schmidt, 2011; Carrière-Swallow & Labbé, 2013; Woo & Owen, 2019; Yu et al., 2019, etc. These studies mostly concluded that the utilization of Google Trends can increase the prediction accuracies. However, as the popularity of different search engines is highly dependent on the region and the time span, it is doubtful if these results on Google Trends can be generalized to other countries where other search engines dominate the market. Indeed, recently researchers have shifted their focus on utilizing other search engines, like Baidu, to analyze if data from other search engines can be used in a similar fashion. This strand of literature has focused on whether Baidu can help predict tourism flows (Yang et al. 2015; Li et al. 2018; Huang, Zhang & Ding, 2017; Sun et al., 2019), stock returns, and stock market volatility (Shen et al., 2017; Fang et al., 2020).

In this thesis, I contribute to this literature by investigating whether internet search volume data can help to forecast macro-level consumption in China by using data from both Baidu and Google. Improvements in the forecasting of macro-level consumption in China mean that policymakers can gain a more accurate insight into what the future path of an economy looks like. This can be useful to both public and private decision-makers.

This thesis focuses on predicting total retail sales of automobile and communication appliances, published by the Chinese Statistical Bureau. Together these two sectors represent a relatively big percentage of total retail sales. If internet search volume data from Baidu and Google contribute positively to predictive performance, this is likely to be evident in these sectors, because people are more likely to engage in pre-shopping research on brands and performance for valuable goods like a car or a mobile phone. In the US context, Choi and Varian (2012) indeed have shown search intensity as measured by Google Trends increases the performance of models predicting US car sales. Similarly, Carrière-Swallow & Labbé (2013) have shown that Google Trends data improves Chilean automobile sales.

In addition, in order to study if my results can be generalized to entire total retail sales in China, I predicted the aggregated retail sales by adopting similar methodologies as predicting sales of the two sectors, using data from Baidu.

Total retail sales between 2011 to 2019 of China are used in this paper. This data has a 1-month delay publication delay. However, search volume data from Baidu and Google are available on a daily basis. These timely data incorporate information not embedded in lagged sales data. This thesis concludes that by adding information from Baidu, the prediction performance of traditional models is improved, this improvement is greater than that from Google Trends or Consumer Confidence index.

4.2 Background and literature review

In 2006, Google started Google Trends at www.google.com/trends/, and Baidu started reporting internet search volume data at http://index.baidu.com/. For many years Google has been the most popular search engine worldwide, while Baidu, its Chinese counterpart, has been the most successful search engine in China. Together these two search engines provide valuable information on people's internet usage and serve as valuable tools for scholars worldwide in various disciplines.

Internet searches can be linked directly with consumers' shopping patterns. Since the internet has grown in popularity in many countries worldwide, more and more people have been conducting pre-shopping research online, especially through search engines.

This research is mainly related to two strands of research, the first one being how researchers can forecast and predict private consumption, and the second strand is how internet search volumes can be used to improve various forecasts.

4.2.1 Forecasting private consumption

Consumption plays a crucial role in determining a country's economic performance. For many countries, private consumption accounts for the majority of their GDP. Therefore, it's a driving force of their economic growth. Individual consumption decisions affect aggregate consumption demand, which in turn shapes the business cycle and the short-term monetary policy. Fiscal policies from the governments also affect demands through changes in taxation, interest rates, and infrastructures, which impact a country's employment, income, and in turn, impacts consumer's spending and investment. Especially since the COVID-19 pandemic hit the world, governments worldwide have been trying to battle the spread of the virus while stimulating consumption. Therefore, making timely and accurate forecasts in consumption is very important in accessing the performance of the economy. Indeed, in recent decades many studies have focused on forecasting private consumption in different contexts.

Some of the typical indexes that are being used to forecast consumption are survey-based indicators like consumer confidence index and consumer sediment index. Various articles have investigated how consumer survey indicators can predict private spending, and whether consumer survey indicators can help nowcasting and forecasting private consumption, the conclusion of these articles mostly conclude that by augmenting the baseline models with survey-based indicators, the nowcasting and forecasting performance of the models is increased, although the magnitude of this increase is still debatable.

One of the most widely cited papers in this strand of literature is Carroll, Fuhrer & Wilcox (1994). They provided evidence that lagged consumer sentiment has explanatory power for changes in current household spending. To be specific, they found that lagged values of the consumer sentiment index can explain about 14 percent of the variations in the growth of total real personal consumption. Similar results are confirmed by Bram & Ludvigson (1998) and Howrey (2001), who concluded that by using survey-based indicators, the forecasting error for private consumption lowered.

While many of the initial studies focused on the US, whether survey-based indicators can forecast household spending has also been investigated for other countries. Cotsomitis & Kwan (2006) used a multi-country framework to examine the ability of consumer confidence to forecast household spending using data between 1980 and 2002, their sample includes 9 European countries, and their results suggested that survey-based indexes only provide limited out of sample predictive capability in Europe. Kwan & Cotsomitis (2007) used the Conference Board of Canada's index of Consumer Attitudes (ICA) to examine whether this index can help predict household spending in the Canadian case. They conducted their research on both the national level and state level of Canada, examining both the prediction outcome of the entire nation as well as several states independently. They concluded that ICA can help predict personal consumption as well as several sub-categories under personal consumption expenditures even after controlling for macroeconomic variables. In addition, the information that ICA contains about household spending seems to be more prominent at the national level compared to the state level.

Dees & Brinca (2013) used seasonally adjusted quarterly data between 1985 and 2010 for Europe and the United States to study the link between consumer confidence and household consumption. After controlling for income and savings, as well as other macroeconomic variables, their results suggested that consumer confidence can in certain cases, be a good predictor of consumption, especially when consumer confidence experience large volatility, like periods of the Global Financial Crisis.

Lahiri, Monokroussos & Zhao (2015) re-examined the existing models on forecasting household consumption by re-estimating the models commonly used in past literature (e.g., Carroll, Fuhrer & Wilcox (1994); Bram & Ludvigson (1998)), they argued that using quarterly data may mask important information contained by the monthly data, so they re-estimated existing models at both quarterly and monthly frequency. By looking at consumption in different sectors, they used a rich dataset to study the role of consumer confidence in predicting private consumption and found that the Consumer Confidence Index contains more information in predicting the service sector, and that the predicting power of consumer confidence is higher during the 2007–2009 global recession.

Recent literature on this area has shifted towards empirically examining how consumer confidence can produce better predictions on household consumption in other countries around the world. Gausden & Hasan (2018) studied how consumer confidence can better the forecasts for household consumption in the UK, and found that models incorporating consumer confidence produce better forecasts. Juhro & Lyke (2020) studied how consumer confidence can help forecast private consumption in the Indonesia case, and concluded that policymakers can increase their forecast accuracy of household consumption by 4 to 13 percentage points by incorporating consumer sentiment and business sentiment into their forecasting models.

Overall, this strand of literature showed a consistent conclusion, that by incorporating survey-based indicators like consumer confidence or consumer sentiment, the forecasting accuracy of private consumption is improved. In the empirical analysis of this paper, I will also include Consumer Confidence Index as a predictor to compare the predicting ability between internet search volume data with consumer confidence.

4.2.2 The use of internet search data

In terms of the second strand of literature that involves the usage of internet search volume data, the papers in this field have been gaining popularity in recent years. A comprehensive paper by Jun et al. (2018) studied the research trends, and utilization of data from Google Trends, their network analysis of 657 papers that used data from Google Trends showed that Google Trends is gaining popularity recently in a wide range of disciplines like IT, communications, medicine, health as well as economics. The authors also indicated that the focus of this research has gradually shifted from narrating and monitoring trends, to forecasting and predictions. Although the majority of the papers used Google Trends and focused on the prediction of the US, only limited papers used Baidu Index or focused on China.

The first paper that suggested using internet search volume data like Google Trends can be useful when making an economic forecast is Ettredge, Gerdes & Karuga (2005), who suggested that internet search volume data can be associated with predicting the unemployment rate. Although due to the short time span of their data, they didn't actually produce any unemployment predictions, they still undertook preliminary research and found an association between unemployment-related searches and the unemployment rate. This link between internet search data and unemployment is later examined by Askitas & Zimmermann (2009), D'Amuri & Marcucci (2017), Fondeur & Karame (2013), Naccarato et al. (2018), Mihaela (2020). These papers all showed an association between internet search data and the unemployment rate and showed that by incorporating data from Google Trends, predictions of the unemployment rate are improved. As one of the key macro-economic variables in economic planning, unemployment is an important indicator to determine the situation of an economy when monetary policies are being made. The fact that the above studies concluded Google Trend is able to better prediction of unemployment, means that policy makers are able to make better decisions when determining out fiscal policies with the help of internet search queries data.

Apart from unemployment, several papers have demonstrated the use of Google Trends in improving other aspects of various forecasts since 2010. Choi & Varian (2012) is one of the first papers to show that by using Google Trends, forecasts of various economic indicators can be improved. They used four examples to demonstrate the improvement of the forecasts while using different categories of Google Trends. The examples include motor vehicle and parts sales, unemployment benefit claims, tourism, and consumer confidence. The results of their examples indicated that all predictions of these four economic indicators are improved after incorporating Google Trends into their model. Goel et al. (2010) showed that consumer behavior can be predicted by what consumers are searching for. They also used several examples, including box-office revenue for feature films, sales of video games, the rank songs on the Billboard chart as well as tracking flu trends. They found that generally, the performance of the baseline models is improved after incorporating Google Trends data, although this increase may vary depending on the question.

Vosen & Schmidt (2011) compared the nowcasting and forecasting performance of models incorporating Google Trends and survey-based indicators. They produced predictions for monthly consumption in the US, and found that models incorporating internet search data outperformed the models using Survey Based Indicators. Their work is later extended by Woo & Owen (2018). Woo & Owen (2018) treated survey-based indicators as complementary rather than substitutes to Google Trends data, and they further tested the forecasts for durable goods, non-durable goods, and services consumption separately, they also examined the incorporating of news related Google Trends data and consumption related Google Trends data separately. They found evidence that Google Trends data increases the accuracy of the predictions in all sectors of consumption, although the magnitude of this increase is subject to model specification.

Carrière-Swallow & Labbé (2013) tested whether by incorporating Google Trends, the nowcasts for Chilian automobile sales are improved. They first introduced an index of automobile purchase using Google Trends, and then showed that models incorporating their Google Trends automobile index improved the nowcasts results.

Guzman (2011) compared how different inflation expectation indexes (including Google data) can help forecast inflation, and found that models with Google data produced the lowest forecast error amongst all the inflation expectation indexes tested.

Some researchers have demonstrated other interesting applications of Google Trends data. Mavragani & Tsagarakis (2016) predicted the 2015 Greece referendum results by using data from Google Trends. Specifically, they found that Google search activity regarding the referendum was rising, and they studied the data from Google Trends regarding people's search hits concerning “Yes” and “No” on the referendum. They consistently find that the number of “No” hits is higher compared to the “Yes” hits. Therefore, the prediction by Google Trends indicates the final voting results for the referendum will be “No”, and this is consistent with the actual referendum results, with 61.31% of the voting population choosing “No” and 38.69% of the people choosing “Yes”. Dos-Santos(2018) Studied how Google Trends data can shed some light on the adaptation and future development of an innovative agriculture system that combines fish farming with vegetable farming called “aquaponics”, and found that there’s been an increase in the popularity of both aquaponics and related search terms, they further proposed that aquaponics firms should seize the opportunity to speed up development, and that public decision-makers should be more concerned about allocating funds and investment in the development of aquaponics.

Some of the more recent applications on predictions with Google Trends include Wu & Brynjolfsson (2015), who showed how Google data could be used to predict housing market sales and prices in the US. They found that their predictions based on models incorporating Google data beat the predictions from experts from the National Association of Realtors. Önder (2017) showed that internet search data could be used to forecast tourism in 2 cities (Vienna and Barcelona) and two countries (Austria and Belgium). Park, Lee & Song (2017) showed that the augmented models incorporating Google Trends could improve the forecasts on the inflow of Japanese tourists to Japan. Castelnuovo & Tran (2017) produced uncertainty indexes for the US and Australia. And Bulut(2017) concluded that the incorporation of Google Trends helps with predicting the direction of exchange rates using data of 11 OECD countries.

Compared to Google Trends, Baidu Index has been receiving far less attention from researchers. Amongst the papers that did incorporate Baidu Index in their analysis, the majority of them focused on predicting the tourism flow of famous travel destinations in China, some empirically tested whether Baidu Index helps to make predictions in the Chinese stock market. Most of these papers found that when incorporating Baidu Index in their models, the forecasting accuracy is increased.

The following papers all studied the predictive power of Baidu Index on forecasting tourism flow. Huang, Zhang & Ding (2017) used Baidu Index to forecast tourism flow to the Forbidden City in Beijing. The authors collected data based on a set of keywords and found that models incorporating the keywords from Baidu Index created more accurate forecasts.

Li et al. (2018) reviewed the existing literature on predicting tourism in China and found that one of the major problems in this area is that the dataset is sometimes too massive and data on tourism is usually highly correlated. They further proposed a framework to create a composite index and tested their index against traditional time series models and models incorporating principal component analysis. They found that the models using their composite tend to outperform the other models.

The above papers only looked at the predictive power of Baidu Index, while none of them used Google Trends. Amongst the papers that looked at tourism flow, Yang et al. (2015) is one of the papers that compared the predictive power of both Baidu Index and Google Trends. They forecasted the tourist volume of Hainan province, a province in the South of China. They found that forecasts produced models augmented by both search engines produced lower forecasting errors, while Baidu performed better due to its large market share in China. Sun et al. (2019) also used both Baidu Index and Google Trends data to forecast the tourism flow in Beijing, the biggest city in China. Instead of treating Baidu Index and Google Trends data as substitutes, they treated both the dataset as complements by incorporating both of the datasets in 1 model, and compared the performance of this model with traditional time series models and models incorporating only Baidu Index or Google Trends, and found that the model using both the datasets performs best. Liu et al. (2018) connected internet search queries from Baidu Index, weather, temperature, and holidays with tourism destination arrivals, and studied the reciprocal predictive power of these factors upon each other using a VAR model. Their estimation results showed that Baidu Index search terms positively impact tickets sold to the tourism destination, holidays also impact ticket sales positively, while both temperature and weather don't affect ticket sales. The authors also mentioned that they used Baidu Index instead of Google Trends because 95% of the ticket sales are to domestic travelers who are from Mainland China, and Baidu has a bigger market share than Google in China.

Apart from tourism flows, Baidu Index is also used by researchers to make forecasts about the stock market. Shen et al. (2017) empirically tested the predictability of the Chinese stock returns and found that the search volume of Baidu index can indeed be used to predict stock returns. Specifically, they found that stock prices go up when less attention is paid to the stocks while stock prices go down when investors put more attention on the stocks. Fang et al. (2020) used a GARCH model to investigate the predictive power of Baidu Index in forecasting the return volatility of the Chinese stock market, and found that the model incorporating Baidu Index produced a more accurate forecast for volatility in the Chinese stock market.

To the best of my knowledge, only one paper looked at whether Baidu Index can help predict consumption or sales in China, which is surprising because, as I showed earlier in the literature review, many researchers looked at how Google can help with consumption and sales-related nowcasts and forecasts. Fang et al. (2019) looked at how Baidu Index can be used to nowcast mobile phone sales in China. Specifically, what they did in this paper is that they used several keywords related to the phone model "Huawei Mate7", to predict the sales of this phone model. Their baseline model is an autoregressive model that incorporates only the AR(1) term with no additional variables. They then augmented the baseline model with two keywords from Baidu Index. “Mate 7” and “Huawei”, and made predictions with the models. They concluded that phone sales correlated highly with the keyword “Mate7”, and found that models incorporating Baidu Index data performed better nowcasts than the baseline model where there’s no Baidu Index.

The lack of attention being put on using Baidu Index to forecast consumption and sales in China is one of the major motivations behind this thesis. As the existing literature suggests, only one paper so far has focused on utilizing Baidu Index to predict Chinese consumption and sales, and this thesis focuses more on predicting sales of an individual model while using limited keywords. In this paper, I make predictions on aggregated sales in China while using more keywords to cover a bigger perspective. The aggregate sectoral sales I predict in this paper is an important component of consumption in China, timely and accurate predictions mean that firms can set their inventories accordingly, and it’ll also better facilitate their decisions to operate in the future. Government and policy makers can also benefit from a better prediction by gaining insight into what the future of consumption might be to make better fiscal decisions.

4.3 Comparison of Baidu and Google

The main function of Google Trends and Baidu Index is to reflect the search volume of user's queries in Google and Baidu, and they are useful sources for data mining. One can access these two data sources at www.google.com/trends/ and http://index.baidu.com/. The data provided by Google Trends dates back to Jan. 2004, while the data from Baidu goes back to Jun. 2006.

Vaughan and Chen (2012) wrote an in-depth comparison between Google Trends and Baidu Index. In this thesis, I update their table and included additional comparisons between Google Trends and Baidu Index. The detailed differences are listed in TABLE 4.1.

4.3.1 Userbase

FIGURE 4.1 and 4.2 show the market share of several search engines worldwide and in China between Jan. 2010 to July. 2020. Figure 1 clearly shows that Google is the search engine with the largest market share worldwide, and its market share has been very stable during the past decade. As of July 2020, Google occupies 92.17 percentage points of the market share, largely exceeds the market share of other search engines like Bing (2.78%) and Yahoo! (1.6%), Globally Baidu takes up 0.92 percentage points of the market share in search engines.

China tells a different story. The search engine market share in China demonstrated a more volatile pattern over the years. As shown in FIGURE 4.2, over the past decade, Baidu’s market share in China has ranged between 50 and 80%. Currently, it stands at around 70%. Baidu's biggest rival before 2013 was Google. From 2010 to 2012, Google had about 40 percent of the Chinese internet search market. During this period, Google transferred service out of mainland China due to a major hack of the company's servers and a dispute over censorship with the Chinese government. Accordingly, they redirected search queries from Google China to Google Hong Kong. However, in 2014, Google China became unavailable to mainland China users. This is clearly seen in FIGURE 2. There was a slow decrease in Google's Chinese market share in China before 2014, and almost no market share afterward. Although Google is no longer available, people from mainland China are still able to access Google by using Virtual Private Networks (VPNs). After Google's exit of the Chinese market, several other search engines began to claim a non-negligible market share, but they have only been popular for a short period of time.

4.3.2 Services

Both Baidu and Google provide information on search volume that its users entered into the search engines, they provide services and options similar in many ways, and they also provide some unique services that can be very useful. In Table 4.1, I compare the services provided by Google Trends and Baidu Index, this table is based on Vaughan and Chen (2015).

Firstly, I updated their table and then extended the comparison to several other aspects that were not included in their paper, I divided the table into the top half and the bottom half panel. In the top panel, I showed the updated results of their original table, while in the bottom panel, I listed some additional comparisons.

A distinct difference between Baidu Index and Google Trends is that Google reports relative volume for a sample of Google searches. Baidu Index reports absolute volume for its whole population of searches. According to Google Trends, they first take a sample of the absolute search volumes. They then normalize the sample by dividing the number of searches by the total search volume for the location and time under consideration. The results are scaled to a range of 0 to 100, with 0 being the lowest and 100 being the highest relative search intensity value. The fact that Baidu reports absolute search volumes is important as this makes it possible to add the search volumes of various keywords, something that is not possible with Google Trends. For simplicity, in the following paragraph, we will refer to both data from Google Trends and Baidu Index as search volume data, although only Baidu Index presents absolute search volumes.

It is important to note that, each time one looks up a search term on Google Trends, they are likely to get slightly different results. This is due to the fact that billions of searches are conducted with Google each day, which makes it difficult to access the entire dataset due to the size of the Google Trends data. As a result, Google Trends reports a small sample of the actual search volume. Although this sample is generally representative of the amount of Google searches conducted. There are still some cases when this feature makes it difficult to yield consistent numbers. This is documented by Medeiros and Pires (2021). However, this seems to only be a problem for less popular search terms, as Medeiros and Pires (2021) also showed that popular terms do not vary much among different samples. Most of the search terms included in this chapter of my thesis are popular car and phone models, which means that the sample inconsistency problem should not affect my results by much. In addition, Medeiros and Pires (2021) proposed that one can solve this problem by recording Google Trends data again and again for the same period, and then taking the average of different Google Trends samples to smooth out this inconsistency in Google data. This is indeed a great way to solve the problem. However, as the main objective of this chapter is to investigate Baidu Index, I do not attempt to focus on Google too much.

Both Google Trends and Baidu Index allow you to limit your search volume data to only a specific region within a country, allowing the limitation to a specific time period. This time period is Jan 2004 for Google Trends and June 2006 for Baidu Index. However, if you wish to include the search volume conducted via phone On Baidu Index, the data only goes back to Dec 2010. Both Google Trends and Baidu Index allow the comparison between keywords, at most you can compare the search volume pattern of 5 keywords on both sites. The maximum number of keywords allows for comparison was 3 for Baidu Index, according to Vaughan and Chen (2015), but when I conducted tests in 2020, this number is now 5. Both the data sources also provide information on average search volumes. Google only provides information on average search volumes when there's a comparison between keywords, while Baidu Index provides both average and daily moving average, as well as year on year and month to month growth rates of a certain search term across the sample period. The total search volume of several search terms is also available for both Google Trends and Baidu Index. In addition, both these services also provide an extensive analysis of related searches that were conducted by people who searched for a specific keyword.

Despite having so many similarities between the services provided by both of these search engines, there are some differences as well as unique options in both Google Trends and Baidu Index. As I mentioned earlier, Baidu is only popular in China and it doesn’t have a big userbase elsewhere. As a result, Baidu only provides search volume data for China. Another major difference between Google and Baidu is that Google Trends allows its users to limit search volume to specific categories. For example, one has the option when collecting search data on Apple to limit the collection the search volumes to reflect queries for Apple, the technology company, as opposed to Apple the fruit. This is very helpful when a specific keyword can represent many different objects. Meanwhile, with Baidu, searches of “Apple” will produce data for both the fruit and the company.[3]

Baidu Index also provides some useful features which are not included in Google Trends. One of these features is the extensive demographic information. Baidu provides a service that translates to "Portrait of the crowd". This feature presents the demographic behind the searches of a certain keyword, including the distribution characteristics of users’ age, gender,

3 Vaughan & Chen (2015) explain the detailed matching mechanism difference between Google Trends and Baidu Index.

region, and what are the interests of the people who searched for this keyword. Google also provides some demographic information of its users behind each keyword, but only information on the geographical distribution of the users is presented. Another option in Baidu Index is that it separates searches conducted on different platforms. Specifically, it distinguishes the search volumes conducted on PCs and Phones, although if you wish to see the search volume conducted on phones, the data only goes back to Dec 2010, while the search volume for PCs goes back to 2004.

Google Trends and Baidu Index is incorporate news-related information differently. Google Trends allows users to look at search volumes for news, pictures, shopping, and YouTube separately, and each of these subcategories may contain different information. For example, news-related Google Trends search volume may have more insight on people's news readership, while shopping-related Google Trends search volume may contain information on people's consumption patterns. Baidu Index, on the other hand, reports what is called the "Information Index" and "Media Index". According to the Baidu Index help book, "Information Index" is calculated by weighting the sum of the numbers of netizens' reading, commenting, forwarding, liking, and disliking, which represents the level of attention and concern people put on a specific keyword. While the "Media Index" is calculated based on the number of news articles reported by major internet media related to a specific keyword, this Index reflects how much new coverage is related to a specific keyword. In addition, although this is information is provided by Baidu Index, “Information Index” and “Media Index” are not directly related to the search volumes in Baidu Index, they are merely Indexes compiled by Baidu to reflect the attention and news coverage of a specific keyword is getting to help users gain more insight into a keyword.

Given that the main purpose of this paper is to look at the predictive power of internet search volumes in forecasting consumption in China, I will use both Baidu Index and Google Trends in the following analysis. Baidu Index is expected to contain more information on consumer's buying patterns mainly due to the fact that Baidu is the most popular search engine for most of the period studied here.

4.4 Data and Methodology

4.4.1 Data

In this thesis, I study whether internet search volumes will improve both the nowcasting and forecasting performance of total retail sales of consumer goods in China., it is published on a monthly basis by the Chinese Statistics Bureau. Total retail sales of consumer goods is the total amount of consumer goods sold directly to urban and rural residents and social groups in various sectors of the national economy. It is an important indicator for studying the changes in the domestic retail market and reflecting the degree of economic prosperity. The Chinese Statistical Bureau publishes reports, usually monthly, detailing the consumption pattern and the status of the Chinese market based on this data. The reports are published at www.stats.gov.cn. The Chinese State Information Center makes predictions of this data, which they use to generate reports of the macro-economic situation. However, to the best of my knowledge, the detailed month-by-month predictions are not available.

To facilitate data collection, the Chinese Statistics Bureau divides total Retail Sales of Consumer Goods into two categories: big enterprises above the quota (a certain size) and small businesses under the quota. Enterprises above the quota directly report their sales information to the statistical bureau, while the sampling survey method is used to collect sales data for small businesses. Adding up the above two parts of retail sales gives the total retail sales of consumer goods.

While sales data for small businesses are not further classified into different categories, the sales data of Enterprises above the quota is classified into the following categories:

Grain and Oil
Garment, footwear, hats, knitwear
Cosmetics
Gold, Silver, Jewelry
Commodities
Sports and recreation
Newspapers and magazines
Communication appliances
Medicine
Cultural and office appliances
Furniture

Household appliances and video equipment
Petroleum and related products
Automobiles
Building and decoration materials

In this study, I aim to forecast the sales of two sectors: the automobile sector, and the communication appliances sector. The reason I chose these two sectors are as follows. Firstly, together these two categories account for a relatively big share of the total sales, which is about 11% in 2019. Secondly, the automobile and communication appliances sectors are dominated by big brands and are difficult for small businesses to gain access to, which means that the sales data of big enterprises above the quota is more likely to reflect the sales of these industries as a whole. Thirdly, because a relatively small number of big brands dominate the market, keywords related to these brands are likely to be able to represent the entire sector. Fourthly, similar sales figures are also used by Choi & Varian (2012), an influential early study focusing on the use of Google.

FIGURES 3 and 4 show the natural logarithm of the sales of the automobile sector and the communication appliance sector. As the figure suggests, the sales in both sectors have been demonstrating strong growth in the past decade, this growth has been slowly moderating, and it's more stabilized in recent years. In addition, there is clear seasonality in both sectors, the sales are the highest in November and December period, while the sales in January and February tend to be low. Both month dummies and quadratic time trends are included in my nowcasting and forecasting models to capture this seasonality and time trend.

Auto regressive (AR) term is also included in the model. The degree of AR term is dependent on when the data are published and available. The sales data on published on a monthly basis with a lag. Typically, in the middle of the current month, sales data of last month are published. This means that if one wants to nowcast sales data at the end of a given month, say August, one can only use sales data from July. Baidu Index and Google Trends make it possible to use data from August. This makes it possible to produce better predictions.

4.4.2 Collection of search query data

The keywords used in this thesis are based on brands and models of automobile and communication appliances. I use these keywords, and the combination of these keywords to obtain the search volume data series from Baidu Index. For example, in the keywords for communication appliances, I included “Huawei” “Mate10” as well as “Huawei Mate10” as search terms (Both English and Chinese languages are used as keywords). Many of these keywords don't form a valid search term when combined together either because the combination is not being searched for or because Baidu (or Google) didn’t record any data for this search term[4]. In addition, some keywords that are associated with buying a new car or a new phone are also included in the search terms, for example: “Car insurance” or “Phone cases”.

When constructing the list keywords, I used brands and models of automobiles and phones from www.autohome.com.cn and www.zol.com.cn. These two websites are widely used in China, and they contain detailed information on brands and models of automobiles and phones that are being sold in China. Other keywords like “car insurance” or “phone cases” are mostly chosen by suggested searches and related searches. The full list of keywords is available on Dataverse[5]. Overall, I included 470 search terms for automobile sales and 727 search terms for communication appliances from Baidu Index.

An additional problem with these search terms is that some of them are only densely searched at a short period, with almost no searches done outside this peak period. This is not surprising because, in reality, phone models or car models are often most popular around a certain time. For example, when a new phone model is being released or sold, more people will do research on it. FIGURE 5 shows the search volume for different models of iPhones between 2011 and 2019. Most of the search volumes for each keyword are highly concentrated around a certain time, and the search volumes before or after this specified time are small. If I run the model with each of these search terms separately, this won't be very useful because each of these search terms only provides information for prediction for a short amount of time. Even if they do have a correlation with sales and belong to the model, this correlation would've been washed away by the small search volumes around other time periods. This problem only seems to affect search terms for communication appliances since each new model under a product line has different names. To solve this problem, I added all

4 For Google, the warning is mostly that there's not enough data, for Baidu, the warning is "Keyword “XXX” is not included or recorded by Baidu Index”, followed by an option to purchase a keyword. Baidu will then start to record it.

5 See the following link for the full list of keywords, https://doi.org/10.7910/DVN/YT25IP, Harvard Dataverse.

the search term for each series of a product together to create a variable that has a long-lasting effect, (for example, Huawei produce several series of phones and they present a new model under this series each year, like the Nova series and the Mate series, I added up all the searches for each model under a series separately as a single variable, in the end, I have one variable for the Nova series and one variable for the Mate series. As a result, I aggregated the keywords into 86 Baidu Index variables for communication appliances.

I will estimate both the baseline models and the models augmented with Baidu search term series, using both OLS and Lasso methodologies, searching over various specifications to find the model that gives the most accurate nowcasts and forecasts.

4.4.3 Baseline models

Both nowcasts and 1-month ahead forecasts of automobile and communication appliance sales in China are produced in this thesis. The difference between nowcasting and forecasting is that nowcasting aims to predict the value for August at the end of August. Forecasting aims to predict the value for August at the beginning of August, when the most recently available data are for June.

To see this, the baseline model for Nowcast and 1-month ahead Forecast are:

Nowcast:

$$C_t = \alpha C_{t-1} + \beta_1 Trend + \beta_2 Trend^2 + \beta_{3\ to\ 12} Month\ Dummies \qquad (4.1)$$

Forecast:

$$C_t = \alpha C_{t-2} + \beta_1 Trend + \beta_2 Trend^2 + \beta_{3-12} Month\ Dummies \qquad (4.2)$$

where C_t is the natural logarithm of total sales for automobile or communication appliances, in real terms[6] at time t; and $Trend$ is the time trend. Note that t stands for the end of each period. I also include ten monthly dummies to account for seasonality. The reason why only 10-month dummies are included is that although automobile and communication appliance consumption are reported monthly, the values for January and February are in most years combined as one, in order to extract as much information as possible without making too

6 The sales for automobile and communication appliances are deflated using the prices index available at the National Bureau of Statistics. The link is: https://data.stats.gov.cn/english/easyquery.htm?cn=A01. Logs of sales are also used in Choi and Varian (2012).

many assumptions about the data, I divided the values for January and February by 2 and used this value as a separate month and treated each year as 11 months.

Expanding window nowcasts and forecasts are being calculated using the baseline models above, where an additional observation is included to train the model as the iteration move from one time period to the next. For example, when one makes a nowcast for the time period *t*, one uses data before *t* to train the model, but when one makes a prediction for *t+1*, data for *t* is added into the training period to train the model. In the first model, I use data between January 2011 to December 2017 to train the model, while predicting the retail sales of January 2018.

To measure the performance of the models, RMSFE (Root Mean Square Forecasting Error) is calculated each time after I run the expanding window nowcasts and forecasts. RMSFE is the standard deviation of the prediction errors, and it's calculated as follows:

$$RMSFE = \sqrt{(f-o)^2} \tag{4.3}$$

where *f* is the prediction, and *o* is the observed value.

To investigate the added value of including information from Baidu, I will augment these baseline models with search query data. I will discuss how are the search query data are incorporated into the models in the next section.

4.4.4 OLS estimations

The equations below show the Baidu Index series augmented nowcasting model and forecasting models.

Nowcast:

$$C_t = \alpha C_{t-1} + \beta_1 Date + \beta_2 Date^2 + \beta_{3\ to\ 12} Month\ Dummies + \beta_{13} Baidu_t \tag{4.4}$$

Forecast:

$$C_t = \alpha C_{t-2} + \beta_1 Date + \beta_2 Date^2 + \beta_{3-12} Month\ Dummies + \beta_{13} Baidu_{t-1} \tag{4.5}$$

In the models above, the lags of C_t, $Date$, $Date^2$, and $Month\ Dummies$ is the same as the baseline model in Equations 4.1 and 4.2, while $Baidu_t$ stands for the different specifications of Baidu Index that are incorporated into the models.

Because of the large quantity of Baidu search terms, OLS models do not have enough degrees of freedom to estimate equations 3 and 4 when all of the search terms are included separately. For example, for the automobile sector, in my total sample, there are 470 Baidu Index keywords but only around 100 observations to train the model, so when all 470 search terms are included separately, I would indeed have more explanatory variables than observations.

I adopted several methodologies to select variables and limit the number of keywords.

First, due to the fact that Baidu Index uses absolute numbers to record search volume, one can add up all the search volume series, although this method is due to lose some of the information embedded in Baidu Index.

Second, a screening process is adopted to pre-select keyword series with a correlation coefficient with retail sales that is bigger than a certain number (This number is usually 0.8, but it can be 0.7 or 0.6 depending on the model specification, for the model to be calibrated successfully), and then the principal component analysis is used to convert the selected Baidu Index series into factor loadings, then these factor loadings are used to augment the baseline models.

Third, I follow a procedure similar to Ginsberg et al. (2009) and run a regression with each Baidu series separately and find the series that individually adds most to the baseline model during the training period. To illustrate, I run the following OLS model:

$$\begin{aligned} C_t \ &= \ \alpha C_{t-1} + \beta_1 Date + \beta_2 Date^2 + \beta_{3\ to\ 12} Month\ Dummies \\ &\quad + \beta_{13} Single\ Baidu\ or\ Google\ Keywords \qquad (4.6) \end{aligned}$$

where each of the individual keywords is included in the model along with the baseline variables. Then the series of the keyword that gives the highest adjusted R square in the training sample is selected, and add this series to the baseline model to nowcast and forecast the retail sales of the next period. I iterate this procedure using the expanding window method, re-selecting at each stage the keyword with the highest adjusted R square for each time period. In addition, I use this method to choose the 3 keywords that, individually, gives the highest adjusted R square, and then evaluate the RMSFE of a model that includes these 3 terms together. The reason that I included this is that I want to try different specifications to see how I can best use the Google Trends series. I included the best 1 series and 3 series to show that sometimes a limited number of keywords may be able to produce the best forecasting results. I also tried other combinations of the number of series to include to see if the results are different, the results are not included in this thesis for simplicity.

In addition to using only the contemporaneous values of the Baidu Index series, in this thesis, I also try experimenting with models that incorporate lagged series of Baidu Index. in my predictions, up to 3 lags for the nowcasting models and up to 4 lags for the forecasting models are incorporated. The aim of adding different lags of Internet search series is to explore how historical search data can potentially be used to increase nowcasting and forecasting accuracy. I only include results up to 3 lags for nowcasting and 4 lags for forecasting to show how the historical data are being utilized in my thesis, more lags can potentially be included as well, but the results of additional lags are not included for simplicity.

4.4.5 Lasso estimations

In this thesis, I also use Lasso models to run expanding window predictions. Lasso models do not suffer from the "large number of explanatory variable problem" I described above for OLS. Lasso is useful when predicting the value of the outcome variables when the number of regressors is large relative to the number of observations in the dataset (Tibshirani (1996)). Lasso is a method popular in model selection and prediction, and the term Lasso is an acronym for "least absolute shrinkage and selection operator."

For example, a regression model with multiple regressors may take the form of:

$$Y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \qquad (4.7)$$

Lasso set a penalty for the sum of the absolute values of the coefficients, to find a solution for the above model while keeping it less complicated. Specifically, it minimizes:

$$\frac{1}{2N}\left(y - x\beta'\right)'\left(y - x\beta'\right) + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \qquad (4.8)$$

The first term of equation 4.8 represents the same value that OLS minimizes, which is the in-sample prediction error. The second term of equation 4.6 is a penalty term that is controlled by the operator λ, and this term increases as more variables are included in the model and the model gets more complex. λ also causes Lasso to omit variables because although none of the coefficients are likely to be exactly 0, the penalty operator λ drives the small coefficients to 0 as equation 4.6 gets minimized when the model is being estimated. The complexity of the model is set by λ, specifically, the larger the λ, the less complex the model will be, while when $\lambda = 0$, the model is the most complex.

To estimate the Lasso models, I use Stata's built-in "Lasso" command, which allows various λ selection methods (CV selection, Adaptive Lasso and plugin formula), grid settings for CV selection and adaptive Lasso, fold of the selection, etc. I tried multiple settings and model specifications and compared the performance of multiple Lasso models. In this study, I will only present the model using CV selection to decide the penalty operator λ because CV selection selects the λ that gives the lowest root mean square forecasting errors.

To facilitate fair comparison of the OLS and Lasso methods, I first run the same specifications as I did for OLS. Firstly, models that include the sum of the Baidu Index series are estimated, then estimate a model that includes the principal components, and finally, the model which includes all Baidu series are estimated.[7]

As Lasso provides a way for model selection when a large amount of regressors is included in the model, I also experiment with models that add both lagged terms of the Baidu series and interaction terms into the model. The interactions included are the interactions of the baseline variable and Baidu Index search terms.[8]

4.4.6 Predictions using Google Trends

In this thesis, I will not only use Baidu Index to predict retail sales in China, search intensity data from Google Trends are also incorporated to evaluate the predictive performance of the models.

Past literature that focuses on predicting consumption-related aggregates with the use of internet search volume data only used Google Trends. Although intuitively, Google Trends will not be as useful as Baidu Index in China because Google is no longer widely used in China since Google quit the Chinese market in 2014. However, for completeness, I also check

7 As a robustness check, we also run the Lasso models with the top 1 and top 3 variables to see how the results of these models compare to the other models. The results of these models don't change the results presented so far. These results as well as the code that produce them can be found on Dataverse at https://doi.org/10.7910/DVN/YT25IP

8 Because up to 3 lags of Baidu Terms will be included in the model, to keep my data comparable, in all of the models in the empirical analysis we exclude the first three time periods.

whether Google Trends can serve as a good forecasting tool in China's automobile and communication appliances retail sales.

The keywords used in Google Trends are collected in a similar fashion as I did with Baidu Index, although I ended up with fewer Google Trends series compared to Baidu, because more keywords are recorded and valid in Baidu Index, partly due to the fact that many of the keywords are in Chinese.

As a unique function of Google Trends, it allows its users to limit search terms into a specific category, and these categories reflect the aggregate search trends of a specific topic. As a result, I included extra series based on Google Trends' "categories." To get a sense of how these "categories" work, for example, if you select the category "automobile and cars," Google Trends will aggregate the data for all searches that fit this category. In addition to the keywords, I added all the categories and sub-categories under "automobile and cars," "internet and telecommunications" as well as other categories that could be associated with automobile and communication appliances consumption like "Shopping," "Travel," "Games" etc.

Overall, 190 variables for automobile sales and 327 variables for communication appliances from Google Trends are incorporated into the models. As mentioned earlier, search queries associated with communication appliances are highly concentrated around specific periods. While for Baidu, I therefore aggregated some series by simply summing, this is not possible for the Google Trends series as Google Trends reports relative search volumes rather than absolute volume.

The same nowcast and forecast models (equations 4 and 5) are used to analyze the predictive performance of the Google Trend series.

I tried to match most of the model specifications between Google Trends and Baidu Index to facilitate comparison. There is one exception, however: since one cannot add the Google Trends series, I cannot run the regression with the sum of all the Google Trends data as I did for Baidu Index.

4.4.7 Predictions using the Consumer Confidence Index

As indicated in the literature review part of this chapter, existing literature that predicts consumption aggregates often uses survey-based indicators like consumer confidence index, and often compares internet search data with survey-based indicators. (Carroll, Fuhrer & Wilcox (1994), Bram & Ludvigson (1998), Howrey (2001), etc.)

In this chapter of my thesis, I will also analyze the predictive performance of adding the survey-based indicator to the baseline model, to compare the predictive power of consumer confidence with Baidu Index and Google Trends. Note that, unlike search intensity data, data for the CCI are available with a delay and hence enter as a lagged variable in the models:

Nowcast:

$$C_t = \alpha C_{t-1} + \beta_1 Date + \beta_2 Date^2 + \beta_{3\ to\ 12} Month\ Dummies + \beta_{13} CCI_{\mathrm{t-1}} \tag{4.9}$$

Forecast:

$$C_t = \alpha C_{t-2} + \beta_1 Date + \beta_2 Date^2 + \beta_{3-12} Month\ Dummies + \beta_{13} CCI_{\mathrm{t-2}} \tag{4.10}$$

This delay in the publication of the CCI data is illustrated in the above equations: in the nowcasting model, only CCI of the previous months is available, while in the forecasting model, only CCI of two months ago is available.

4.5 Empirical results: Automobile and Communication Appliances

4.5.1 Nowcasting results: Baidu Index

TABLE 4.2 shows the RMSFE of nowcasts of retail sales using different regression methodologies and specifications. The top panel shows the results for the OLS models. The bottom panel shows the results for the Lasso models. The nowcast results of automobile sales are shown in the left panel of the table, while the results for communication appliances are shown in the right panel. Besides the absolute RMSFE, for the models that include information from Baidu, I also show the reduction in RMSFE relative to the RMSFE of the baseline model (OLS or Lasso). Positive numbers show the percentage improvement in predictive performance, negative numbers mean that adding Baidu information decreased predictive performance.

Suppose we look at the OLS model that adds, to the baseline model, the sum of the Baidu Indices of the search terms. The left panel of TABLE 4.2 shows that including the sum of Baidu Index into the baseline model can improve nowcasting performance for automobile sales: including the contemporaneous Baidu sum improves forecasting accuracy by 3.45%. However, the same panel shows that adding Baidu information is not a guarantee to get a more accurate forecast: if lags of the Baidu sum are added, in addition to the contemporaneous values, the RMSFEs become worse than the RMSFE of the baseline model.

In the case of communication appliances (the right panel of TABLE 4.2), I find that including the Baidu sum always improves the nowcasts, and that including 3 lags of the sum of Baidu Index into the model improves the accuracy the most, reducing the RMSFE by 7.54% compared to the baseline model.

Using the sum of the Baidu series is unlikely to exploit all available information, so next, I try alternative ways of adding the information from Baidu. Including the first 3 principal components.

When using, instead of the sum of the search terms, the PCA factors of Baidu Index, I find that nowcasting accuracy of automobile sales can be further improved, when using up to 1 lag of PCA factors of Baidu Index, this improves predictive accuracy by 11.25%. Similarly, for the automobile sales, I also found that adding Baidu PCA factors contribute to predictive performance. However, the decrease in RMSFE for the PCA models are smaller for the communication appliances sector.

Next, rather than aggregating the Baidu series, I analyze what happens if I add, to the baseline model, the individual keyword series that gives the highest adjusted R square during the training period. TABLE 4.2 shows that when only keyword with the highest adjusted R square is included for each training period, I do not see a reduction in RMSFE for the automobile sector. However, if I add the 3 series, that individually gives the highest adjusted R square in the training period, jointly into one model, I get a reduction in the nowcast errors of 10.53% as compared to the baseline model.

For sales of communication appliances, including the Baidu Index keyword series with the highest adjusted R square reduces nowcasting errors by about 12.85%, while including the 3 best individual series jointly, nowcasting errors are reduced by about 13.84%.[9] Note that both these improvements are smaller improvements than the improvement I obtained when using the PCA method.

9 Note that the RMSFE of the nowcasts is sometimes the same when an extra lag of Baidu Index is included. This happens when the extra lag of the Baidu Index series does not improve forecasting accuracy over the best model with one less lag.

Next, I turn to the Lasso models at the bottom half of TABLE 4.2. In theory, Lasso models should be able to do better, because, unlike OLS, Lasso models do not force us to select ex-ante which individual series to include. Instead, Lasso models use the data to select the best models.

Similar to the results from the OLS models, I again find that adding Baidu information to the baseline model can improve forecasting accuracy. If the sum of the Baidu indices is added, I improve forecasting accuracy by about 9% for the automobile sales and about 7% for the communication appliances sales.

The best Lasso model for automobile sales is the model that adds PCA factors of the Baidu series. Specifically, the Lasso model for automobile sales that adds 1 lag of Baidu series PCA to the base model performed best, leading to a reduction of the RMSFE by 20.36%.

Adding the Baidu series jointly into the model also improves the model RMSFE by about 10% to 20%. Adding interactions between the Baidu series and the base model variables reduces the RMSFE further to about a 20% improvement compared to the base model.

Note further that adding all Baidu series jointly to the base model does not work for communication appliances as it has worse predictive accuracy than the base model. For communication appliances, the best Lasso model is the model that includes the sum of the Baidu series, rather than all series individually. Hence, comprehensive models are not always the better models.

In fact, TABLE 4.2 shows that, for both communication appliances and automobile sales, the overall best model is not the more complex Lasso model. The model with the lowest RMSFE is in both cases, an OLS model. They are the top 3 OLS model for automobile sales (a RMSFE of 0.0495, compared to the best baseline (the OLS baseline model) of 0.0558, an 11.25% improvement), and the sum of Baidu Index model for communication appliances (a RMSFE of 0.1034, compared to 0.1118 for the best baseline model (the OLS baseline model), a 7.54% improvement).[10]

10 While, for automobile sales, the Lasso model with all individual Baidu series shows the highest improvement over the OLS baseline model, the Lasso baseline model has a higher RMSFE than the baseline OLS model, allowing the best OLS model to be the overall best model even for automobile sales.

One possible reason for the relatively poor performance of the Lasso model is that the Lasso model can have difficulties handling highly correlated variables (Hastie, Tibshirani & Wainwright (2015)). Theoretically, when one has a large enough sample size, highly correlated explanatory variables will not cause problems. However, in my sample, there are at most 10 years of monthly data, so the sample size is relatively small.[11]

Summarizing my findings so far, the evidence suggests that nowcasting of Chinese consumption series can be improved by including search intensity information from Baidu, but also that there is no guarantee that adding such information will always improve predictive performance. In fact, I find that the best models are relatively simple models with some Baidu information rather than models with lots of Baidu series.

4.5.2 Forecasting results: Baidu Index

TABLE 4.3 shows the RMSFE and the reduction in RMSFE of the various forecasting models. Similar to TABLE 4.2, the top panel shows the forecasting results for OLS models while the bottom panel shows the results for Lasso models.

Overall, I observe the following pattern: OLS models with the sum and PCA factors of Baidu Index don't help much when forecasting sales in the automobile sector, and help somewhat when forecasting sales in communication appliances. But when I include the top 1 and top 3 most useful Baidu variables in the baseline model, there is a bigger reduction in forecasting errors. For the automobile series, the best OLS model incorporates the top 1 Baidu variables, reducing the RMSFE by 1470% compared to the baseline model. For the communication appliances sales, the model with the best individual Baidu series, reduces RMSFE by 12.34% compared to the baseline model without Baidu information.

As for the Lasso models, Lasso models do better than OLS models for automobile sales but do worse for communication appliances sales. The best model incorporates 3 additional lags of the Baidu Index series, reducing the RMSFE by 19.42% compared to the baseline Lasso model for the automobile sales.

11 Hastie, Tibshirani & Wainwright (2015) suggest using elastic nets rather than Lasso when variables are highly correlated. We also experimented with elastic net models but none of the elastic net models with Baidu information outperformed the Lasso baseline model. The results of the elastic net predictions are not listed here but are available upon request.

The results presented so far are based on data starting in 2011. However, the market share of Baidu was substantially lower in the early years of the sample because of the competition of Google, so Baidu search volumes tend to be low compared to more recent years. This structural change can affect the forecasting ability of the forecasting models. To check this, I will next focus on the period since 2015, after Google quit China.

4.5.3 Limiting the sample period to 2015–2019: Baidu Index

TABLES 4.4 and 4.5 show the RMSFE for expanding window nowcasts and forecasts of OLS and Lasso models when using data from 2015 to 2019. In this analysis, the model is thus trained initially using 3 years of data (January 2015 to December 2017), further adding one more observation into the training period each time a prediction is made (expanding window).

TABLE 4.4 shows that Baidu Index series contain extra information that can help nowcast sales in both the automobile sector and, to a lesser extent, the communication appliances sector. While including Baidu information does not always improve forecasting accuracy over the baseline model, the models with the lowest RMSFE indeed again include Baidu series.

When using the shorter time period, the best model for automobile sales is the Lasso model that includes 1 lag of the individual Baidu series PCA factors. This model has a RMSFE of 0.0454, an improvement of about 44% over the baseline Lasso model and an improvement of about 24% over the OLS baseline model. The best model for communication appliances is the OLS model that includes 3 lags of the sum of Baidu index but in this case, adding Baidu information only improves forecast accuracy by about 1.5% compared to the OLS baseline model.

TABLE 4.5 shows the prediction results for the shorter sample for both sectors and presents evidence that including the Baidu series in predictive models for the Chinese consumption series can improve predictive accuracy. For the automobile sector, the model with the lowest RMSFE is the OLS model incorporating the top 3 Baidu series, reducing forecasting errors by 24.39% relative to the best baseline model. For the communication appliances, the best OLS and the best Lasso models give similar improvements in accuracy over the best baseline model, improving forecasts about 3%.

4.5.4 Google Trends

So far, I have focused on whether Baidu search data can help to improve predictions. In this section, I use search intensities from Google Trends for the period 2015 to 2019. While Google was no longer available in mainland China after 2015, it could still be accessed using VPNs, so some data are available.

As mentioned earlier, Google Trends only provide a sample of the data, which can mean that the search volume numbers may be different depending on when the data is acquired, although this problem is only prominent with less popular search terms. This is likely to be a problem when using Google Trends in China, as this search engine is not very popular. Although my primary focus on this thesis has been on Baidu, nonetheless it may be useful to use Google Trends as an alternative to see if Google Trends can improve forecasting accuracy in consumption in China.

TABLES 4.6 and 4.7 show the nowcasting and forecasting results for the baseline and augmented models. The results of the baseline models are exactly the same as in TABLES 4.4 and 4.5. The RSMFSs and reductions in prediction errors for the OLS and Lasso models are listed, where once again, up to 3 lags for the nowcasting models and 4 lags for the forecasting models are included. The left and right panels correspond to the prediction results for the automobile and communication appliances sectors, respectively. Using PCA factors of Google Trends in Lasso model is able to reduce RMSFE by a little. Other than that, there is hardly any reduction in RMSFE from the Google augmented models. Adding Google Trends information thus does not improve predictive accuracy by much, especially when compared to the Baidu Index augmented models.

4.5.5 Consumer Confidence Index (long sample)

Survey-based indicators like the Consumer Confidence Index (CCI) and the Consumer Sediment Index are often linked with forecasting sales and consumption before other datasets like Baidu Index and Google Scholar are incorporated. Therefore, in the following section CCI in China is used to predict sectoral retail sales.

TABLES 4.8 and 4.9 show prediction accuracy and decrease in forecasting errors of CCI augmented models relative to the baseline models, using data between 2011-2019. The results provided limited evidence that CCI improves the nowcasting accuracy of automobile and communication appliance sales. CCI also doesn't seem to improve forecasts of communication appliances sales, as the OLS baseline model has the lowest RMSFE for all these models. When forecasting automobile sales, results are somewhat improved by including CCI information. The best model, the Lasso model that includes the CCI, improves accuracy by about 12.5% over the baseline Lasso model, and about 6% over the baseline OLS model. If I compare the added value of CCI to the added value of the Baidu series, however, the added value of the CCI is smaller, as the Baidu series is able to reduce a bigger percentage of the forecasting errors in several model specifications.

4.5.6 CCI and Baidu Index (long sample)

In my previous results, I showed that models with Baidu Index or CCI information can sometimes produce more accurate results than the baseline models. In the following section, I explore if, by adding both Baidu Index and CCI information into the models, prediction accuracy can be improved, and how do these improvements compare to my previous models. Prediction results using the long sample are presented in TABLE 4.10 and TABLE 4.11.

To see if there's indeed any added value when both Baidu Index and CCI are included, firstly I compare the results from TABLE 4.2 (nowcasting with Baidu but without CCI) and TABLE 4.10 (nowcasting with Baidu and CCI), it's evident that in most cases results in TABLE 4.2 are better than that in TABLE 4.10, this indicates that in many cases models incorporating only Baidu Index does a better job compared to the models using both CCI and Baidu Index. In terms of the forecasting results, however, if I compare the forecasting results of TABLE 4.3 (Baidu, no CCI) with TABLE 4.11 (Baidu and CCI), for the automobile sector the best performing forecasting model with the smallest RMSFE is the model that includes 3 additional lags of both CCI and Baidu Index, which improves accuracy by about 22.9% over the baseline Lasso model, and about 17.2% over the baseline OLS model. This is not surprising because both the results in TABLE 4.3 and TABLE 4.9 suggest that when 3 additional lags of CCI or 3 additional lags of Baidu Index increases forecasting accuracy by a lot. However, the same cannot be said for the communication appliances sector, as none of the models performed better than the baseline model when both CCI and Baidu Index were included.

In conclusion, when it comes to using both CCI and Baidu Index information, the tables indicated that in most cases, CCI is not very useful in predictions in both sectors, however, this is subject to the specification of the model used. In a few scenarios, there is a bigger improvement in prediction accuracy when both Baidu Index and CCI are incorporated.

4.6 Empirical results: Total Retail Sales

4.6.1 Collection of data and methodology

In the previous analysis, I focused on predicting only two sectors of the retail sales of consumer goods in China. The reasons why only these two sectors are chosen are detailed in section 4.1.

Although the previous results already indicated that keyword series related to brands, models, and related searches are able to reduce prediction errors in retail sales of the two sectors, it will be interesting to see if my findings can be generalized to the entire total retail sales of consumer goods. For instance, government agencies and policy makers are often more interested in aggregated measures tracking the entire consumer demand, which in turn provides an overall indicator for economic health and domestic consumption.

The difficulties associated with using Baidu Index to predict total retail sales in China lies in the ambiguity of the related keywords. For my previous models, I was able to assemble a list of keywords by using the limited brands and models that are associated with both sectors. However, it is indeed substantially more difficult to come up with a list of keywords that correlates to the entire total retail sales.

This, however, doesn't seem to be an issue when using Google Trends, because, unlike Baidu Index, Google Trends provides the unique function of limiting keywords into a specific category. Using the example in section 3 of this chapter, one has the option when collecting search data on Apple to limit the collection the search volumes to reflect queries for Apple, the technology company, as opposed to Apple the fruit. This function is not only very helpful when a specific keyword can represent many different objects, but it also comes in handy as a measure of the popularity of aggregate searches conducted under this category. To be precise, when one selects a category without imputing a specific keyword, Google Trends will show an aggregate measure for the search volume of all the related keywords under this category for the chosen time span. Google Trends has 1132 categories and sub-categories as of Feb,2021, which provides a comprehensive categorical system that measures all the searchers conducted by its users. This category data has facilitated past papers to predict aggregate consumption using Google Trends. (Vosen & Schmidt, 2011; Woo & Owen, 2019)

This unique function in Google Trends, although unavailable in Baidu Index, provide me with the opportunity to construct a list of keywords based on the comprehensive categories. Specifically, I used the names of the categories and sub-categories as keywords to collect search volume series from Baidu Index, and used these series to nowcast and forecast total retail sales of consumer goods in China. I started with a list of 1132 keywords, corresponding to the 1132 categories and sub-categories from Google Trends, and started gathering data by imputing these keywords in Baidu Index, many of these keywords don't form a valid search on Baidu, when this happens, Baidu usually issues the following warning: "Keyword "XXX" is not included or recorded by Baidu Index," followed by an option to purchase a keyword. Baidu will then start to record it after the keyword is purchased. As a result, only 982 of the keywords yield useable search volume series. In the following section, I attempt to nowcast and forecast total retail sales using these 982 keywords series from Baidu Index[12].

The model specifications are similar to the ones previously used in this chapter of my thesis. Specifically, the following methods are used:

First, by using the sum of the Baidu Index series.

Secondly, by adopting principal component analysis to transform all the Baidu Index series that has a correlation coefficient above 0.9, into factor loadings.

Thirdly, by running a regression with each Baidu series separately and find the series that individually adds most to the baseline model during the training period.

The models are calibrated using both OLS and Lasso, allowing up to 3 lags for nowcasting and 4 lags for forecasting to explore any information embedded in the lagged series of Baidu Index. The specification of the models is consistent with the previous methodologies.

4.6.2 Nowcasting results:

Similar to my previous tables, TABLE 4.12 shows the RMSFE of nowcasts of total retail sales using different regression methodologies and specifications, with the top panel shows the results for the OLS models, and the bottom panel shows the results for the Lasso models.

12 See the following link for the full list of Google categories, https://doi.org/10.7910/DVN/YT25IP, Harvard Dataverse.

Both absolute RMSFE and the reduction in RMSFE relative to the RMSFE of the baseline model (OLS or Lasso) are listed.

As the table suggests, in most cases adding Baidu Index into the baseline model can improve nowcasting performance. The best OLS model is the one that adds, to the baseline model, 3 series that individually increase the adjusted R square the most. Specifically, when 3 additional lags of Baidu Index are included, this model is able to reduce 24.51% of the nowcasting errors.

In the case of Lasso models, I find that including the Baidu Index series always improves the nowcast, and that including 3 lags of the individual Baidu Index series into the model improves the accuracy the most, reducing the RMSFE by 42.28% compared to the baseline Lasso model. This model is also the model with the overall lowest RMSFE, which is 0.0203.

4.6.3 Forecasting results:

TABLE 4.13 shows the forecasting results for total retail sales. This table is structured similarly to my previous results.

In terms of the OLS models, I found that in all cases, the incorporation of Baidu Index series decreases forecasting errors, and this decrease seems to be more prominent as more lags are included in the models. Specifically, if the sum of Baidu Index series is added to the model, the forecasting accuracy is improved by around 0.5% to 8%. If principal components are added to the models, forecasting accuracies are improved by between 5.49% to 31.27%. While if 3 series that individually increased the in sample adjusted R square the most are added to the model, this reduces around 19.1% to 31.64% of the forecasting errors. The model with the smallest RMSFE is the model that adds, to the baseline model, 1 series that individually increases the most adjusted R square in the training period, at most this specification (with 3 additional lags) is able to reduce 48.84% of the forecasting errors.

Similarly, the results of the Lasso models also indicated that in all cases Baidu Index series decrease RMSFE. The Lasso model with individual factors is able to reduce 34.15% of the forecasting errors.

In summary, the evidence suggests that when Baidu Index series are included in the models, in most cases, I see a prominent improvement in prediction accuracy. To be precise, in some specifications, the inclusion of Baidu Index is able to reduce 42% of the nowcasting errors and 48% of the forecasting errors compared to the baseline models.

4.7 Conclusion

In this chapter of my thesis, my main contribution is to analyze the use of internet search intensity data from search engines in producing consumption related economic nowcasts and forecasts. Contrary to the previous studies, that focused primarily on Google Trends, this chapter of my thesis looked at the potential of Baidu Index, a leading search engine and data source that is popular in China.

My results indicate that Baidu Index contains information on sales for both the automobile sector and communication appliances sector. The models show clear evidence that incorporating search intensity data from Baidu can improve the nowcasts and forecasts of the sales of automobiles and the sales of communication appliances in China. These improvements are substantial, as adding information from Baidu can reduce predictive errors by 10% or more. In addition, my results indicated that simple models that incorporated Baidu Index typically perform better than complicated models with a lot of Baidu Index series, and that OLS models mostly do a better job in nowcasting and forecasting than LASSO models. When comparing the added value of Google Trends and CCI to Baidu Index, I show that Baidu Index augmented models performed better than Google Trends models or CCI models.

Finally, by collecting keywords associated with the categories in Google Trends, and imputing these keywords in Baidu Index. I tried to generalize my finding to the entire retail sales of consumer goods in China. The results suggested that in almost every model specification, predictions on the total retail sales are improved by a substantial amount.

My results will benefit both private companies and government organizations in China. Private companies may find my results useful in forecasting demand and making operation and inventory decisions, while government agencies may use better forecasts to understand how demand will evolve in the coming months, to determine future monetary and fiscal policies.

4.8 Appendix

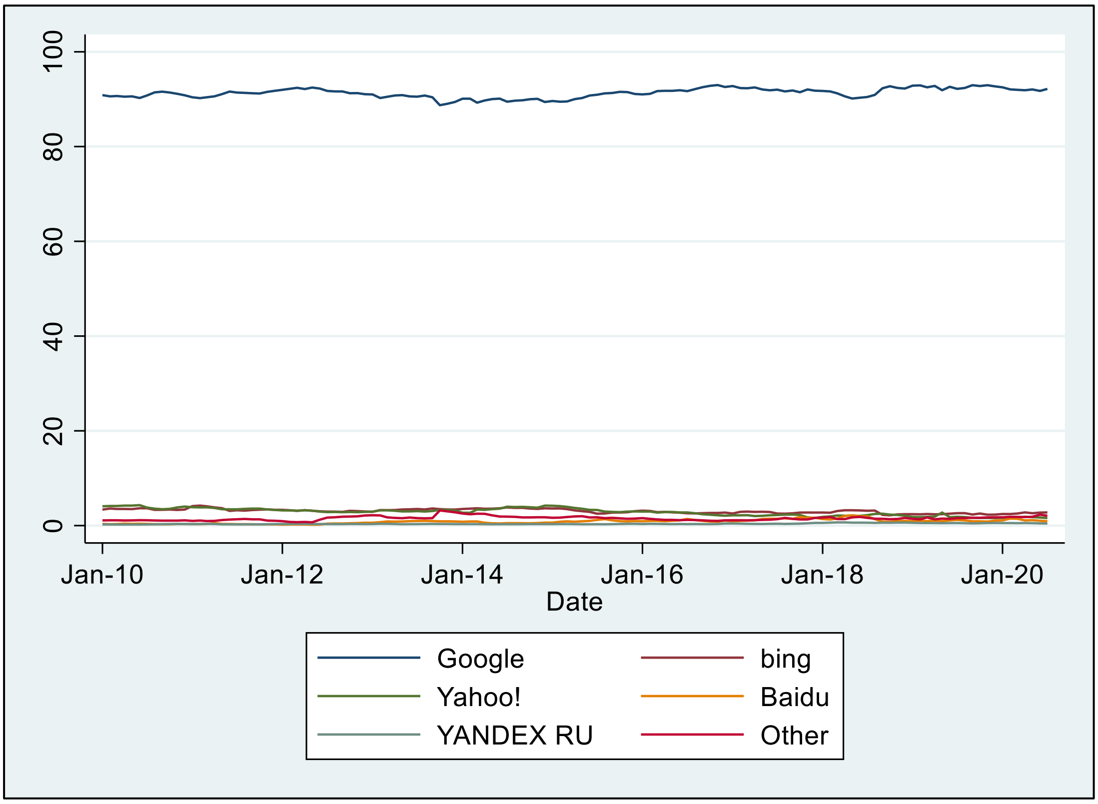
Table 4.1 Comparison between Google Trends and Baidu Index

Features	***Google Trends***	***Baidu Index***
• Limit to a specific country	Yes	No, only China
• Limit to a specific region within that country	Yes	Yes
• Limit to a specific time period	Yes, earliest Jan.2004	Yes, earliest June. 2006 for PC and Dec.2010 for phones.
• Limit to a specific category	Yes	No
• Maximum number of terms that can be compared	5	5
• Search volume reported	Relative volume	Absolute volume
• Average search volume	Reports average across the sample period	Reports both average and daily moving average across the sample period.
• Report total search volume of several terms	Yes	Yes
• Method of matching	Partial matching	Complete matching
• Related searches	Yes	Yes

• Demography of the people	Only shows the region where the searches are from	Region, age, gender, and information on what sectors are people interested in when they search for a certain keyword.
• Separate searches from different user platforms (PC or phones)	No	Yes
• Show the news headlines related to the search terms	No	Yes, but only when there's a spike in the search volume.
• Limit to a specific search option (News, Pictures, etc.)	Yes	No
• Measure of popularity amongst internet users and news outlets	No	Yes

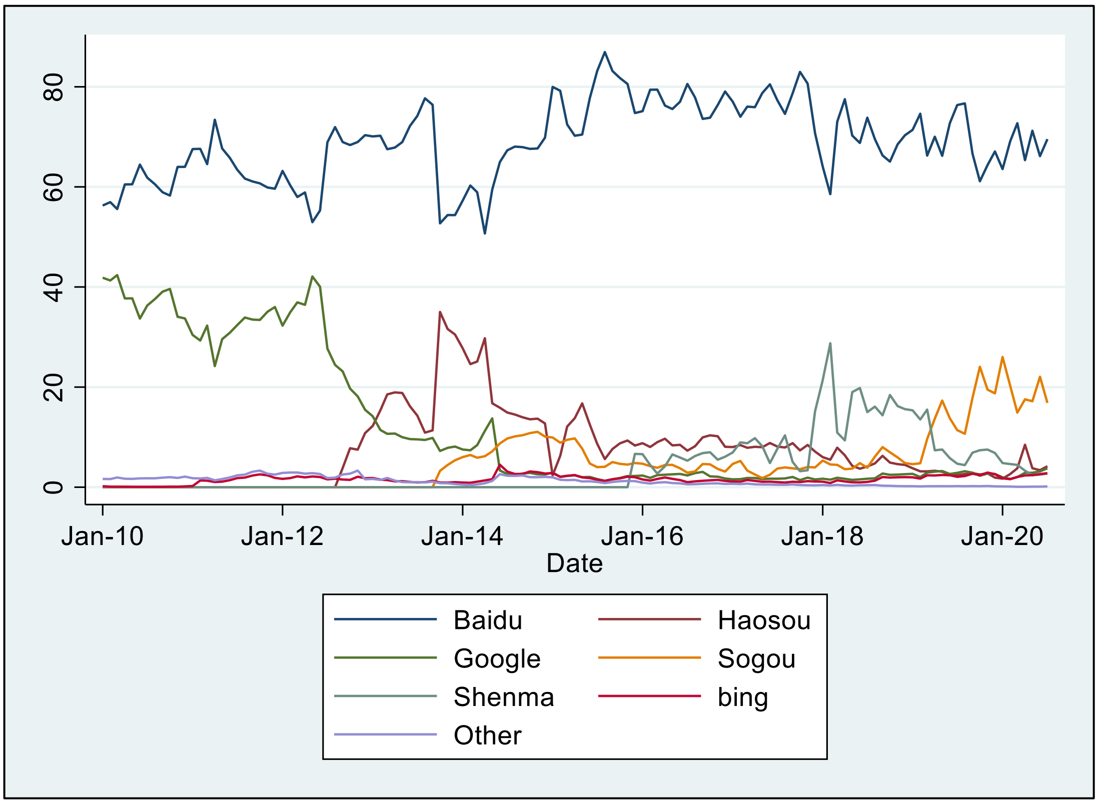
Source: updated from Vaughn and Chen (2015), Google Trends, Baidu Index.

Figure 4. 1 Search engine market share worldwide



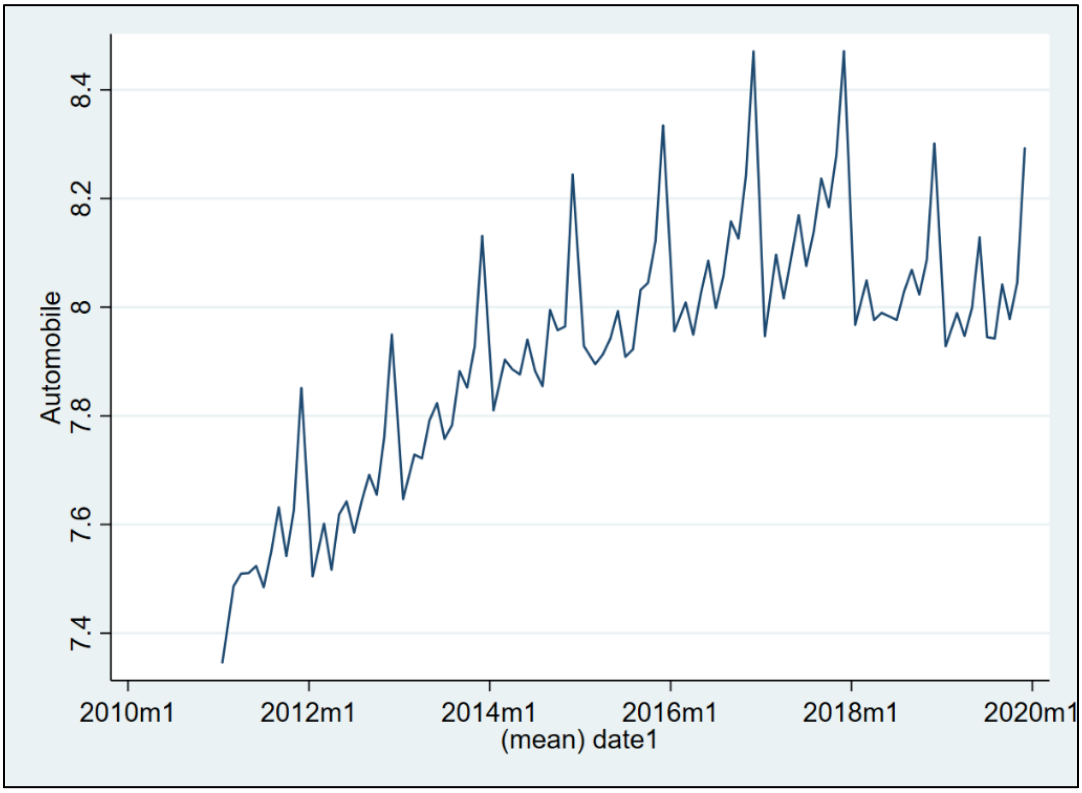
(<u>Source</u>: Statcounter GlobalStats 2020)

Figure 4. 2 Search engine market share in China



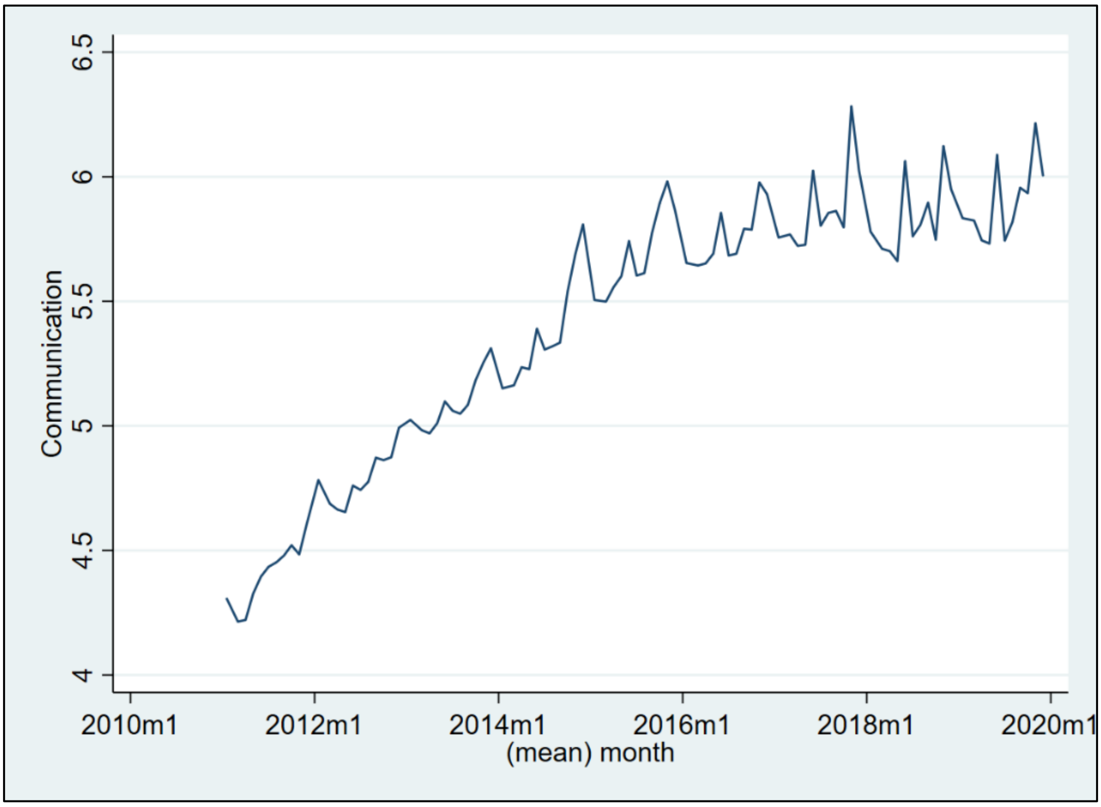
(Source: Statcounter GlobalStats 2020)

Figure 4. 3 Natural logarithm of automobile sales in China



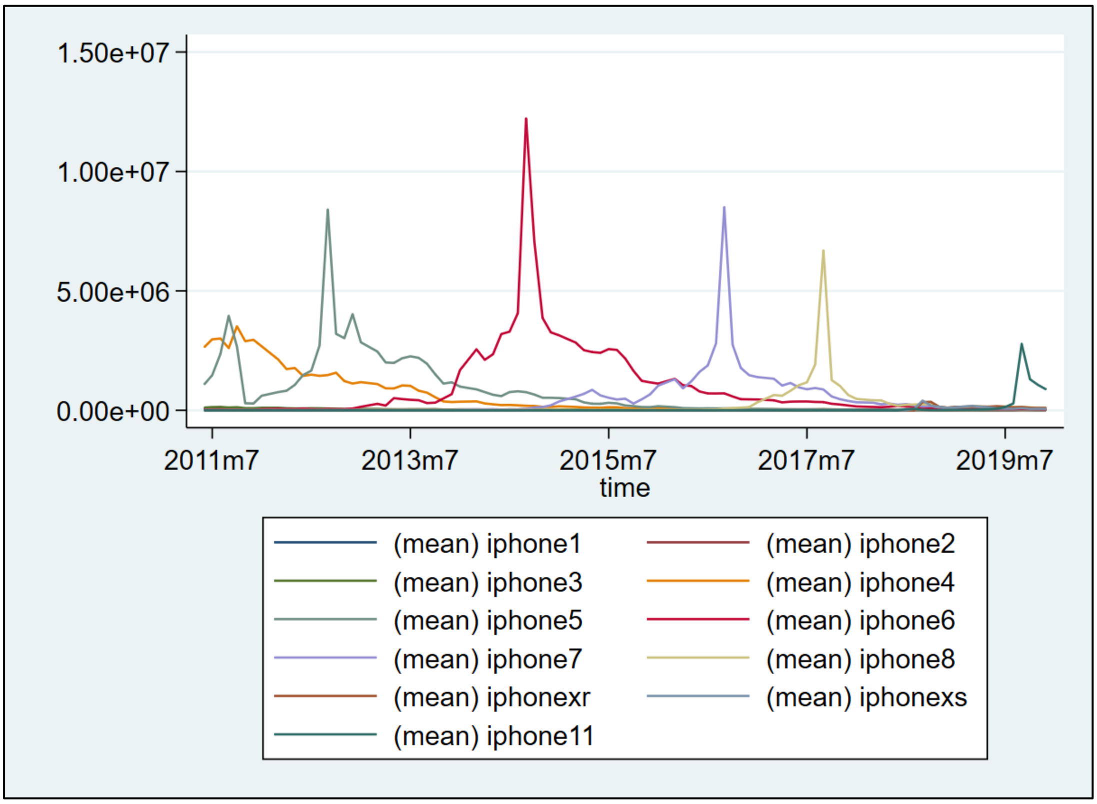
(Source: National Bureau of Statistics of China)

Figure 4. 4 Natural logarithm of communication appliances sales in China



(Source: National Bureau of Statistics of China)

Figure 4. 5 Search Volumes for iPhone Related Keywords In China



(Source: Baidu Index)

Table 4.2 Nowcasting with Information from Baidu (full sample)

	Lags	Automobile RMSFE	Automobile Reduction	Communication RMSFE	Communication Reduction
A) OLS					
Baseline		0.0558		0.1118	
Sum	L0	0.0538	**0.0345**	0.1077	0.0364
	L1	0.0570	-0.0224	0.1055	0.0558
	L2	0.0598	-0.0719	0.1036	0.0731
	L3	0.0622	-0.1147	0.1034	**0.0754**
PCA	L0	**0.0495**	**0.1125**	0.1103	0.0132
	L1	0.0504	0.0957	0.1104	0.0128
	L2	0.0531	0.0481	0.1110	0.0070
	L3	0.0579	-0.0387	0.1100	0.0161
Best series	L0	0.0642	-0.1505	0.0974	**0.1286**
	L1	0.0634	-0.1367	0.0974	0.1286
	L2	0.0637	-0.1420	0.1065	0.0475
	L3	0.0623	-0.1176	0.1065	0.0475
Top 3 series	L0	0.0500	0.1033	0.0983	0.1205
	L1	0.0499	**0.1053**	0.0963	**0.1384**
	L2	0.0539	0.0340	0.1043	0.0674
	L3	0.0538	0.0352	0.1043	0.0674
B) LASSO					
Baseline		0.0649		0.1177	
Sum	L0	0.0592	0.0885	0.1090	**0.0736**
	L1	0.0608	0.0636	0.1188	-0.0093
	L2	0.0627	0.0351	0.1159	0.0151
	L3	0.0657	-0.0120	0.1166	0.0093
PCA	L0	0.05174	**0.2032**	0.1175	0.0016
	L1	0.05171	0.2036	0.1195	-0.0157
	L2	0.0540	0.1684	0.1185	-0.0071
	L3	0.0609	0.0617	0.1117	**0.0507**
Individual Factors	L0	0.0580	0.1067	0.1459	-0.2397
	L1	0.0568	0.1250	0.1596	-0.3564
	L2	0.0599	0.0775	0.1713	-0.4560
	L3	0.0533	**0.1794**	0.1811	-0.5391
Interactions	L0	0.0522	**0.1965**	0.1138	**0.0325**
	L1	0.0567	0.1268	0.1342	-0.1405
	L2	0.0523	0.1950	0.1216	-0.0339
	L3	0.0535	0.1760	0.1509	-0.2826

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.3 Forecasting with Information from Baidu (full sample)

		Automobile		***Communication***	
	Lags	***RMSFE***	***Reduction***	***RMSFE***	***Reduction***
A) OLS					
Baseline		0.0665		0.1102	
Sum	L0	0.0671	-0.0086	0.1015	0.0789
	L1	0.0697	-0.0472	0.0987	0.1044
	L2	0.0728	-0.0937	0.0973	**0.1176**
	L3	0.0778	-0.1688	0.0979	0.1121
PCA	L0	0.0666	-0.0008	0.1162	-0.0543
	L1	0.0661	0.0067	0.1169	-0.0606
	L2	0.0648	0.0259	0.1168	-0.0593
	L3	0.0736	-0.1059	0.1142	-0.0364
Best series	L0	0.0713	-0.0722	**0.0966**	**0.1234**
	L1	0.0665	0.0009	0.1094	0.0070
	L2	0.0625	0.0603	0.1094	0.0070
	L3	**0.0567**	**0.1470**	0.1094	0.0070
Top 3 series	L0	0.0649	0.0249	0.0983	**0.1085**
	L1	0.0634	0.0466	0.1036	0.0600
	L2	0.0571	**0.1412**	0.1036	0.0600
	L3	0.0601	0.0966	0.1036	0.0600
B) LASSO					
Baseline		0.0714		0.1110	
Sum	L0	0.0685	0.0407	0.1169	-0.0539
	L1	0.0623	**0.1276**	0.1167	-0.0514
	L2	0.0652	0.0861	0.1190	-0.0724
	L3	0.0687	0.0374	0.1195	-0.0767
PCA	L0	0.0677	0.0510	0.1232	-0.1103
	L1	0.0748	-0.0488	0.1160	-0.0453
	L2	0.0714	-0.0001	0.1229	-0.1080
	L3	0.0708	0.0078	0.1170	-0.0545
Individual Factors	L0	0.0866	-0.2134	0.1568	-0.4133
	L1	0.0753	-0.0545	0.1529	-0.3777
	L2	0.0604	0.1537	0.1781	-0.6050
	L3	0.0575	**0.1942**	0.1651	-0.4881
Interactions	L0	0.0730	-0.0224	0.1620	-0.4597
	L1	0.0783	-0.0968	0.1686	-0.5194
	L2	0.0601	**0.1577**	0.1993	-0.7958
	L3	0.0610	0.1449	0.1927	-0.7362

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.4 Nowcasting with Information from Baidu (short sample)

		Automobile		Communication	
	Lags	***RMSFE***	***Reduction***	***RMSFE***	***Reduction***
A) OLS					
Baseline		0.0601		0.0777	
Sum	L0	0.0556	**0.0750**	0.0815	-0.0494
	L1	0.0581	0.0329	0.0792	-0.0193
	L2	0.0585	0.0257	0.0782	-0.0074
	L3	0.0613	-0.0196	**0.0765**	**0.0142**
PCA	L0	0.0461	**0.2333**	0.0816	-0.0512
	L1	0.0464	0.2284	0.0845	-0.0876
	L2	0.0462	0.2305	0.0994	-0.2801
	L3	0.0467	0.2222	0.0994	-0.2801
Best series	L0	0.0621	-0.0328	0.1644	-1.1174
	L1	0.0620	-0.0315	0.1673	-1.1550
	L2	0.0580	**0.0345**	0.1767	-1.2758
	L3	0.0580	0.0345	0.1041	-0.3408
Top 3 series	L0	0.0631	-0.0495	0.1245	-0.6037
	L1	0.0640	-0.0657	0.1300	-0.6742
	L2	0.0611	-0.0165	0.1367	-0.7602
	L3	0.0638	-0.0610	0.0971	-0.2506
B) LASSO					
Baseline		0.0815		0.0845	
Sum	L0	0.0659	**0.1919**	0.0847	-0.0026
	L1	0.0676	0.1705	0.0815	0.0357
	L2	0.0713	0.1249	0.0786	0.0702
	L3	0.0729	0.1056	0.0769	**0.0903**
PCA	L0	0.0479	0.4124	0.0874	-0.0348
	L1	0.0454	**0.4430**	0.0829	0.0189
	L2	0.0472	0.4214	0.0930	-0.1008
	L3	0.0469	0.4245	0.0861	-0.0184
Individual Factors	L0	0.0659	0.1919	0.2264	-1.6794
	L1	0.0627	0.2307	0.1061	-0.2558
	L2	**0.0526**	**0.3548**	0.1171	-0.3861
	L3	0.0552	0.3234	0.1199	-0.4188
Interactions	L0	0.0718	0.1198	0.2165	-1.5622
	L1	0.0945	-0.1587	0.1183	-0.3998
	L2	0.0761	**0.0667**	0.1267	-0.4992
	L3	0.1068	-0.3107	0.1494	-0.7677

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.5 Forecasting with Information from Baidu (short sample)

	Lags	*Automobile* RMSFE	*Automobile* Reduction	*Communication* RMSFE	*Communication* Reduction
A) OLS					
Baseline		0.0707		0.0785	
Sum	L0	0.0638	**0.0971**	0.0783	0.0027
	L1	0.0659	0.0676	**0.0759**	**0.0325**
	L2	0.0673	0.0475	0.0762	0.0286
	L3	0.0685	0.0314	0.0780	0.0058
PCA	L0	0.0692	0.0209	0.0803	-0.0232
	L1	0.0660	0.0661	0.0815	-0.0388
	L2	0.0667	0.0562	0.0848	-0.0805
	L3	0.0667	0.0562	0.0925	-0.1788
Best series	L0	0.0599	**0.1523**	0.0901	-0.1473
	L1	0.0615	0.1295	0.0867	-0.1049
	L2	0.0611	0.1366	0.0957	-0.2194
	L3	0.0616	0.1294	0.0957	-0.2194
Top 3 series	L0	0.0671	0.0507	0.0932	-0.1874
	L1	0.0583	0.1758	0.0886	-0.1291
	L2	**0.0535**	**0.2439**	0.0858	-0.0935
	L3	0.0584	0.1739	0.0858	-0.0935
B) LASSO					
Baseline		0.0789		0.0823	
Sum	L0	0.0716	0.0929	0.0805	0.0220
	L1	0.0719	0.0898	**0.0759**	**0.0775**
	L2	0.0711	**0.0997**	0.0781	0.0512
	L3	0.0717	0.0920	0.0785	0.0466
PCA	L0	0.0698	0.1154	0.0838	-0.0180
	L1	0.0676	0.1436	0.0845	-0.0269
	L2	0.0668	**0.1533**	0.0907	-0.1025
	L3	0.0734	0.0698	0.0904	-0.0991
Individual Factors	L0	0.0781	0.0113	0.1092	-0.3275
	L1	0.0803	-0.0166	0.1184	-0.4393
	L2	0.1020	-0.2916	0.1164	-0.4151
	L3	0.0905	-0.1468	0.1150	-0.3971
Interactions	L0	0.1228	-0.5551	0.1382	-0.6800
	L1	0.0876	-0.1093	0.1149	-0.3958
	L2	0.0829	-0.0499	0.1018	-0.2371
	L3	0.0858	-0.0868	0.1124	-0.3661

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.6 Nowcasting with Information from Google (short sample)

		Automobile		***Communication***	
	Lags	***RMSFE***	***Reduction***	***RMSFE***	***Reduction***
A) OLS					
Baseline		0.0601		0.0777	
PCA	L0	0.0681	-0.1327	0.0897	-0.1554
	L1	0.0663	-0.1026	0.0845	-0.0885
	L2	0.0664	-0.1053	0.0840	-0.0821
	L3	0.0632	-0.0512	0.0787	-0.0134
Best series	L0	0.0618	-0.0281	0.0768	0.0106
	L1	0.0667	-0.1097	0.0843	-0.0851
	L2	0.0612	-0.0181	0.0868	-0.1181
	L3	0.0623	-0.0369	0.0883	-0.1369
Top 3 series	L0	0.0615	-0.0239	0.1761	-1.2679
	L1	0.0667	-0.1098	0.0877	-0.1288
	L2	0.0707	-0.1772	0.0931	-0.1987
	L3	0.0729	-0.2135	0.0881	-0.1348
B) LASSO					
Baseline		0.0815		0.0845	
PCA	L0	0.0727	0.1081	0.0850	-0.0062
	L1	0.0702	**0.1390**	0.0910	-0.0764
	L2	0.0837	-0.0266	0.0827	**0.0214**
	L3	0.0744	0.0869	0.0841	0.0046
Individual Factors	L0	0.1048	-0.2856	0.1234	-0.4601
	L1	0.1094	-0.3420	0.1439	-0.7029
	L2	0.1109	-0.3608	0.1496	-0.7703
	L3	0.1090	-0.3371	0.1625	-0.9233
Interactions	L0	0.1070	-0.3129	0.1247	-0.4754
	L1	0.1094	-0.3418	0.1363	-0.6133
	L2	0.1097	-0.3456	0.1409	-0.6679
	L3	0.1039	-0.2749	0.1324	-0.5673

NOTE: L stands for the number of lagged Google series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. PCA adds the first 8 principal component. Best series adds the Google series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors add all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.7 Forecasting with Information from Google (short sample)

		Automobile		*Communication*	
	Lags	*RMSFE*	*Reduction*	*RMSFE*	*Reduction*
		A) OLS			
Baseline	L0	0.0707		0.0785	
PCA	L1	0.0745	-0.0542	0.0866	-0.1039
	L2	0.0740	-0.0464	0.0776	0.0108
	L3	0.0775	-0.0957	0.0797	-0.0158
	L0	0.0747	-0.0570	0.0924	-0.1777
Best series	L0	0.0800	-0.1312	0.0883	-0.1246
	L1	0.0861	-0.2171	0.0948	-0.2079
	L2	0.0852	-0.2050	0.0946	-0.2051
	L3	0.0852	-0.2050	0.0874	-0.1138
Top 3 series	L0	0.0761	-0.0756	0.0896	-0.1413
	L1	0.0799	-0.1296	0.0893	-0.1380
	L2	0.0853	-0.2060	0.0893	-0.1372
	L3	0.0892	-0.2622	0.0990	-0.2615
		B) LASSO			
Baseline		0.0789		0.0823	
PCA	L0	0.0911	-0.1540	0.0767	0.0676
	L1	0.0849	-0.0758	0.0764	**0.0713**
	L2	0.0888	-0.1252	0.0853	-0.0362
	L3	0.0841	-0.0652	0.0852	-0.0349
Individual Factors	L0	0.0898	-0.1378	0.1241	-0.5087
	L1	0.0946	-0.1987	0.0818	**0.0060**
	L2	0.0987	-0.2507	0.1693	-1.0579
	L3	0.1016	-0.2869	0.1550	-0.8831
Interactions	L0	0.0942	-0.1930	0.1324	-0.6092
	L1	0.0864	-0.0947	0.1538	-0.8696
	L2	0.1036	-0.3120	0.1454	-0.7674
	L3	0.1047	-0.3258	0.1263	-0.5352

NOTE: L stands for the number of lagged Google series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. PCA adds the first 8 principal component. Best series adds the Google series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors add all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.8 Nowcasting with CCI Information (long sample)

		Automobile		***Communication***	
	Lags	***RMSFE***	***Reduction***	***RMSFE***	***Reduction***
		A) OLS			
Baseline		0.0558		0.1118	
CCI	L1	0.0575	-0.0306	0.1162	-0.0395
	L2	0.0565	-0.0135	0.1161	-0.0382
	L3	0.0574	-0.0293	0.1167	-0.0438
	L4	0.0576	-0.0327	0.1205	-0.0780
		B) LASSO			
Baseline		0.0649		0.1177	
CCI	L1	0.0654	-0.0077	0.1306	-0.1100
	L2	0.0636	**0.0212**	0.1152	**0.0212**
	L3	0.0637	0.0183	0.1192	-0.0128
	L4	0.0637	0.0198	0.1388	-0.1798

NOTE: L stands for the number of lagged CCI series. L2 means both lags 1 and 2 are included. Forecasting models goes back more lag than nowcasting models.

Table 4.9 Forecasting with CCI Information (long sample)

		Automobile		*Communication*	
	Lags	*RMSFE*	*Reduction*	*RMSFE*	*Reduction*
A) OLS					
Baseline		0.0665		0.1102	
CCI	L2	0.0659	0.0092	0.1141	-0.0353
	L3	0.0660	0.0087	0.1173	-0.0642
	L4	0.0665	0.0004	0.1226	-0.1122
	L5	0.0645	0.0303	0.1253	-0.1365
B) LASSO					
Baseline		0.0714		0.1110	
CCI	L2	0.0705	0.0122	0.1133	-0.0208
	L3	0.0645	0.0969	0.1149	-0.0358
	L4	0.0643	0.0994	0.1453	-0.3098
	L5	**0.0624**	**0.1250**	0.1450	-0.3063

NOTE: L stands for the number of lagged CCI series. L3 means both lags 2 and 3 are included. Forecasting models go back more lags than nowcasting models.

Table 4.10 Nowcasting with CCI and Baidu (full sample)

	Lags	*Automobile* RMSFE	*Automobile* Reduction	*Communication* RMSFE	*Communication* Reduction
A) OLS					
Baseline		0.0558		0.1118	
Sum	L0	0.0544	0.0238	0.1126	-0.0077
	L1	0.0569	-0.0201	0.1107	0.0101
	L2	0.0616	-0.1041	0.1096	0.0192
	L3	0.0660	-0.1829	0.1109	0.0076
PCA	L0	0.0547	0.0193	0.1038	0.0715
	L1	0.0535	0.0398	0.1059	0.0524
	L2	0.0558	-0.0014	0.1086	0.0283
	L3	0.0585	-0.0491	0.1099	0.0171
Best series	L0	0.0654	-0.1726	0.1019	**0.0888**
	L1	0.0634	-0.1366	0.1027	0.0812
	L2	0.0649	-0.1633	0.1086	0.0284
	L3	0.0666	-0.1949	0.1099	0.0166
Top 3 series	L0	0.0526	0.0564	0.1035	0.0743
	L1	**0.0502**	**0.1001**	0.1042	0.0679
	L2	0.0566	-0.0151	0.1087	0.0272
	L3	0.0595	-0.0664	0.1072	0.0406
B) LASSO					
Baseline		0.0649		0.1177	
Sum	L0	0.0596	**0.0827**	0.1261	-0.0716
	L1	0.0602	0.0723	0.1311	-0.1141
	L2	0.0628	0.0333	0.1260	-0.0711
	L3	0.0683	-0.0521	0.1357	-0.1534
PCA	L0	0.0543	0.1632	0.1219	-0.0362
	L1	0.0531	0.1830	0.1250	-0.0625
	L2	0.0556	0.1436	0.1261	-0.0721
	L3	0.0573	0.1171	0.1167	0.0081
Individual Factors	L0	0.0591	0.0903	0.1456	-0.2373
	L1	0.0575	0.1143	0.1587	-0.3488
	L2	0.0599	0.0775	0.1713	-0.4560
	L3	**0.0505**	**0.2225**	0.1827	-0.5526
Interactions	L0	0.0585	0.0995	0.1138	0.0325
	L1	0.0647	0.0037	0.1342	-0.1405
	L2	0.0647	0.0044	0.1216	-0.0339
	L3	0.0528	**0.1874**	0.1515	-0.2876

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.11 Forecasting with CCI and Baidu (full sample)

		Automobile		***Communication***	
	Lags	***RMSFE***	***Reduction***	***RMSFE***	***Reduction***
A) OLS					
Baseline		0.0665		0.1102	
Sum	L0	0.0671	-0.0092	0.1068	**0.0312**
	L1	0.0710	-0.0670	0.1069	0.0299
	L2	0.0754	-0.1330	0.1077	0.0227
	L3	0.0780	-0.1717	0.1114	-0.0110
PCA	L0	0.0683	-0.0269	0.1189	-0.0788
	L1	0.0661	0.0068	0.1211	-0.0988
	L2	0.0663	0.0036	0.1210	-0.0978
	L3	0.0700	-0.0517	0.1181	-0.0712
Best series	L0	0.0631	0.0519	0.9073	0.0927
	L1	0.0623	0.0634	1.0013	-0.0013
	L2	0.0601	0.0962	1.0450	-0.0450
	L3	**0.0583**	**0.1236**	1.0294	-0.0294
Top 3 series	L0	0.0703	-0.0562	**0.9438**	**0.0562**
	L1	0.0616	0.0740	0.9744	0.0256
	L2	0.0635	0.0453	0.9899	0.0101
	L3	0.0592	**0.1104**	1.0162	-0.0162
B) LASSO					
Baseline		0.0714		0.1110	
Sum	L0	0.0683	0.0428	0.1200	-0.0810
	L1	0.0632	**0.1142**	0.1170	-0.0539
	L2	0.0662	0.0724	0.1229	-0.1077
	L3	0.0679	0.0481	0.1301	-0.1726
PCA	L0	0.0670	0.0615	0.1260	-0.1355
	L1	0.0657	0.0793	0.1284	-0.1572
	L2	0.0670	0.0606	0.1263	-0.1381
	L3	0.0664	0.0694	0.1199	-0.0807
Individual Factors	L0	0.0866	-0.2134	0.1568	-0.4133
	L1	0.0753	-0.0545	0.1529	-0.3777
	L2	0.0604	0.1537	0.1780	-0.6046
	L3	**0.0551**	**0.2285**	0.1517	-0.3670
Interactions	L0	0.0730	-0.0224	0.1620	-0.4597
	L1	0.0783	-0.0968	0.1686	-0.5194
	L2	0.0601	**0.1577**	0.1993	-0.7958
	L3	0.0638	0.1056	0.1862	-0.6784

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.12 Nowcasting with Information from Baidu (full sample)

Total Retail Sales of Consumer Goods			
	Lags	***RMSFE***	***Reduction***
		OLS	
Baseline		0.0349	
Sum	L0	0.0324	0.0709
	L1	0.0352	-0.0070
	L2	0.0353	-0.0113
	L3	0.0352	-0.0085
PCA	L0	0.0296	0.1533
	L1	0.0319	0.0871
	L2	0.0308	0.1172
	L3	0.0268	**0.2314**
Best series	L0	0.0317	0.0919
	L1	0.0334	0.0446
	L2	0.0302	0.1358
	L3	0.0307	0.1214
Top 3 series	L0	0.0286	0.1803
	L1	0.0309	0.1156
	L2	0.0482	-0.3821
	L3	0.0264	**0.2451**
		LASSO	
Baseline		0.0352	
Sum	L0	0.0320	0.0920
	L1	0.0330	0.0619
	L2	0.0332	0.0556
	L3	0.0326	0.0751
PCA	L0	0.0263	0.2534
	L1	0.0292	0.1698
	L2	0.0289	0.1793
	L3	0.0248	0.2968
Individual Factors	L0	0.0260	0.2623
	L1	0.0259	0.2653
	L2	0.0219	0.3770
	L3	0.0203	**0.4228**

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 4.13 Forecasting with Information from Baidu (full sample)

Total Retail Sales of Consumer Goods			
	Lags	***RMSFE***	***Reduction***
		OLS	
Baseline		0.0466	
Sum	L0	0.0420	0.0994
	L1	0.0444	0.0483
	L2	0.0451	0.0314
	L3	0.0448	0.0381
PCA	L0	0.0369	0.2083
	L1	0.0410	0.1202
	L2	0.0341	0.2685
	L3	0.0298	0.3602
Best series	L0	0.0351	0.2469
	L1	0.0335	0.2813
	L2	0.0325	0.3017
	L3	0.0297	0.3637
Top 3 series	L0	0.0251	0.4609
	L1	0.0241	0.4825
	L2	0.0222	**0.5238**
	L3	0.0225	0.5162
		LASSO	
Baseline		0.0435	
Sum	L0	0.0396	0.0884
	L1	0.0409	0.0588
	L2	0.0407	0.0638
	L3	0.0382	0.1221
PCA	L0	0.0352	0.1912
	L1	0.0411	0.0536
	L2	0.0345	0.2062
	L3	0.0310	0.2876
Individual Factors	L0	0.0421	0.0304
	L1	0.0297	0.3174
	L2	0.0296	0.3194
	L3	0.0286	**0.3415**

NOTE: L stands for the number of lagged Baidu series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. Sum includes the sum of Baidu series as an additional variable to the baseline model. PCA adds the first 8 principal component. Best series adds the Baidu series that gives the highest adjusted R square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

4.9 Reference

Askitas, N., & Zimmermann, K. (2009). Google Econometrics and Unemployment Forecasting. *Institute of Labor Economics (IZA) Working Paper.* No. 4201.

Armstrong, P. (2016). Why your business should be using Google Trends. *Forbes.* Retrieved 27 December 2020 from https://www.forbes.com/sites/paularmstrongtech/2016/04/01/why-your-business-should-be-using-google-trends/?sh=129725d98268.

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can google data improve the forecasting performance of tourist arrivals? mixed-data sampling approach. *Tourism Management (1982), 46*, 454-464.

Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in google trends data - an analysis of the effects on tourism demand forecasting in germany. *Tourism Management (1982), 75*, 1-12.

Bram, J., & Ludvigson, S. C. (1998). Does consumer confidence forecast household expenditure? A sentiment index horse race. *Economic Policy Review*, *4*(2).

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why?. *The American Economic Review*, *84*(5), 1397-1408.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2-9.

Cotsomitis, J. A., & Kwan, A. C. (2006). Can consumer confidence forecast household spending? Evidence from the European commission business and consumer surveys. *Southern Economic Journal*, 597-610.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33*(4), 801-816.

Dees, S., & Brinca, P. S. (2013). Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area. *International Economics*, *134*, 1-14.

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, *48*(11), 87-92.

Fang, J., Gozgor, G., Lau, C. K. M., & Lu, Z. (2020). The impact of Baidu Index sentiment on the volatility of China's stock markets. *Finance Research Letters*, *32*, 101099.

Fang, J., Wu, W., Lu, Z., & Cho, E. (2019). Using Baidu index to nowcast mobile phone sales in China. *The Singapore Economic Review*, *64*(01), 83-96.

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment?. *Economic Modelling*, *30*, 117-125.

Gausden, R., & Hasan, M. S. (2018). An assessment of the contribution of consumer confidence towards household spending decisions using UK data. *Applied Economics*, *50*(12), 1395-1411.

Ginsberg, J., Mohebbi, M., Patel, R. et al. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457, 1012–1014.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National academy of sciences*, *107*(41), 17486-17490.

Guzman, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of Economic and Social Measurement*, *36*(3), 119-167.

Hand, C., & Judge, G. (2012). Searching for the picture: Forecasting UK cinema admissions using google trends data. *Applied Economics Letters, 19*(11), 1051-1055.

Hastie, T., Tibshirani, R. and Wainwright, M., 2019. *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.

Howrey, E. P. (2001). The predictive power of the index of consumer sentiment. *Brookings Papers on Economic Activity*, *2001*(1), 175-207.

Huang, X., Zhang, L., & Ding, Y. (2017). The Baidu Index: Uses in predicting tourism flows–A case study of the Forbidden City. *Tourism Management*, *58*, 301-306.

Juhro, S. M., & Iyke, B. N. (2020). Consumer confidence and consumption expenditure in Indonesia. *Economic Modelling*, *89*, 367-377.

Jun, S., Yoo, H. S., & Choi, S. (2018). Ten years of research change using google trends: From the perspective of big data utilizations and applications. *Technological Forecasting & Social Change, 130*, 69-87.

Krugel, L. & Viljoen, C. (2019). Post-Black Friday analysis. PWC. Retrieved 27 December 2020, from https://www.pwc.co.za/en/press-room/post-black-friday-analysis.html

Kwan, A. C., & Cotsomitis, J. A. (2006). The usefulness of consumer confidence in forecasting household spending in Canada: A national and regional analysis. *Economic Inquiry*, *44*(1), 185-197.

Lahiri, K., Monokroussos, G., & Zhao, Y. (2016). Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics*, *31*(7), 1254-1275.

Li, S., Chen, T., Wang, L., & Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management*, *68*, 116-126.

Liu, Y., Tseng, F., & Tseng, Y. (2018). Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model. *Technological Forecasting & Social Change, 130*, 123-134.

Medeiros, M.C. and Pires, H.F., (2021). The Proper Use of Google Trends in Forecasting Models. ResearchGate. Retrieved from https://www.researchgate.net/publication/350720260_The_Properly_Use_of_Google_Trends_in_Forecasting_Models.

Mihaela, S. (2020). Improving unemployment rate forecasts at regional level in romania using google trends. *Technological Forecasting & Social Change, 155*, 120026.

Morris, D., 2012. Google Searches Central Banks' Latest Economic Tool. The Financial Post. Available at: https://financialpost.com/news/economy/google-searches-central-banks-latest-economic-tool.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, *130*, 114-122.

Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*, *19*(6), 648-660.

Park, S., Lee, J., & Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing*, *34*(3), 357-368.

Shen, D., Zhang, Y., Xiong, X., & Zhang, W. (2017). Baidu index and predictability of Chinese stock returns. *Financial Innovation*, *3*(1), 4.

Sun, S., Wei, Y., Tsui, K. L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management, 70*, 1-10.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 5*8(1), 267-288.

Vaughan, L., & Chen, Y. (2015). Data mining from web search queries: A comparison of google trends and baidu index. *Journal of the Association for Information Science and Technology*, *66*(1), 13-22.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of forecasting*, *30*(6), 565-578.

Woo, J., & Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting*, *38*(2), 81-91.

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. *In Economic analysis of the digital economy* (pp. 89-118). University of Chicago Press.

Chapter 5: Predicting New Zealand Consumption with Google Trends

5.1 Introduction

The literature that explored the use of Google Trends to predict economic indicators so far has focused on a limited number of countries. There are studies for the US (Choi & Varian, 2012; Vosen & Schmidt, 2011; Choi & Varian, 2012; Woo & Owen, 2019), Chile (Carrière-Swallow & Labbé, 2013), Germany (Vosen & Schmidt, 2012), and China (Song, 2021, chapter 4 of this thesis). Although these studies typically conclude that internet search volume series can be used to produce more accurate predictions, it remains to be determined if these results on Google Trends can be generalized to a wider set of countries.

In the previous chapter of this thesis, I explored the utilization of Internet search volume data from Baidu Index and Google Trends in making predictions of Chinese consumption aggregates. Although my results concluded that Baidu Index increases prediction accuracy on both sectoral and total retail sales of consumer goods in China, this increase is not evident when using Google Trends. The low market share of Google in China can be the reason why Google Trends augmented prediction models didn't improve prediction accuracies in the Chinese context.

In this chapter, I focus on New Zealand. New Zealand is a small country, and hence produces a relatively limited search volume. Multiple machine learning techniques like OLS and Lasso will be used in order to fully exploit any information embedded in Google Trends to predict New Zealand household consumption.

Consumption data used in this chapter are from Statistics New Zealand for the period 2005 Q1 to 2020 Q4. These data have a 1-quarter publication delay. However, search volume data from Google are available on a daily basis. The timeliness of Google Trends data thus can contain information that is not in the lagged consumption data. In this chapter, I show that models with Google Trends reduce prediction errors by 18% for nowcasting and up to 45% for forecasting over a standard OLS model with AR terms.

5.2 Background and literature review

Many organizations in New Zealand, both in the public sector, like the Reserve Bank of New Zealand (RBNZ), the Treasury, and in the private sector like ASB, ANZ or BNZ,[13] make

[13] These forecasts can be found here:
RBNZ: https://www.rbnz.govt.nz/monetary-policy/monetary-policy-statement
Treasury: https://www.treasury.govt.nz/publications/efu/budget-economic-and-fiscal-update-2021

regular forecasts of major economic indicators such as GDP and its components. Detailed information about the forecasting models of private organizations were mostly unavailable, so in this section, I will give an overview of the forecasting models used by RBNZ, followed by some forecasting progress in both the public and the private sector.[14]

At the heart of New Zealand's monetary policy, the RBNZ has used several economic models to make forecasts of the economy. These models have greatly evolved over the years. Bloor (2009) described some of the statistical models that are being used by the RBNZ and how forecasts are produced. In the 1970s, the RBNZ calculated their forecasts based on systems of equations based on a series of empirically-based macro-economic models.

General equilibrium modelling called Forecasting and Policy Systems (FPS) was introduced between 1997 and 2009, taking the place of the equation-by-equation modeling method previously used. These modelling approaches traded off empirical fit and theoretical principles. Later on, Kiwi Inflation-Targeting Technology (KITT) was introduced. KITT took into account multiple sectors of the economy as well as structural economic shocks. KITT also enabled model parameters to be estimated from data, rather than being calibrated by model-builders. KITT was praised for its comprehensive ability to model New Zealand's economy. However, its complexity made it difficult to draw intuitive forecasts (Austin & Reid, 2017).

The RBNZ currently uses a Dynamic Stochastic General Equilibrium model (DSGE) which is called The New Zealand Structural Inflation Model, or NZSIM for short. NZSIM is a theory-based model that incorporates micro-economic agents to represent behaviors of each group in the economy. The goal of this model is to capture important dynamics in the New Zealand economy, while keeping the model simple and tractable. The core of NZSIM represents the economy using optimizing economic agents; namely, households, domestic firms, and import-distributing firms. These agents maximize their utility subject to constraints. NZSIM captures the interaction between these economic agents while taking into account equations that capture important empirical relationships that the Reserve Bank desires to consider. The

ASB: https://www.asb.co.nz/documents/economic-research/quarterly-economic-forecasts.html
ANZ: https://www.anz.co.nz/about-us/economic-markets-research/economic-outlook/
BNZ: https://www.bnz.co.nz/institutional-banking/research/publications#publication-Markets_Outlook

[14] This chapter focuses more on the forecasting practice by RBNZ, because the RBNZ make the details of their forecasting models available to the public. Details of the forecasting models used by private organizations are mostly unavailable. So I didn't write about the private organizations in detail.

assumptions of the NZSIM are updated regularly, and the forecasting system is reviewed at least every five years.[15] The technical details of the NZSIM model can be found in Kamber et al. (2014).

In addition to NZSIM, there are resources on official websites about some of the indexes used in the forecasts, as well as some articles and working papers published by the Reserve Bank and Treasury, which shed light on the calibration of their forecasts. On the forecasting section of the Reserve Bank of New Zealand, a list of indicators is presented as examples of how economic indicators can be used as a source of information for future economic conditions.[16] Specifically, two indicators are introduced and plotted against consumption by RBNZ: (i) the Westpac Trust-McDermott Miller Consumer Confidence Index (CCI), and (ii) Retail Sales of economically significant retailing enterprises, published by Statistics New Zealand.

Indeed, Consumer Confidence Indexes are historically associated with predictions of consumption. Details on how this strand of literature evolved can be found in Chapter 4 of this thesis. Carroll, Fuhrer & Wilcox (1994) found that lagged values of the consumer sentiment index explained about 14 percent of the variations in the growth of total real personal consumption. Other early studies followed by using CCI to explain private consumption and make predictions in the US. (Bram & Ludvigson, 1998; Howrey, 2001; Lahiri, Monokroussos & Zhao, 2015). While initial studies focused on the US, other countries have also been investigated. This international evidence includes Kwan & Cotsomitis (2007) for Canada, Gausden & Hasan (2018) for the UK, Juhro & Lyke (2020) for Indonesia, and Dees & Brinca (2013), using European data (as well as US). The overall conclusion from the literature is that survey-based indicators like consumer confidence or consumer sentiment improve the forecasting accuracy of private consumption.

Some working papers from the New Zealand Treasury also investigated using Consumer Confidence to forecast consumption expenditure using NZ data. A study by Goh (2003) was conducted in an earlier period, and it concluded that consumer confidence in New Zealand

15 An introduction of the NZSIM can be found at https://www.rbnz.govt.nz/research-and-publications/research-programme/additional-research/nzsim-our-macroeconomic-model
16The forecasting section of the Reserve Bank of New Zealand can be found at the following link: https://www.rbnz.govt.nz/challenge/team-resources/forecasting

doesn't contain a lot of information on consumer expenditure, and therefore adds very little forecasting value to consumption.

In addition to predicting consumption, other papers from RBNZ and Treasury also looked at the potential of using other data and estimation methods to boost the accuracy of the economic forecasts made by RBNZ. Some economists focused on the use of other important economic predictors. Matheson (2006) is the first to try using a wide range of predictors in a dynamic factor model to forecast GDP, inflation, interest rates, and exchange rates for New Zealand. He found that this model performs well in long horizons, outperforming the forecasts from the RBNZ. Other papers followed by looking at other potential predictors. Krippner & Thorsrud (2009) looked at the use of yield curve information in forecasting the New Zealand economy. They used various OLS and VAR model specifications and found that both in-sample fit and out-of-sample predictions are improved when yield curve data are incorporated.

Progress has also been made by using more advanced estimation techniques. McDonald & Thorsrud (2010) used several weighting methods to combine the forecasts of a number of economic models and compared the performance with the forecasts published by the RBNZ. They used past forecasts to construct model densities and found that the weighted model makes comparable or more accurate forecasts of the New Zealand economy. Eickmeier & NG (2011) used shrinkage estimation methods like principal components, elastic net, and ridge regression to forecast New Zealand GDP with big data sets that contain domestic and international data series on prices, monetary and financial activities. They found that the use of large datasets with shrinkage models is able to substantially improve forecasts of New Zealand GDP growth. Most recently, Richardson et al. (2021) studied the used a wide range of machine learning algorithms to test if vintage data of more than 600 economic indicators can be used to improve forecasts of the Reserve Bank. They used machine learning algorithms like Lasso, Ridge, elastic net, and neural networks to produce forecasts of the economy. They compared the forecasting accuracy of these models to a simple autoregressive model and found that when incorporating economic indicators, these models show significant improvement over a simple autoregressive model. Furthermore, they concluded that their models can increase the forecasting accuracy of the official forecasts released by RBNZ.

Progress has also been made in private organizations in predicting New Zealand economic growth, including the creation of timely indicators by ANZ and the live nowcasting

AI developed by Massy University. ANZ produced timely indexes on traffic volume as a proxy for economic activity across New Zealand. These indexes named "Truckometer" show a strong correlation with GDP, and they are used to match with GDP growth. ANZ also makes monthly reports with these indexes[17].

The abundant sources of data as well as the development of machine learning methodology also enabled the potential for real-time GDP nowcasting. GDPlive[18] is the world's first GDP nowcasting system developed and used in New Zealand to nowcast GDP in real-time. GDPlive utilizes high frequency and timely industry data with key economic variables, and combined this data with diverse machine learning algorithms. Specifically, they partnered with several industry giants to acquire timely data on financial transactions, import and export shipping container movements, KiwiRail shipments, as well as traffic volume, and used these data to produce predictions of GDP using recent advancements in machine learning algorithms like XGBoost, Light GBM, and Kernel Ridge Regression. These models are set up to produce daily nowcasts of GDP growth and are trained on a daily basis. Overall, their results are encouraging, although some predictions can be poor.

In terms of the use of Internet search engine data like Google Trends, Haworth et al. (2018) used Google Trends data to investigate how market participants respond to risk events related to insurers and bank stakeholders. Apart from that, no paper has been published on the use of internet search engine data in forecasting the New Zealand economy.

International evidence of search engine data in nowcasting and forecasting a wide range of economic activities have been carefully reviewed in chapter 4 of this thesis. In this chapter of my thesis, I contribute to the existing literature in exploring the use of such data in a New Zealand scenario by adopting a wide range of modeling techniques.

For many years Google has been the most popular search engine in New Zealand. FIGURE 5.1 shows the search engine market share in New Zealand between Jan. 2010 and Dec. 2020. As the figure suggests, Google has been the most successful search engine, occupying more than 90% of the market in New Zealand. This stable and prominent market share of Google

17 These reports can be found here: https://www.anz.co.nz/about-us/economic-markets-research/truckometer/
18 Details of GDPlive can be found here: https://www.gdplive.net/

Trends, means that there's potential for both the government and private organizations to make use of such data sources in analyzing consumer choice and economic activities.

5.3 Data and Methodology

5.3.1 Data

The main goal of this chapter is to study whether internet search volume can be used to improve both nowcasting and forecasting performance of private consumption in New Zealand. Quarterly seasonally adjusted consumption data [19] published by Statistics New Zealand between 2005 – 2020 is used to facilitate this objective. This data can be found at http://infoshare.stats.govt.nz/.

FIGURE 5.2 shows the seasonally adjusted real household final consumption expenditure in New Zealand. As the figure suggests, despite setbacks during the global financial crisis, this figure demonstrates strong growth in the past decade. This growth has slowly moderated and stabilized in recent years. There is no seasonality in the data because it has already been seasonally adjusted. As a result, in addition to the AR term, only time trends and are included in my nowcasting and forecasting models.

5.3.2 Collection of search query data

The internet search volume series used in this chapter are based on categories data from Google Trends. As a unique function of Google Trends, the searches on Google Trends allow users to limit search terms into a specific category, and when you select one of these categories without listing a specific keyword, the data reflect the aggregate search trends of the category that you've chosen.

Some of the detailed features of this category option are listed in chapter 3 of this thesis. Specifically, to get a sense of how these 'categories' work, for example, if one selects the category "automobile and cars," Google Trends will aggregate the data for all searches that fit this category. There are also sub-categories under some of the categories, for example, if you select the category "Auto & Vehicles," there are further sub-categories that allow you to limit search volume into a sub-category, like brands of automobile like "Mercedes," "Toyota"

19 Data used in the last chapter for China are not seasonally adjusted so I included month dummies. Stat NZ has seasonally adjusted consumption data, so I used it here. Using seasonally adjusted data also means the models are less complex.

or types of automobiles like "SUVs" or "city cars." I included all of these categories and used them to nowcast and forecast consumption. To match the quarterly data of consumption, I took the average of Google Trends data for each quarter to aggregate it into quarterly intervals.

One thing to note is that, by using Google 'categories' data, the problem I mentioned in the previous chapter, documented by Medeiros and Pires (2021), is likely to be mitigated. Medeiros and Pires (2021) showed that each time one looks up a search term on Google Trends, they are likely to get slightly different results for the less popular search terms. However, this is not likely to be a problem when Google Trends are being used to predict consumption of New Zealand in this thesis. As the data I use are the aggregated Google Categories data, not individual keywords.

In my previous chapter, I aggregated some series by simply summing the data from Baidu Index. However, this is not possible for Google Trends as Google Trends reports relative search volumes rather than absolute values. For this reason, I didn't use all of the specifications as I previously did in chapter 4. Specifically, I didn't include any models that use the sum, because summing Google Trends doesn't make sense.

Similar to the structure of chapter 4, I estimate both the baseline models and the models augmented with Google Trends term series, using both OLS and Lasso methodologies, and searching over various specifications to find the model that gives the most accurate nowcasts and forecasts.

5.3.3 Baseline models

The models used in this chapter follow a similar structure as chapter 4. In this chapter, I produce both nowcasts and 1-quarter ahead forecasts of consumption. Nowcasting aims to predict the value of the current quarter, while 1-quarter ahead forecast aims to predict the consumption of the next quarter.

To see this, the baseline model for nowcast and 1-quarter ahead forecast are:

Nowcast:

$$C_t = \alpha C_{t-1} + \beta_1 Date + \varepsilon_t \qquad (5.1)$$

Forecast:

$$C_t = \alpha C_{t-2} + \beta_1 Date + \varepsilon_t \qquad (5.2)$$

where C_t is the seasonally adjusted household final consumption expenditure in real terms at time *t*; and $Date$ is the time trend. Note that t stands for the end of each period.

The forecast horizon is between Q1 2017 to Q4 2020. To do this, I calculate expanding window nowcasts and forecasts with the baseline models above, where an additional observation is included to train the model as the iteration moves from one time period to the next. For example, when I make a nowcast for the time period *t*, I use data before *t* to train the model. But when I make a prediction for *t+1*, data for *t* is added into the training period to train the model. In the first model, I use data between Q1 2005 to Q4 2017 to train the model while predicting the consumption of Q1 2018.

One thing to note here is that because of the COVID pandemic in 2020, consumption in New Zealand took a major hit in 2020. This is evident in FIGURE 2. The simple model used in this thesis will not be able to pick up this consumption shock. Nonetheless, I do check whether Google Trends can be useful in extreme situations like COVID. To compare the predictability of Google Trends before and after the pandemic era, I study the performance of the prediction models separately for the pre-pandemic years (2018, 2019), and after the pandemic (2020).

I use the same method as chapter 4 to measure the performance of the models, by calculating RMSFE (Root Mean Square Forecasting Error). RMSFE is calculated each time after I run the expanding window nowcasts and forecasts. It is the standard deviation of the prediction errors, and is calculated as follows:

$$RMSFE = \sqrt{(f - o)^2} \qquad (5.3)$$

where *f* is the prediction, and *o* is the observed value.

To investigate the added value of including information from Google Trends, I will augment these baseline models with search volume data. In the following section, I discuss how the search volume data are incorporated into the models.

5.3.4 OLS estimations

The equations below show Google Trends series augmented nowcasting and forecasting models.

Nowcast:

$$C_t = \alpha C_{t-1} + \beta_1 Date + \beta_2 Google_t + \varepsilon_t \qquad (5.4)$$

Forecast:

$$C_t = \alpha C_{t-2} + \beta_1 Date + \beta_2 Google_{t-1} + \varepsilon_t \qquad (5.5)$$

In the models above, the lags of C_t, $Date$ are the same as the baseline model in Equations 5.1 and 5.2, while $Google_t$ stands for the different specifications of Google Trends that are incorporated into the models.

Due to the large quantity of Google Trends terms, several shrinkage estimation methods are used in this chapter to make sure there are enough degrees of freedom to estimate the prediction models. To be exact, the following methods are used.

Firstly, a screening process is adopted to pre-select the Google Trends series having a correlation coefficient with consumption that is larger than 0.9. Principal component analysis is then used to convert the selected series into factor loadings. These factor loadings are then used to augment the baseline models[20].

Secondly, I follow a procedure similar to Ginsberg et al. (2009) and run a regression with each Google series separately and find the series that individually adds the most to the baseline model during the training period. To illustrate, I run the following OLS model:

$$C_t = \alpha C_{t-1} + \beta_1 Date + \beta_{13} Single\ \ Google\, Series + \varepsilon_t \qquad (5.6)$$

Where each of the individual series is included in the model along with the baseline variables. The series that gives the highest adjusted R square in the training sample is selected and added to the baseline model to nowcast and forecast the consumption of the next period. I iterate this procedure using the expanding window method, re-selecting at each stage the

20 Similar to the previous chapter, the first 3 factors are used. I find that usually less factors like 3 do better in predicting than a lot of factors like 8 or 9. Less factors also means our models are estimable when we have small degrees of freedom.

series with the highest adjusted R square for each time period. In addition, I use this method to choose the three series that, individually, gives the highest adjusted R square, and then evaluate the RMSFE of a model that includes these three terms together.

In addition to using only the contemporaneous values of the Google Trends series, lagged series of Google Trends are also incorporated into the models, allowing up to three lags for nowcasting models and up to four lags for forecasting models.

Similar to the previous chapter, I added 1 and 3 individual series to explore how sometimes limited number of search terms can result in significant improvements in prediction results. I also included different lags of Internet search series is to explore how past values can be used to inform a more accurate forecast. Potentially more lags and more series can be included in the models, but the results are not included in this thesis for simplicity.

5.3.5 Lasso estimation

Lasso models will also be used to run expanding window predictions. Detailed descriptions for the mechanism of Lasso models as well as some of its limitations are listed in chapter 4 of this thesis.

The lasso specifications include (1) Lasso models with PCA factors of Google Trends and (2) Lasso models that include all the Google Trends series. In addition, I also experiment with models that add lagged terms of the Google Trends into the models.[21]

As Lasso allows for model selection when facing a large number of variables, I interacted Google Trends series with the baseline variables, which are the lagged dependent variable and the time trend variable, and run the models with these interaction terms to see if the performance of the models can be further improved.

21 Because up to 3 lags of Google Trends will be included in the nowcasting model and up to 4 lags of Google Trends will be included in the forecasting model, to keep my data comparable, in all of the models in the empirical analysis I exclude the first 3 time periods for nowcasting and excluded the first 4 periods for forecasting.

5.4 Empirical results

5.4.1 Nowcasting results: 2018-2019

In TABLE 5.1, I list the nowcast RMSFE of consumption using different methodologies and specifications. The nowcasting horizon is between Q1 2018 and Q4 2019[22]. The tables are structured in a similar fashion as the results in chapter 4, with the top panel showing results for the OLS models and the bottom panel showing the results of the Lasso models. In addition to the absolute RMSFE, the reduction of RMSFE relative to the RMSFE of the baseline model (OLS or Lasso) is also listed. When the reduction of RMSFE is positive, it indicates the percentage improvement in predictive performance. If this number is negative, it means adding Google information decreases predictive performance.

The top of TABLE 5.1 shows that including the PCA of Google Trends series into the baseline model can improve nowcasting performance for consumption. When 3 lags of the Google Trends series are included in the model, this specification improves nowcasting accuracy the most. With the absolute RMSFE of around 150, this specification reduces 18.09% of the baseline nowcasting error. This model is also the overall best performing nowcasting model.

Instead of aggregating the Google series, I then analyze the nowcasting results when I add the individual series that gave the highest adjusted R square during the training period. According to TABLE 5.1, when only the series with the highest adjusted R square is included for each training period, I do not see a reduction in RMSFE for consumption. I get a similar result when all 3 individual series with the highest adjusted R square are included. In all cases, adding keywords with the highest adjusted R square seems to make my nowcasting performance worse.

The results of the Lasso models are shown in the bottom panel of Table 5.1. When I used the Lasso models with PCA factors for nowcasting, I find that as more lags of Google Trends are involved, the performance of the model is improved. In most cases, the inclusion of Google Trends PCA factors is able to reduce nowcasting errors by between 17% and 20%.

As for the models that add Google Trends series jointly into the models, I find that this is able to reduce between 5% and 8% of the nowcasting errors. The lasso model with interaction terms of Google Trends reduces between 4% and 10% of the nowcasting errors.

22 This is also the forecasting horizon of the Baidu chapter (For the Baidu Chapter its monthly).

Overall, for the Lasso models, I find that including the Google Trends series almost always improves the nowcast, and that 2 lags of the PCA of Google Trends into the model improves the accuracy the most for the Lasso models, reducing the RMSFE by 20.14% compared to the baseline Lasso model. Note that the overall best performing model is still the OLS model with 3 additional lags of PCA factors, reducing 18.09% of the nowcasting error compared to the OLS baseline model.

5.4.2 Forecasting results: 2018-2019

TABLE 5.2 shows the RMSFE and the reduction in RMSFE of the forecasting models that incorporate the Google Trends series. The forecasting horizon is between 2018 and 2019. This table is structured similarly to table 1, with the top panel showing the forecasting results for OLS models while the bottom panel showing the results for Lasso models.

Similar to the results from the OLS nowcasting models, I again find that adding PCA of Google Trends to the baseline model can improve prediction accuracy. If 3 additional lags of the Google Trends series are added, I improve the forecasting accuracy by 43.68%. This is a very significant increase in forecasting accuracy. I also find that when the individual series that has the highest in-sample adjusted R square is added to the model, forecasting accuracy is also improved.

As for the Lasso models, once again, I see improvements in forecasting accuracy across the table. Compared to the baseline Lasso model, the PCA models are able to reduce around 23% to 40% of the forecasting error. The Lasso models with individual factors can reduce up to 19% of the forecasting error, while the Lasso models with interaction terms can reduce up to 34% of the forecasting error. However, the best forecasting model is still the OLS model with 3 additional lags of PCA factors.

TABLE 5.1 and 5.2 show that the OLS models with PCA do really well and that these models have the lowest RMSFE in both nowcasting and forecasting. For nowcasting, the best model is the OLS model with 3 additional lags of PCA. It has an RMSFE of 150.1524, compared to the best baseline (the OLS baseline model) of 183.3139, an 18.09% improvement. For forecasting, the best model is the model with 3 lags of PCA. It has an RMSFE of 161.1721, compared to 286.1872 for the best baseline model (the OLS baseline model), which is a 43.68% improvement.

Summarizing my findings so far, the evidence suggests that both the nowcasting and forecasting of consumption in New Zealand can be improved using Google Trends. In most model specifications, I can conclude that adding such information will improve predictive performance. In addition, I find that the best models are relatively simple OLS models with the PCA factors of Google Trends. The model with Google Trends PCA factors can decrease 18.09% of the nowcasting errors and 43.68% of the forecasting errors.

5.4.3 Nowcasting and forecasting results: 2020

In 2020, the COVID pandemic hit, causing global shocks in demand and supply. Although New Zealand is one of the few countries that handled the pandemic well, a decrease in consumption was nonetheless evident. As Figure 5.2 shows, the pandemic, as well as the associated lockdown measures that were taken due to COVID, had a big impact on the economy, resulting in a major decrease in private consumption. Of course, due to its nature, a simple AR fails to capture this huge decrease in consumption. As a result, my predictions for 2020 are way worse compared to my predictions for the pre-pandemic era. Nonetheless, I report my results in TABLE 5.3 and TABLE 5.4, to see even if the results for 2020 are bad, does Google Trends still improves prediction accuracy compared to the baseline models.

TABLE 5.3 and TABLE 5.4 show the RMSFE and the reduction in RMSFE of the various models for 2020. The top panel shows the results for OLS models, while the bottom panel shows the results for the Lasso models.

For the nowcasting results listed in TABLE 5.3, I observe the following pattern: OLS models with the PCA factors of Google Trends can still improve nowcasting accuracy. When one additional lag of Google Trends PCA factors is included, nowcasting accuracy was improved 5.49%, although this improvement is much smaller than my results for the 2018-2019 period. When I focus on the other model specifications, I see that most of the other models are doing worse compared to the baseline models.

As for the forecasting results in TABLE 5.4, I see a similar result, with the OLS model with PCA factors reducing 12.65% of the forecasting error compared to the baseline OLS model. Apart from the PCA models, all the other model specifications produced worse results compared to the baseline models.

This reduction in the predictive ability of Google Trends may be due to COVID. For example, because of the COVID lockdown or traveling restrictions, people are more likely to stay at home. While staying at home will most likely not have a negative impact on the usage of Google. However, the travel restrictions might make it difficult for people to transfer their pre-shopping research into actually buying goods. Nonetheless, if this is indeed the case, there are no reasons to believe that after the COVID pandemic, the usefulness of Google Trends in increasing the prediction accuracy of consumption will still be affected.

Overall, for the period of 2020, my results indicate that the use of Google Trends can increase prediction accuracy, although this increase is smaller compared to the pre-pandemic period of 2018 and 2019.

5.5 Conclusion

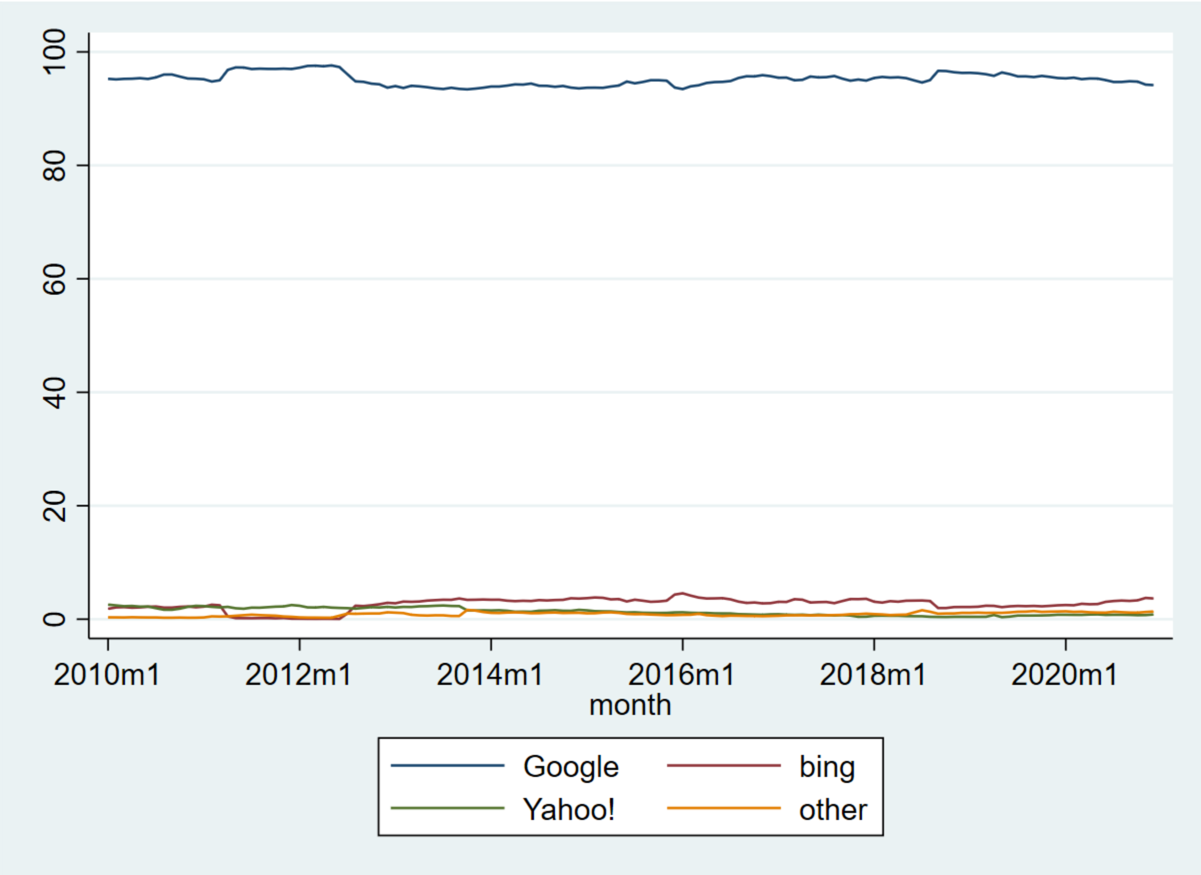
This chapter of my thesis complements my findings in chapter 4. I analyze the use of internet search intensity data from Google Trends to improve nowcasts and forecasts of private consumption in New Zealand.

My results are consistent with the previous chapter, and it shows clear evidence that incorporating search intensity data can improve the nowcasts and forecasts of consumption in the New Zealand context. These improvements can be substantial. In some of my specifications, using PCA factors of Google Trends can reduce predictive errors by 40% or more. My results indicate that simple models typically perform better than complicated models with a lot of Google Trends series, and that OLS models do a better job of nowcasting and forecasting than LASSO models. In addition, I conclude that Google Trends was able to improve prediction accuracy during the pandemic period of 2020, although the improvement during 2020 was smaller compared to the improvement between 2018 and 2019.

This chapter of my thesis is the first to look at the use of Google Trends in making economic forecasts in New Zealand. As many New Zealand government and private organizations make routine forecasts of the New Zealand economy, these organizations may find my results useful in improving their forecasting accuracy.

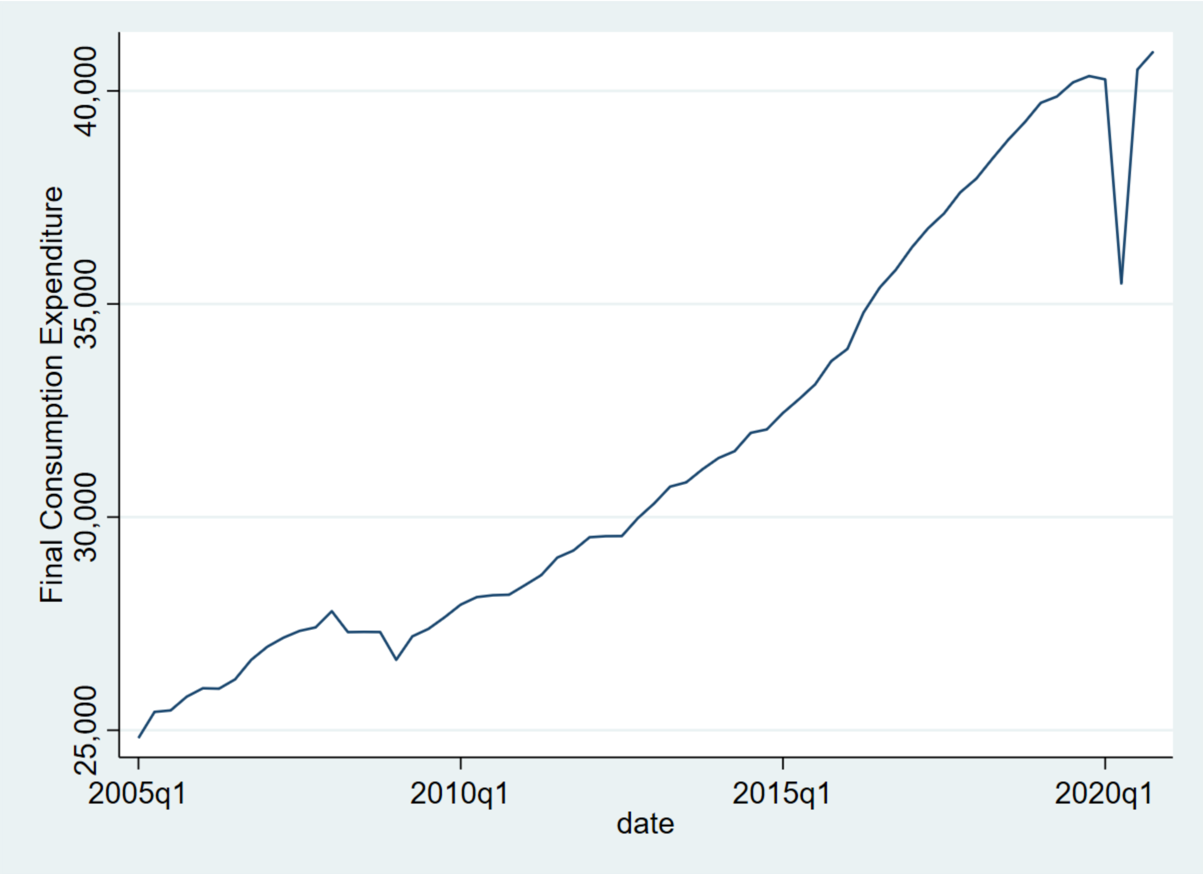
5.6 Appendix

Figure 5. 1 Search engine market share in New Zealand between Jan 2010 to Dec 2020



(Source: Statcounter GlobalStats 2021)

Figure 5. 2 New Zealand seasonally adjusted household final consumption expenditure in real term



(Source: Statistics New Zealand)

Table 5.1 Nowcasting with Information from Google, prediction horizon: 2018-2019

Household Consumption Expenditure			
	Lags	***RMSFE***	***Reduction***
OLS			
Baseline		183.3139	
PCA	L0	157.6904	0.1398
	L1	175.6344	0.0419
	L2	164.3943	0.1032
	L3	**150.1524**	**0.1809**
Best series	L0	195.6619	-0.0674
	L1	195.6619	-0.0674
	L2	205.8350	-0.1229
	L3	205.8350	-0.1229
Top 3 series	L0	236.1224	-0.2881
	L1	205.8127	-0.1227
	L2	201.6753	-0.1002
	L3	201.6753	-0.1002
LASSO			
Baseline		217.4005	
PCA	L0	288.6399	-0.3277
	L1	180.0941	0.1716
	L2	173.6071	0.2014
	L3	178.9010	0.1771
Individual Factors	L0	200.1600	0.0793
	L1	202.2199	0.0698
	L2	204.2373	0.0605
	L3	205.7503	0.0536
Interaction	L0	195.1760	0.1022
	L1	208.6059	0.0405
	L2	204.0918	0.0612
	L3	203.8579	0.0623

NOTE: L stands for the number of lagged Google series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. PCA adds the first 3 principal component. Best series adds the Google series that gives the highest adjusted R

square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 5.2 Forecasting with Information from Google, prediction horizon: 2018-2019

Household Consumption Expenditure			
	Lags	***RMSFE***	***Reduction***
OLS			
Baseline		286.1872	
PCA	L0	208.4512	0.2716
	L1	213.2462	0.2549
	L2	183.5229	0.3587
	L3	**161.1721**	**0.4368**
Best series	L0	281.1070	0.0178
	L1	281.1070	0.0178
	L2	281.1070	0.0178
	L3	281.1070	0.0178
Top 3 series	L0	257.0244	0.1019
	L1	239.8979	0.1617
	L2	239.8979	0.1617
	L3	239.8979	0.1617
LASSO			
Baseline		339.0225	
PCA	L0	260.5767	0.2314
	L1	214.2594	0.3680
	L2	237.3795	0.2998
	L3	202.6267	0.4023
Individual Factors	L0	332.7151	0.0186
	L1	301.3469	0.1111
	L2	275.6511	0.1869
	L3	315.3363	0.0699
Interaction	L0	328.5369	0.0309
	L1	273.2312	0.1941
	L2	223.9690	0.3394
	L3	263.2592	0.2235

NOTE: L stands for the number of lagged Google series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. PCA adds the first 3 principal component. Best series adds the Google series that gives the highest adjusted R

square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 5.3 Nowcasting with Information from Google, prediction horizon: 2020

	Lags	RMSFE	Reduction
Household Consumption Expenditure			
	OLS		
Baseline		3464.1840	
PCA	L0	3844.5320	-0.1098
	L1	**3274.0590**	**0.0549**
	L2	3301.5990	0.0469
	L3	3468.7980	-0.0013
Best series	L0	4135.4580	-0.1938
	L1	5445.3700	-0.5719
	L2	5445.3700	-0.5719
	L3	5445.3700	-0.5719
Top 3 series	L0	4845.2490	-0.3987
	L1	5327.1900	-0.5378
	L2	5329.1710	-0.5384
	L3	5061.6910	-0.4611
	LASSO		
Baseline		3534.3250	
PCA	L0	3545.0290	-0.0030
	L1	3478.0970	0.0159
	L2	3511.4600	0.0065
	L3	3468.9650	0.0185
Individual Factors	L0	3566.2700	-0.0090
	L1	4349.3000	-0.2306
	L2	4351.5510	-0.2312
	L3	4256.7760	-0.2044
Interaction	L0	3591.3320	-0.0161
	L1	3772.8670	-0.0675
	L2	3777.754	-0.0689
	L3	3595.536	-0.0173

NOTE: L stands for the number of lagged Google series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. PCA adds the first 3 principal component. Best series adds the Google series that gives the highest adjusted R

square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

Table 5.4 Forecasting with Information from Google, prediction horizon: 2020

Household Consumption Expenditure			
	Lags	***RMSFE***	***Reduction***
OLS			
Baseline		3598.4110	
PCA	L0	**3143.0370**	**0.1265**
	L1	3179.7500	0.1163
	L2	3178.2100	0.1168
	L3	3272.3980	0.0906
Best series	L0	4435.7760	-0.2327
	L1	4435.7760	-0.2327
	L2	4307.7720	-0.1971
	L3	4307.7720	-0.1971
Top 3 series	L0	4274.4900	-0.1879
	L1	4189.7760	-0.1643
	L2	4222.0490	-0.1733
	L3	4222.0490	-0.1733
LASSO			
Baseline		3647.4690	
PCA	L0	3596.6200	0.0139
	L1	3703.2990	-0.0153
	L2	3696.1170	-0.0133
	L3	3634.0620	0.0037
Individual Factors	L0	4491.2620	-0.2313
	L1	4277.6600	-0.1728
	L2	3726.9420	-0.0218
	L3	3818.8900	-0.0470
Interaction	L0	4307.3200	-0.1809
	L1	4278.9970	-0.1731
	L2	3676.7050	-0.0080
	L3	3688.0600	-0.0111

NOTE: L stands for the number of lagged Google series. L1 means both lags 0 and 1 are included. Forecasting models use one more lag than nowcasting models. PCA adds the first 3 principal component. Best series adds the Google series that gives the highest adjusted R

square in the training period. Top 3 series adds the 3 series that individually gives the highest adjusted R square in the training period. Individual factors adds all series separately, while interactions in addition interacts these separate series with the baseline variables.

5.7 Reference

Austin, N., & Reid, G. (2017). NZSIM: A model of the New Zealand economy for forecasting and policy analysis. *The Reserve Bank of New Zealand Bulletin*, 80(1), 1.

Bloor, C. (2009). The use of statistical forecasting models at the Reserve Bank of New Zealand. *The Reserve Bank of New Zealand Bulletin*, 72(2), 21.

Bram, J., & Ludvigson, S. C. (1998). Does consumer confidence forecast household expenditure? A sentiment index horse race. *Economic Policy Review*, *4*(2).

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why?. *The American Economic Review*, *84*(5), 1397-1408.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2-9.

Dees, S., & Brinca, P. S. (2013). Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area. *International Economics*, *134*, 1-14.

Eickmeier, S., & Ng, T. (2011). Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting*, 27(2), 496-511.

Gausden, R., & Hasan, M. S. (2018). An assessment of the contribution of consumer confidence towards household spending decisions using UK data. *Applied Economics*, *50*(12), 1395-1411.

Ginsberg, J., Mohebbi, M., Patel, R. et al. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457, 1012–1014.

Goh, K. L. (2003). *Does consumer confidence forecast consumption expenditure in New Zealand?*. *New Zealand Treasury Working Paper*. No. 03/22.

Haworth, C., Gillies, L., & Irrcher, T. (2018). Measuring Market Discipline in New Zealand. *Reserve Bank of New Zealand Discussion Paper.* No. AN2018/07

Howrey, E. P. (2001). The predictive power of the index of consumer sentiment. *Brookings Papers on Economic Activity*, *2001*(1), 175-207.

Juhro, S. M., & Iyke, B. N. (2020). Consumer confidence and consumption expenditure in Indonesia. *Economic Modelling*, *89*, 367-377.

Kamber, G., McDonald, C., Sander, N., & Theodoridis, K. (2015). A structural forecasting model for New Zealand: NZSIM. Reserve Bank of New Zealand.

Krippner, L., & Thorsrud, L. A. (2009). Forecasting New Zealand's economic growth using yield curve information. *Reserve Bank of New Zealand Discussion Paper*. No. DP2009/18.

Kwan, A. C., & Cotsomitis, J. A. (2006). The usefulness of consumer confidence in forecasting household spending in Canada: A national and regional analysis. *Economic Inquiry*, *44*(1), 185-197.

Lahiri, K., Monokroussos, G., & Zhao, Y. (2016). Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics*, *31*(7), 1254-1275.

Medeiros, M.C. and Pires, H.F., (2021). The Proper Use of Google Trends in Forecasting Models. ResearchGate. Retrieved from https://www.researchgate.net/publication/350720260_The_Properly_Use_of_Google_Trends_in_Forecasting_Models.

Matheson, T. D. (2006). Factor Model Forecasts for New Zealand. *International Journal of Central Banking*, 169–237.

McDonald, C., & Thorsrud, L. A. (2011). Evaluating density forecasts: model combination strategies versus the RBNZ. *Reserve Bank of New Zealand Discussion Paper*. No. DP2011/03.

Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, 37(2), 941-948.

Susnjak, T., & Schumacher, C. Nowcasting: Towards Real-time GDP Prediction. GDPLive. Retrieved from https://www.gdplive.net/static/media/GDP_Live_Working_Paper.63d13a68.pdf

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of forecasting*, *30*(6), 565-578.

Vosen, S., & Schmidt, T. (2012). A monthly consumption indicator for Germany based on Internet search query data. *Applied Economics Letters*, *19*(7), 683-687.

Woo, J., & Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting*, *38*(2), 81-91.

Chapter 6: Conclusion

In this thesis, I undertook four studies looking at how consumption and savings were affected by a wide range of factors, as well as how internet search engines can be used to better nowcasts and forecasts of consumption. My thesis is mainly structured around my attempts to understand the determinants and behaviour of consumption. I started my first step of this effort by replicating a well-cited paper on government consumption and private consumption, F&K. This was the focus of Chapter Two of my thesis.

F&K presented evidence on the extent of the "crowding out effect" between government and private consumption. To do this, they split government consumption into "Public goods" that are difficult to supply privately, such as defence, public order, and justice, and "Merit goods" such as health, education, and other services that can be provided privately. The motivation behind splitting government consumption into several categories stems from the fact that government consumption in different areas is likely to affect private consumption differently. F&K estimated their model using difference GMM for 12 OECD countries.

In my replication of F&K's paper, I followed their estimation process and attempted to reproduce their estimation results. I then extended their analysis using more recent data. In order to re-create their results, I first contacted the authors in an effort to obtain the original data. Unfortunately, the original data no longer exists. I then collected data using the same sources they identified in the statistical appendix to their paper.

In my replication study, I was able to match F&K's model specification and produce very similar results. I also confirm their results when I use more recent data to run the same model specifications. This is very impressive because I get results very close to what F&K obtained even when I use entirely new data from more recent years. My replication confirms that government spending tends to affect private consumption differently depending on the composition of government expenditure.

My next step focused on answering to what extent the One-Child Policy (OCP) was responsible for China's high savings rate. The question is of current interest because there is speculation about the extent to which relaxation of the OCP is likely to boost China's future consumption. Chapter Three of this thesis took up this subject. I found that although multiple papers conclude that the OCP contributed to China's high savings rate, they assumed that the only channel of effect of the OCP on savings behaviour was through the number of children. They ignored the fact that with everybody having fewer children, the nature of the relationship between children and savings might change. Ideally, to study how the OCP affects savings, comparisons should be made between Chinese households with and without OCP conditions. As the perfect counterfactual doesn't exist, my approach was to use other countries that have similar cultures and backgrounds as a counterfactual for China without the OCP.

By adopting a Blinder-Oaxaca decomposition procedure, I am able to disentangle the different channels by which OCP can affect savings. In particular, I can separate the impact of the OCP on savings via its effect on the number of children (the "endowment effect"), and the impact of the OCP on savings via its effect on the relationship between savings and children, holding the number of children constant (the "coefficient effect"). I estimate this model using data from the 2014 Gallup World Poll and the 2014 Global Findex database.

My analysis indicates that there's little difference in saving behavior between China and its "counterfactuals". The Blinder-Oaxaca decomposition shows that while children account for some of the variations in saving behaviour between China and its counterfactuals, the estimated effects are too small to explain the high rate of Chinese savings. As a result, my results suggest that the relaxation of the OCP will not provide a major boost to Chinese household consumption.

Afterwards, I turned my attention to the use of internet search volume data like Baidu Index and Google Trends to improve the forecasting accuracy of economic aggregates. In Chapter Four of this thesis, I investigated whether Baidu Index and Google Trends volume data could improve the nowcasting and forecasting of Chinese automobile and communication appliances sales. To do this, I collected total retail sales data from the Chinese Statistical Bureau from 2011 to 2019. These data have a 1-month publication delay. In contrast, search volume data from Baidu Index and Google Trends are available on a daily basis. These timely data incorporate information not included in lagged sales data.

I estimate both the baseline models and the models augmented with Baidu and Google search term series. Various specifications estimated with both OLS and Lasso methodologies were used to find the model that gave the most accurate nowcasts and forecasts. I find that Baidu Index-augmented models tend to perform better compared to the baseline models. The improvement was greater than that from using Google Trends or Consumer Confidence index-augmented models.

Chapter Five of this thesis looks at whether Google Trends can be used to predict quarterly consumption in New Zealand. This chapter is motivated by the fact that existing literature only focused on a handful of countries such as the US, Germany and Chile. While these studies conclude that Google Trends can be used to produce more accurate forecasts, more evidence is needed to determine if similar conclusions apply to other countries. In this chapter, I focus on the New Zealand economy while using search volume data from Google Trends to produce nowcasts and forecasts of New Zealand quarterly consumption. The consumption data used in this chapter were taken from Statistics New Zealand for the period 2005: Q1 to 2020: Q4. Multiple machine learning algorithms similar to Chapter Four were used to exploit the information incorporated in Google Trends. My results show that Google Trends-augmented models can reduce 18% of the nowcasting errors and up to 45% of the forecasting errors compared to the OLS baseline model with AR terms.

In conclusion, the research in this thesis provides insights into the general determinants of consumption and savings. I provide evidence on how government can potentially allocate government spending to stimulate private consumption by spending more on “merit goods” like health and education services. I find that the One-Child Policy in China didn’t play a major role in people’s saving decisions, and that the relaxation of OCP is unlikely to affect people’s spending behavior in the future. I also provide evidence on how internet search engine data can be used to generate more accurate nowcasts and forecasts for China and New Zealand. My results may be useful to policy-makers in understanding the different channels by which consumption can be affected, which in turn can provide insights into the evolution of the future economy.