Essays in Econometrics

Galen Matthew Sher

Submitted for the degree of Doctor of Philosophy in Economics
Michaelmas Term 2015

Department of Economics
Social Sciences Division
University of Oxford

Acknowledgements

I am heavily indebted to my supervisor, Professor Steve Bond, for investing countless hours over the last three years of my DPhil and one year of my MPhil. He taught me what it means to do applied econometrics and what research integrity looks like.

Thank you to Professor Bent Nielsen for encouragement and guidance over the years. I am grateful to my college advisors Christopher Bowdler and Rui Esteves for advising me well.

Thank you to the Oxford–Man Institute for Quantitative Finance for hosting me for three years. The excellent facilities and access to Mathematicians developed my quantitative skills and the research in this thesis.

I made many good friends at the Oxford–Man Institute who supported me. Thank you to Arnaud, Bary, Bertrand, Francesca, Lawrence, Mathias, Matthieu, Miha, Pedro, Vimal, Wande, Yoel and Yves–Laurent, to name a few.

The first chapter of this thesis emerged from an internship at the Bank of England. Thank you to Oliver Bush for introducing me to the topic of that chapter. Steve Bond provided extensive guidance, and the chapter also benefitted from comments from John Muellbauer, Simon Price and seminar participants at the Bank of England and University of Oxford. I alone am responsible for any errors or omissions.

The second chapter of this thesis emerged from an internship at the European Central Bank under the supervision of Christoffer Kok and Matthias Sydow. For comments I would like to thank Sylvain Benoit, Steve Bond, Alan Morrison, Marek Musiela, Hairui Zhang and seminar participants at the Department of Economics and Oxford–Man Institute at the University of Oxford, Bank of England, European Central Bank, 2015 European Winter Meeting of the Econometric Society at Bocconi University, 9^{th} International Conference on Computational and Financial Econometrics at the University of London, 4^{th} PhD Student Conference in International Macroeconomics and Financial Econometrics at Université Paris-Ouest, 15^{th} Finance, Risk and Accounting Perspectives Conference at the University of Applied Sciences Upper Austria. Thank you to Pascal Paul for many interesting discussions on this topic. Any errors that remain are my own.

The third chapter of this thesis is based on Sher and Vitoria [2013] and Pedro Vitoria provided many illuminating discussions. I received helpful comments from Steve Bond and Bent Nielsen and from seminar participants at the University of Oxford, Bank of England, 2014

Society for Financial Econometrics Summer School at Harvard University, 19^{th} International Conference on Computing in Economics and Finance in Vancouver and 7^{th} International Conference on Computational and Financial Econometrics at the University of London.

The Econometrics Working Group in the Department of Economics stimulated my interest in Econometrics and exposed me to a variety of topics. Thank you to some of my fellow DPhil candidates and friends in the Economics Department including Anthony, Belinda, Claudia, Cornelius, Dani, Fede, Felix, James, Nick, Nikki, Oleg, Pascal, Pawel and Rachel for sharing this experience with me.

The five years I spent at Oxford have been formative. They never would have been possible without the support I received from the Rhodes Scholarships, the Economics Department Doctoral Studentship and the Herbert and Ilse Frankel Memorial Studentship at Oriel College.

A very special thank you to Natalie for her encouragement, support and sense of humour over the past three years that made completing this research a lot easier.

Galen Sher

Abstract

Chapter 1

I revisit claims in the literature about the relationship between financial development and economic growth. I find that the Solow–Swan growth determinants are important omitted variables in the existing empirical literature. The role for credit in determining growth is attenuated once one controls for investment. The positive relationship between growth and credit in the literature is explained by a combination of the well known positive relationship between growth and investment and a positive relationship between investment and credit. However, there is evidence for a negative relationship between growth and credit at levels of credit above 90% of output and this relationship is not explained by any declining effect of credit on investment at these high levels of credit.

Chapter 2

I present a model for describing the dependence of financial intermediary net worth on the term to maturity structure of nominal interest rates. Such a model can be used to predict the change in net worth caused by a change in nominal interest rates. I show how to quantify the uncertainty in this change in net worth that arises from term structure parameter estimation and measurement error in accounting data. I apply the methods to measure the exposure of US banks to changes in nominal interest rates.

Chapter 3

I consider a measure of serial dependence in stochastic processes based on the compressibility of samples under modern compression algorithms. I relate the measure, which depends on the entropy rate of the stochastic process, to traditional measures of dependence, including those based on entropy like Hong and White [2004]. I present numerical evidence on the finite and large sample distributions of the measure under serial independence, that can be used to choose critical values for hypothesis tests for dependence.

289 words

Contents

Introduction

The title of this thesis is quite general, reflecting the fact that its three chapters are quite distinct from one another. The first chapter contributes to the economics literature on the empirical relationship between financial development and long-run economic growth, the second chapter contributes to the finance and economics literature on measuring the risk of changes in nominal interest rates for financial intermediaries and the third chapter contributes to the statistics literature on measuring dependence in a random process using the entropy rate of that process. In the acknowledgements, I explain the genesis of each chapter. Despite their differences, these chapters are not completely distinct. All three chapters are exercises in measurement. The first chapter measures the relationship between credit, a particular measure of financial development, and long-run economic growth, using data across countries and time periods and using regression models that control for important, unobservable characteristics of long-run economic growth. The second chapter shows how to measure the risk of changes in nominal interest rates on the net worth of a financial intermediary and how to quantify the uncertainty in this risk measure due to uncertainty in parameter estimation and accounting data. The third chapter shows how to measure the dependence in a sample from a random process and how to quantify the uncertainty in this measure of dependence. The first and second chapters both relate to the role of the financial sector in macroeconomics, which is a literature that has gained prominence in the last eight years since the start of the global financial crisis. While the first chapter is concerned with the effect of the financial sector on the broader macroeconomy, the second chapter addresses an effect in the reverse direction. In particular, the second chapter focuses on the effect of a change in the structure of nominal interest rates, a particular aspect of the macroeconomy, on the financial sector. Another commonality among the three chapters is that they all emerged from applied work. While the first chapter primarily applies existing theory, the second and third chapters offer some methodology of their own. Nevertheless, even the second and third chapters emerged from a need to support applied work. I prefer my research to be practically useful, so the second and third chapters also contain application sections.

Finance, investment and growth

The first chapter revisits the literature on the evidence for a relationship between financial development and long-run economic growth. I compile a dataset from modern versions of the original data sources used by papers in the literature, which allows me to have more confidence in the data than if I had reused data from these papers.[1] With this compiled dataset aggregated into 5-year non-overlapping windows, I reassess claims in the literature about the relationship between financial development, as measured by the ratio of the nominal stock of credit by non-central bank intermediaries to nominal gross domestic product ('credit'), and real per capita growth in gross domestic product ('growth').

In particular, I revisit the finding by Beck et al. [2000] and Levine et al. [2000] of a positive log-linear relationship between credit and growth, which the authors show does not seem to be driven by any unmeasured time-invariant country-specific factors affecting growth or by endogeneity in the explanatory variables. This finding is doubtful for at least three reasons. First, it relies on standard errors that are known to be biased down in small samples. Second, it is based on constant price US dollar measures of output that are not comparable across countries because of purchasing power differences between them. Third, it is based on specifications that omit the Solow–Swan growth determinants, namely the expenditure share of investment in output and the rate of population growth. In particular, we may be concerned that the relationship between credit and growth is proxying for the well known relationship between investment and growth.

Using standard errors that are corrected for finite-sample bias and various measures of output from the Penn World Tables that are adjusted for purchasing power differences across countries, I find that the positive log-linear relationship between credit and growth prevails. However, the positive log-linear relationship between credit and growth becomes attenuated and statistically insignificant in the presence of an investment control variable. Combined with findings of statistically significant positive log-linear relationships between credit and investment and between investment and growth, and under the maintained identifying assumptions of the specifications, this suggests that any positive effect of credit on growth found in earlier work is proxying for a combination of a positive effect of credit on investment and the well-known positive effect of investment on growth.

In addition, I revisit the finding by Arcand et al. [2012] of a non-monotonic concave relationship between credit and growth. In the first chapter I explain five reasons for doubting this finding, but in this introduction I mention just three. Like Beck et al. [2000], the authors use constant price US dollar measures of output that are not comparable across countries. Second,

[1] For example, the national accounts data used in Beck et al. [2000] and Levine et al. [2000] come from the World Saving Database of the World Bank, which is not maintained today or publicly available, making it difficult to assess the veracity of these data or to replicate these papers. The World Bank national accounts data that I use comes from the World Development Indicators.

the authors' choice of a quadratic specification does not help them to distinguish between the positive log-linear relationship found in Beck et al. [2000] and a non-monotonic relationship between growth and credit. In fact, estimating a concave quadratic relationship is exactly what one would expect if the relationship between growth and credit were monotonic, increasing and logarithmic. Third, the estimate that Arcand et al. [2012] obtain for the coefficient on the initial income variable using their estimator is below what one would obtain with their data if one were to use a within groups estimator instead. Since the within groups estimator is known to be biased downward for the coefficient on initial income in a growth regression, this suggests that the estimator used by Arcand et al. [2012] is inconsistent, which could be due to misspecification of the explanatory variables or instrumental variables.

I find a specification that includes the Solow–Swan growth determinants and does not seem to be invalidated by any diagnostic tests. I use a piecewise log-linear specification for the relationship between growth and credit that allows to test for departures from monotonicity and the log-linear relationship found in Beck et al. [2000]. I find stronger evidence for the piecewise log-linear relationship between growth and credit than for the quadratic relationship between growth and credit in the sense that the evidence for a quadratic relationship becomes weak once I control for the Solow–Swan growth determinants. Estimates of the piecewise log-linear specification show a negative relationship between growth and credit at levels of credit above 90% of output. I cannot find evidence for departures from log-linearity in the relationships between growth and investment or between investment and credit. Together these findings suggest, under the identification assumptions, that the negative effect of credit on growth at high levels of credit is not explained by a negative effect of investment on growth at high levels of investment or by a negative effect of credit on investment at high levels of credit.

Measuring nominal yield risk

The second chapter in this thesis contributes to the literature on measuring the effect of changes in nominal interest rates on financial intermediaries. I introduce a general framework for expressing a financial intermediary's net worth in terms of the term to maturity structures of nominal interest rates. The static version of this framework generalises Piazzesi and Schneider [2010] and Sher and Loiacono [2013] and predicts the change in net worth due to any counterfactual change in nominal interest rates. The dynamic version of this framework generalises Drehmann et al. [2010] and describes how the evolution of net worth over time depends on the paths of nominal interest rates.

In practice we might be concerned about parameter estimation uncertainty and measurement error in the accounting data on which these calculations are based. In the context of the static version of the framework, these two types of uncertainty induce sampling variation in the prediction of the change in net worth under a given scenario of a counterfactual change in nominal interest rates. In simple terms, the measure of risk is itself uncertain due to parameter estimation and accounting data measurement error. Therefore, in the context of the static version of this general framework, I derive the bias and variance of the predicted change in net worth under a given counterfactual change in nominal interest rates. The bias formula reveals that the predicted change in net worth is unbiased for the change in net worth that would be estimated if there were no measurement error in the accounting data in the special case where the observed accounting data are unbiased for the true accounting data. The variance formula can be used to compute standard errors for the measure of change in net worth; these standard errors could in turn be used to compare the changes in net worth across intermediaries and over time.

I demonstrate these methods in an application to $10,759$ US banks, quarterly between 2005 and 2010. Using observed yields to maturity on syndicated loans originated in the US between 2005 and 2010, I estimate Svensson [1994] specifications each quarter for the term to maturity structure of nominal interest rates underlying loan contracts. I consider a counterfactual scenario for nominal interest rates on loan contracts published by the Bank for International Settlements where the short term rate increases by 85% of its original level and the long term interest rate is held fixed. In the scenario, I hold fixed other asset prices and in particular nominal interest rates on contract types other than loan contracts, so that I only consider changes in the values of loan contracts. In this special case of the model, the static version gives the predicted change in net worth due to a change in the values of loan contracts only.

Applying this methodology to the balance sheets of individual banks in the second quarter of 2010, I find that the median bank loses about 2.5% of the value of its loan book under the counterfactual interest rate scenario, which corresponds to about 17.5% of its net worth. In other words, its ratio of equity to assets falls from about 10% to about $8.25\% = 10\% \times (1 - 17.5\%)$. The banks in the second quarter of 2010 with the largest losses lose about three times more than the median bank. These losses are large enough to question whether they may have implications for the intermediary's lending behaviour or borrowing costs. I compute standard errors for these predicted changes in net worth using the following calibrations for parameter estimation and accounting valuation uncertainty. White [1982] provides a sandwich covariance matrix estimator for the quasi-likelihood estimator that I use to estimate the parameters underlying the Svensson [1994] model of the term to maturity structure for loan contracts, and this covariance matrix estimator provides a calibration for the parameter estimation uncertainty underlying the predicted change in net worth. Bank of England [2011] provides survey measures of accounting valuation uncertainty that I use to calibrate the variance parameters characterising measurement error in the accounting data underlying the computation of the predicted change in net worth. I find that the resulting standard errors associated with the change in net worth are large relative to the measures of change in net worth, so it is important to calculate them in applications.

I also apply the above methodology to the average balance sheet across all banks in each quarter, to assess the riskiness of the overall banking system. A law of large numbers ensures that measurement error in the accounting data is negligible after averaging across all $10,759$ banks, so the uncertainty in the measures of loss for the aggregate banking system primarily reflects parameter estimation uncertainty. These predicted measures of loss display some cyclicality and peak just before the onset of the financial crisis, which suggests that they may contain some useful information.

Measuring dependence using the entropy rate

The third chapter of this thesis contributes to the statistics literature on measuring dependence in time series or cross sections. Independence in this context can be understood as the extent to which the joint distribution function can be factorised into a product of marginal distribution functions. I consider a population measure of dependence based on the entropy rate, which is known in the computer science literature but contrasts with the population measures of dependence based on correlation or entropy that appear in the econometrics literature. I review population measures of dependence based on correlation or entropy, including the large sample properties of their estimators, and I compare them to the population measure of dependence based on the entropy rate.

I show that the amount of compression achieved by common data compression algorithms can be used as a consistent estimator for the entropy rate of a stationary and ergodic data generating process and therefore also for the degree of serial dependence in the process. The intuition is that data compression algorithms achieve compression by finding repeated subsequences within the sample, and repeated subsequences are more likely to occur if the data generating process is more serially dependent. I consider the application of this method to random vectors with continuous state spaces, to cross sections and to Markov processes. The case of a continuous state space requires careful attention because some discretisation of the state space needs to be performed in order to use data compression algorithms, in which case the entropy rate being estimated is that of the discretised data generating process rather than that of the original data generating process.

An analytical result by Jacquet and Szpankowski [2011] provides the large sample distribution for the entropy rate estimator in the case of the classic Lempel–Ziv 1978 compression algorithm applied to an independent and identically distributed discrete state space stochastic process. However, the literature does not provide evidence on the distribution of the estimator when applied using other compression algorithms or to other stochastic processes. I provide numerical evidence for the distribution of the entropy rate estimator in the case of two specific modern compression algorithms, the Lempel–Ziv Markov Algorithm (LZMA) and the Burrows–Wheeler BZip2 algorithm (BZ2), when applied to independent and identically distributed discrete state space stochastic processes. I find that both of these algorithms achieve convergence in mean in polynomial-log time. The standard deviation of the LZMA entropy rate estimator decreases approximately with the square root of the sample size and the standard deviation of the BZ2 entropy rate estimator decreases faster than a square-root rate, but slower than a linear rate.

The above numerical (simulated) distributions can be used to obtain critical values in a hypothesis test for serial dependence for stationary and ergodic discrete state space stochastic processes. Levels of compression in the upper tail of the distribution under the null of no serial dependence should be regarded as unlikely draws under the null and lead to a rejection of the null hypothesis. Using critical values from the above numerical distributions as distributions under the null of no serial dependence, I conduct Monte Carlo experiments to compute the size and power of tests for serial dependence for a Markov chain and an AR(1) process. These experiments reveal that the hypothesis test controls size and achieves unit power with sufficiently many observations. However, the discrete nature of the distribution of the entropy rate estimator in finite samples can make it difficult to achieve a desired size for the test.[2] I provide some suggestions for further work that could be done to overcome this problem. The experiments also reveal that the test based on the entropy rate outperforms a test based on correlations for the Markov chain example, even in small samples.

[2] In other words, a desired size like 5% is not available in some cases, but smaller and larger sizes are available.

Chapter 1

Finance, investment and growth

I construct a dataset from modern sources and use it to revisit claims in the literature about the relationship between financial development and economic growth. I focus on the empirical literature that uses data aggregated into non-overlapping 5-year windows and financial development measured by the stock of credit extended by deposit money banks to the private sector. I show that it is important to allow for differences in purchasing power across countries when studying the relationship between growth and credit. I find that the Solow–Swan growth determinants are important omitted variables in the existing empirical literature. The role for credit in determining growth is attenuated once one controls for investment. The positive relationship between growth and credit in the literature is explained by a combination of the well known positive relationship between growth and investment and a positive relationship between investment and credit. However, there is evidence for a negative relationship between growth and credit at levels of credit above 90% of output and this relationship is not explained by any declining effect of credit on investment at these high levels of credit.

Keywords: Financial development, private credit, capital accumulation, economic growth
JEL Classification: O16, O47, E44

1.1 Introduction

Summarising more than thirty years of research on the long-run linkages between financial development and economic development, Levine [2005] concludes that countries with better functioning financial intermediaries grow faster and this finding is not likely to be driven by confounding factors or reverse causation. This stylised fact encourages policymakers to find ways of promoting financial development and academics to discover the determinants of financial development. After the global financial crisis, this stylised fact has been scrutinised and several papers, exemplified by Sahay et al. [2015], have concluded that financial development can become excessive for economic growth. This conclusion is convenient for policymakers in advanced economies at a time when tighter capital regulation on financial intermediaries threatens their provision of credit to the real economy.

Empirical work on the relationship between financial development and economic growth should be seen in the context of a larger literature on the empirics of economic growth.[1] Doing so raises questions. Firstly, does the economic growth that accompanies financial development also raise living standards? To compare living standards associated with the levels of output across countries, it is important to adjust current local currency measures of output for price differences across countries and over time. King and Levine [1993] do not explain the sense in which they measure output in 'real' units, while Beck et al. [2000] measure output in constant price US dollars. Similarly, a role for financial development in determining output along a balanced growth path can only be determined from a dynamic model in which output is measured in comparable units across countries and time. A second question raised by viewing the empirical finance–growth literature in a broader context is whether credit might by 'proxying' for other well known determinants of economic growth. For example, the share of investment expenditure in output is commonly thought to explain growth,[2] yet is absent from the specifications in King and Levine [1993], Beck et al. [2000], Loayza and Ranciere [2006] and Arcand et al. [2012].

In this paper I focus on the relationship between economic growth and credit (sometimes called financial 'depth'), which is the most commonly used measure of financial development, by applying instrumental variables procedures to country panel data in non-overlapping 5-year averages. I show that the positive association between credit and economic growth prevails when output and control variables are measured in constant price international dollars. When output is measured in constant price international dollars, the evidence for a quadratic relationship between growth and credit is weak, but there is strong evidence for a piecewise log-linear concave relationship between growth and credit, with a turning point at a 90% ratio of credit to output. However, these relationships could be proxying for the well known relationships between growth and the Solow–Swan model growth determinants, which are the share of investment expenditure in output and the population growth rate. I show that these are indeed omitted variables in the specifications used in the literature to study the effect of financial development on growth. Once investment and population growth are controlled for, the role for credit in explaining growth is attenuated. In particular, the evidence for a positive log-linear relationship between growth and credit disappears in the presence of investment, although evidence for a negative relationship between growth and credit at levels of credit above 90% of output remains even in the presence of investment. If the instrumental variables I use here are valid, which is not contradicted by the tests of overidentifying restrictions, this finding suggests there is a positive effect of credit on growth that operates through the positive effect of credit on investment, and that there is a negative relationship between credit and growth at high (above 90% of output) levels of credit

[1] For a review of the broader empirical growth literature, see, for example, Durlauf et al. [2005].

[2] See, for example, Mankiw et al. [1992], Levine and Renelt [1992], Durlauf et al. [2005].

that does not operate through an effect of credit on investment, population growth or any of the other control variables commonly used in the literature.

The positive component of the relationship between growth and credit is 'explained' by investment in the sense that it disappears in the presence of investment. However, to argue that credit affects growth through investment, one also needs to find a relationship between investment and credit. I show evidence for positive relationships between growth and investment and between investment and credit. I therefore find that the relationship between credit and economic growth found in earlier work is proxying for the well known relationship between investment and economic growth. If the instrumental variables used here are valid, the negative effect of credit on growth at levels of credit above 90% of output is not due to any negative effect of credit on investment at levels of credit above 90% of output.

Section 1.2 describes the data. Section 1.3 builds on Beck et al. [2000] to arrive at an admissible specification for an empirical model of economic growth with credit and other explanatory control variables. Section 1.4 uses this specification to re-examine stylised facts in the finance–growth literature. Section 1.5 concludes.

1.2 Data

1.2.1 Sources

I construct a large country–year dataset on the variables used in previous papers from the latest versions of authors' original sources. I obtain national accounts data on output, expenditure shares in output, terms of trade and life expectancy from the World Development Indicators (WDI),[3] I also obtain national accounts data on output and expenditure shares in output from various editions of the Penn World Table (PWT). I obtain data on output, credit and inflation from the International Financial Statistics (IFS).[4] I obtain data on the number of years of schooling attained in the population aged at least 25 from Barro and Lee [2013]. I obtain data on the black market currency premium from the Global Development Network Growth Database of the World Bank. I obtain data on the numbers of revolutions from the 2013 edition of the Cross-National Time Series Data Archive.

1.2.2 Credit

I construct country–year measures of the ratio of credit to output from the IFS by following footnote 12 in Beck et al. [2000], which is similar to the definition of variable `GFDD.DI.12` in

[3] I use the online edition of the World Development Indicators accessed from the World Bank's website in August 2015.

[4] I use the online edition of the International Financial Statistics database accessed from the International Monetary Fund's website in August 2015.

the Global Financial Development Database (GFDD) of the World Bank. The numerator of the ratio is credit extended by deposit-taking banks and other financial intermediaries to the private sector, which appear in the IFS as lines `22D` and `42D` respectively. The IFS defines line `22D` as claims by financial institutions whose liabilities are included in the country's definition of broad money on the private sector and line `42D` as claims by financial intermediaries, other than the central bank or those institutions whose liabilities are included in the country's definition of broad money, on the private sector.[5] The remainder of this subsection describes the construction of the ratio of credit to output in detail.

In the notation of the documentation to the GFDD, I construct the ratio of credit to output in year t using the formula

$$\frac{1}{4}\left(\frac{F_t}{P_t}+\frac{F_{t-1}}{P_{t-1}}\right)\div\frac{\mathrm{GDP}_t}{P_t+P_{t-1}} \tag{1.1}$$

where F denotes credit in current price local currency units, P denotes a consumer price index in current price local currency units, GDP denotes output in current price local currency units and the country index is suppressed.[6] Beck et al. [2000] claim that the reason for introducing consumer price indices is to match credit, the stock variable in the numerator measured at the end of the year, with output, the flow variable in the denominator measured during the year.

Lines `22D` and `42D` of the IFS are superceded by `FOSAOP` and `FFSAP` respectively; for Euro countries, line `22D` is superceded by `FOSAOEA`. If we take the notation $c(x, y, z)$ to mean the variable x if it is not missing, otherwise the variable y if it is not missing, otherwise the variable z if it is not missing, then my definition of credit can be written

$$F_t = c(\mathtt{FOSAOEA}_t, \mathtt{FOSAOP}_t, \mathtt{22D}_t) + c(\mathtt{FFSAP}_t, \mathtt{42D}_t, 0) \tag{1.2}$$

again with the country index suppressed, so that F_t is missing if and only if $\mathtt{22D}_t$ and its successors are missing.[7] I then set F_t to missing if it is equal to zero in any country–year, because it is not realistic to suppose that any country has exactly zero credit.

I set the consumer price index variable P_t to the IFS variable `PCPI`.[8] In the analyses below, I vary the source of the current price local currency output variable in the denominator of the ratio (1.1) of credit to output according to the source of the dependent variable in the growth specification. In the cases where an output measure from IFS is warranted, I use the current

[5] The IFS explicitly includes insurance companies and pension funds in the definition of the institutions whose claims on the private sector are included in line `42D`.

[6] The constant of proportionality 1/4 ensures that if $P_t = P_{t-1}$ and $F_t = F_{t-1}$ then the ratio (1.1) reduces to F_t/GDP_t.

[7] I follow Beck et al. [2000] in treating $\mathtt{42D}_t$ as zero when it is missing for the purposes of constructing F_t, which they presumably do because line `42D` is missing for many country–years. The results reported in this paper using a narrower definition of credit consisting only of line `22D` are very similar.

[8] The variable `PCPI` in the IFS supercedes the various lines `64*` from the IFS. For the UK, I set P_t to the retail price index IFS variable `CPRPTT01` because `PCPI` is not available.

price local currency gross domestic product variable `NGDP` from that source.[9] Within the IFS and between the IFS and PWT I take great care to ensure the numerator and denominator currencies and scaling units (e.g. thousands, millions) match. I provide summary statistics for my variables alongside the variables from Beck et al. [2000] and Arcand et al. [2012] in Section 1.2.4.

1.2.3 Aggregation and transformations

For comparability with previous research, I aggregate the underlying country–year data into non-overlapping 5-year intervals that are open on the left: $(1945, 1950], (1950, 1955], \ldots, (2005, 2010]$. I aggregate by picking out the final annual observation in each 5-year interval for the natural logarithm of real GDP per capita, the natural logarithm of the number of years of schooling and the number of years of life expectancy, and by averaging annual observations for all other variables. Aggregation loses information at the higher frequency, but provides some abstraction from phenomena at the business cycle frequency and from the inherent noise in price comparisons across countries at the annual frequency.[10]

Variable transformations, and the order in which I perform aggregation and transformation, are designed to be as close as possible to the existing literature. To match Beck et al. [2000], I average the annual ratios of credit, government consumption and exports plus imports to output in percentage point units and then apply the natural logarithm. By contrast, I average the annual black market premium in hundredths of percent (i.e. a one hundred percent black market premium is represented as the number 1) before applying the natural logarithm. For consistency, I treat the ratio of investment to output in the same way as the ratio of government consumption to output. I average the effective annual inflation rate in percentage point units before converting to a continuously compounded (i.e. logarithmic) inflation rate, to be consistent with the other variables that are expressed in logarithmic units.[11] I construct the Solow–Swan population growth variable $n + g + \delta$ as the natural logarithm of the sum of 5 and the average annual population growth rate in percentage point units.

1.2.4 Variable summaries

In this subsection I provide summary statistics for the variables used in this paper. Here I scale the variables in units that are natural to interpret, rather than the units explained in Section

[9] The IFS variable `NGDP` supercedes the various current price local currency gross domestic product variables in lines `99B`* from the IFS.

[10] Deaton and Heston [2010] caution against the application of Penn World Tables data to empirical growth research at the annual frequency.

[11] Based on replication exercises, Beck et al. [2000] apply the logarithm transformation to annual inflation before averaging to the 5-year frequency, and then apply a second logarithm transformation to the aggregated log inflation measure to obtain a 'log-log' inflation measure. There is no precedent in the literature for using log-log inflation.

1.2.3 above that I use in the regression estimations in Sections 1.3 and 1.4, so that the reader can quickly interrogate the ranges of the variables. I present the summary statistics in the samples in which the variables first appear in regression tables in Sections 1.3 and 1.4, so the reader can use the 'period' and 'count' columns to match variable summary statistics with regression estimation results in Sections 1.3 and 1.4. For comparison, I provide summary statistics for the variables from Beck et al. [2000] and Arcand et al. [2012] on the authors' original samples, which are also used in the replication exercises in Sections 1.3 and 1.4 of this paper.[12]

The initial income variables in Table 1.1 are the lagged final annual observation in each 5-year window of the natural logarithm of 'real' output per person, where 'real' in some cases means constant price US dollars and in other cases means constant price international dollars.[13,14] I explain these currency units in the two subsequent paragraphs. The growth variables in this table are one fifth of the change in the initial income variable over the 5-year window, multiplied by one hundred to convert to percentage point units.

The WDI measures a variable in constant 2005 price US dollars by converting the variable from current price local currency to constant 2005 price local currency (using components of the national consumer price index as specific as possible to the variable in question) and then by converting the variable from constant 2005 price local currency to constant 2005 price US dollars using 2005 market exchange rates. The PWT uses variables published in constant 2005 price US dollars in the United Nations' National Accounts Main Aggregates Database, which are in turn computed analogously to the World Bank's procedure just described.

The PWT measures a component of the national accounts in constant price international dollars in a benchmark year of the International Comparison Program by dividing the measure of the component of national accounts in current price local currency units by the aggregate, across a basket of goods representing that component of the national accounts, of relative prices in local currency units per US dollar. In years that are not benchmark years of the International Comparisons Program, the PWT extrapolates from the nearest adjacent benchmark years using changes over time in ratios of the consumer price index (or GDP deflator) in local currency to US dollars.

The ratio of credit to output requires sources for the numerator and denominator in the ratio. Where I have constructed the credit variable myself, the source for the numerator is always the IFS. In specifications where the growth rate and initial income variable are from a PWT source, the denominator of the ratio of credit to output is also taken from the PWT source, in the appropriate current price local currency units.

[12] I thank Thorsten Beck and Ugo Panizza for making their data and code available.

[13] PWT8.1 provides output in constant price international dollars on two different bases: output-based and expenditure-based, with variable names `rgdpo` and `rgdpe` in that dataset. The former adjusts for relative prices of exports and imports in output while the latter does not. The abbreviation 'PPPo' in the 'units' column of the table of summary statistics indicates `rgdpo`.

[14] I use population in the denominator of the ratio of output per person, rather than using numbers of working age people or numbers of workers, to match Beck et al. [2000].

Table 1.1: Key variables in the specifications of this paper and their summary statistics in the country–5-year samples on which I estimate models with specifications that include them. I provide summary statistics based on multiple sources so that the summary statistics of variables can readily be compared.

	source	units	period	count	mean	sd	min	max
growth	BLL (2000)	%	1960–1995	288	1.29	2.73	-10.02	8.68
growth	WDI USD	%	1960–1995	235	1.59	2.50	-5.55	8.93
growth	PWT8.1 USD	%	1960–1995	231	1.66	2.49	-6.70	8.73
growth	PWT8.1 PPPo	%	1960–1995	232	2.01	3.26	-10.89	12.65
growth	PWT8.1 PPPo	%	1960–2010	528	2.27	3.33	-19.21	17.31
growth	ABP (2012)	%	1960–2010	917	2.02	2.77	-21.00	13.86
initial income	BLL (2000)	log USD pp	1960–1995	288	7.81	1.32	5.24	9.91
initial income	WDI	log 2000 USD pp	1960–1995	235	8.57	1.52	5.48	10.88
initial income	PWT8.1	log 2000 USD pp	1960–1995	231	8.32	1.45	5.29	10.50
initial income	PWT8.1	log 2005 PPPo pp	1960–1995	232	8.88	1.03	6.02	10.43
initial income	PWT8.1	log 2005 PPPo pp	1960–2010	528	8.80	1.13	5.79	11.09
initial income	ABP (2012)	log 2000 USD pp	1960–2010	917	7.80	1.55	4.61	10.89
school	BLL (2000)	years	1960–1995	288	5.32	2.80	0.39	12.00
school	BL (2013)	years	1960–1995	235	6.45	2.75	1.09	12.69
school	BL (2013)	years	1960–2010	528	6.57	3.09	0.48	13.42
school	ABP (2012)	years	1960–2010	917	5.79	3.10	0.27	13.19
credit	BLL (2000)	%	1960–1995	288	45.89	36.89	1.56	205.95
credit	IFS	%	1960–1995	235	44.05	31.27	2.70	175.04
credit	IFS PWT8.1	%	1960–1995	232	44.09	33.08	1.78	174.21
credit	IFS PWT8.1	%	1960–2010	528	49.16	45.76	1.40	283.57
credit	ABP (2012)	%	1960–2010	917	39.97	37.04	0.74	269.76
government	BLL (2000)	%	1960–1995	288	15.03	5.46	5.66	38.02
government	WDI USD	%	1960–1995	235	16.93	5.96	4.62	32.86
government	ABP (2012)	%	1960–2010	917	15.27	5.84	3.22	45.96
trade	BLL (2000)	%	1960–1995	288	56.03	28.13	9.29	180.09
trade	WDI USD	%	1960–1995	235	50.27	33.23	10.01	187.54
trade	ABP (2012)	%	1960–2010	917	73.17	46.82	7.76	438.09
inflation	BLL (2000)	%	1960–1995	288	19.71	36.21	-3.06	344.40
inflation	IFS	%	1960–1995	235	20.01	45.95	0.20	389.40
inflation	ABP (2012)	%	1960–2010	917	18.38	152.70	-17.63	4523.70
black market premium	BLL (2000)	%	1960–1995	288	81.53	671.25	-3.68	10990.70
black market premium	GDNGD	%	1960–1995	235	85.47	738.09	-3.68	10990.70
investment	WDI USD	%	1960–1995	229	21.43	6.66	6.45	56.29
investment	WDI USD	%	1960–2010	528	21.67	6.55	4.83	56.29
population growth	PWT8.1	%	1960–1995	229	6.53	1.02	4.85	8.78
population growth	PWT8.1	%	1960–2010	528	6.62	1.06	0.12	12.48
growth in terms of trade	BLL (2000)	%	1960–1995	221	-0.56	3.53	-15.76	15.23
life expectancy	WDI	years	1960–1995	221	69.08	7.74	32.61	79.54
revolutions	CNTSDA	number	1960–1995	221	0.18	0.37	0.00	2.60

The summary statistics of my variables agree closely with those of Beck et al. [2000] on the subsets of their sample used in the regressions in Sections 1.3 and 1.4. The summary statistics of my variables are similar between the 1960–1995 and 1960–2010 samples. On the 1960–2010 sample, the summary statistics of Arcand et al. [2012] tend to be more extreme because the authors include 133 countries in their 1960–2010 sample.

1.3 A specification for finance–growth

1.3.1 Replicating, fixing and updating BLL (2000)

I start from the pair of equivalent specifications for all $(i,t) \in J \subseteq \{1,\ldots,N\} \times \{1,\ldots,T\}$

$$y_{it} - y_{i,t-1} = (\alpha - 1)y_{i,t-1} + \beta' x_{it} + \theta_t + c_i + u_{it} \tag{1.3}$$
$$y_{it} = \alpha y_{i,t-1} + \beta' x_{it} + \theta_t + c_i + u_{it} \tag{1.4}$$

due to Beck et al. [2000], where the index (i,t) denotes country i in 5-year period t, y is a random variable, x denotes a random column vector of measurements on other variables (including credit), α denotes a deterministic real number, β denotes a deterministic column vector of real numbers, θ_t denotes a latent deterministic real number that depends on t, c_i denotes a latent random variable depending on i and u_{it} denotes a latent random variable depending on (i,t). In this context, y is the logarithm of constant price US dollar output per person. The time-specific component θ_t captures all global, time-varying factors like common technology improvements across countries, commodity prices and world climate. The country-specific component c_i captures all time-invariant country-specific features that could explain variation in output like the level of technology in each country at the start of the sample period, geographic characteristics (including natural resources and access to oceans) and the legal or political institutions in each country at the start of the sample period. Beck et al. [2000] consider two specifications for the vector x: a full specification in their Table 2 Fit 4 that includes initial income, number of years of schooling attained, the ratio of credit to output, the share of government consumption in output, the ratio of the sum of exports and imports to output (trade openness), the annual rate of consumer price index inflation and the black market premium, and a restricted specification in their Table 2 Fit 3 that includes only the first three of these variables.

Assume $\mathbb{E}c_i = \mathbb{E}u_{it} = \mathbb{E}c_i u_{it} = 0$ and $T < \infty$. Then the ordinary least squares and within groups estimators applied to (1.3) are not consistent estimators of the parameter vector $(\alpha\,\beta')'$ as $N \to \infty$ respectively because the random variable $y_{i,t-1}$ is correlated with the random variables c_i and $u_{i,t-1}$. Even in the case where we impose $\alpha = 0$, the within groups estimator of β requires the exogeneity assumption $\mathbb{E}u_{it}|x_{i1}\ldots x_{iT} = 0$ for all i and t, which is untenable in this context because we would expect past shocks u_{is} to output to be correlated with future realisations of the variables x_{it} (where $s < t$). For example, we would expect events in the sample space that lead to large realisations of output through idiosyncratic shocks u_{is} to lead to higher saving and, to the extent that these savings are channelled through financial intermediaries, to a higher stock of credit at future times $t > s$.

Beck et al. [2000] use an instrumental variables estimator for $(\alpha\ \beta')$ due to Arellano and

Bover [1995] and Blundell and Bond [1998].[15] For notational convenience in explaining the estimator, let us assume that all variables are expressed as deviations from time-specific means so that $\theta_t = 0$ in (1.3). For their full specification, Beck et al. [2000] partition the explanatory variables on the right hand side of (1.3) into $(y_{i,t-1}\, x'_{it}) = (\pi'_{it}\, \eta'_{it})$ where π contains initial income and schooling and η contains credit, government, trade openness, inflation and the black market premium. Beck et al. [2000] then assume the moment conditions[16]

$$\begin{aligned} \mathbb{E}(\pi'_{i,t-1}\, \eta'_{i,t-2})\Delta u_{it} &= 0 \quad \text{for all } (i,t) \\ \mathbb{E}(\Delta\pi'_{it}\, \Delta\eta'_{i,t-1})(c_i + u_{it}) &= 0 \quad \text{for all } (i,t), \end{aligned} \tag{1.5}$$

which can also be written as $\mathbb{E}Z_i'e_i = 0$ for each i if we define

$$Z_i = \operatorname{diag}(Z_{i\Delta}, Z_{i\lambda}) \tag{1.6}$$

$$Z_{i\Delta} = \begin{bmatrix} \pi'_{i1} & \eta'_{i0} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \pi'_{i2} & \eta'_{i1} & 0 & 0 & & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi'_{i3} & \eta'_{i2} & & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \pi'_{i,T-1} & \eta'_{i,T-2} \end{bmatrix} \tag{1.7}$$

$$Z_{i\lambda} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \Delta\pi'_{i2} & \Delta\eta'_{i1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \Delta\pi'_{i3} & \Delta\eta'_{i2} & & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \Delta\pi'_{iT} & \Delta\eta'_{i,T-1} \end{bmatrix} \tag{1.8}$$

$$e_i = h \begin{pmatrix} c_i + u_{i1} \\ c_i + u_{i2} \\ \vdots \\ c_i + u_{iT} \end{pmatrix} \tag{1.9}$$

$$h = \begin{pmatrix} M_\Delta \\ I_{T\times T} \end{pmatrix} \tag{1.10}$$

where M_Δ is a $(T-1)\times T$ matrix with -1 on the diagonal, 1 on the superdiagonal and zeros elsewhere.

Bond et al. [2001] discuss the validity of the conditions (1.5) in the context of the data

[15] The estimator is based on earlier work in Holtz-Eakin et al. [1988] and Arellano and Bond [1991].

[16] The letters π and η allude to the terminology ‘predetermined’ and ‘endogenous’.

generating process (1.3) where the variables in x_{it} are the share of investment expenditure in output and the population growth rate, which determine the steady state balanced growth path for output in the Solow–Swan growth model.[17] Bond et al. [2001] argue that the correlation between the first lag of the change in real per capita output growth $\Delta y_{i,t-1}$ and the time-invariant technology, geography and resources determinants of growth in c_i can be made arbitrarily small by extending back in time the initial time period over which the model (1.3) is defined.[18] Bond et al. [2001] also argue that uncorrelatedness between investment expenditure share growth rates in Δx_{it} and the time-invariant technology, geography and resource determinants of growth in c_i is not unreasonable and that such correlations would lead to implausible long-run behaviour in the model (1.3). Such claims could also be made in the context of x_{it} in (1.3) being defined according to Beck et al. [2000].

By stacking (1.3) across t, we can see that e_i depends (linearly) on $(\alpha\ \beta')$, the parameter vector to be estimated. We could therefore imagine choosing parameters to solve the orthogonality conditions $\mathbb{E}Z_i'e_i = 0$ for any i. In this application, given the sizes of N, T and x, the number of columns of Z_i exceeds the number of parameters to be estimated so the orthogonality conditions $\mathbb{E}Z_i'e_i = 0$ for any i do not admit a unique solution for the parameters. However, if we define $Z = (Z_1' \cdots Z_N')'$, $e = (e_1' \cdots e_N')'$ and the generalised method of moments (GMM) estimator

$$\hat{\gamma} = \operatorname{argmin}\{e'Za_N'a_NZ'e :\ (\alpha\ \beta') \in S\} \tag{1.11}$$

for a sequence $a_1, a_2, \ldots$ of random matrices and $S \subseteq \mathbb{R}^k$ then under further conditions given in Hansen [1982],

$$\sqrt{N}\left(\hat{\gamma} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right) \tag{1.12}$$

converges in distribution to a normal distribution. This estimator finds the parameters α and β in the sample that minimise the size of the vector $\frac{1}{N}Z'e = \frac{1}{N}\sum_{i=1}^{N} Z_i'e_i$ that is the sample analog of the population vector $\mathbb{E}Z_i'e_i$ for any i, where size is measured by the norm $||\xi|| = \sqrt{\xi' a_N' a_N \xi}$. Since e is linear in the parameters, $||\frac{1}{N}Z'e||^2$ can be minimised with respect to the parameters using a first order condition that is linear in the parameters, from which $\hat{\gamma}$ is the (linear) weighted least squares estimator

$$\hat{\gamma} = \left((y_{(-1)}\ x)'Za_N'a_NZ'(y_{(-1)}\ x)\right)^{-1} (y_{(-1)}\ x)'Za_N'a_NZ'y \tag{1.13}$$

[17] Mankiw et al. [1992], Islam [1995] and Durlauf et al. [2005] derive the equation (1.3) with these x_{it} variables from the classical Solow–Swan growth model.

[18] We conduct this thought experiment without changing the period $1, \ldots, T$ over which y_{it} is observed.

where $y = ((hy_1)' \cdots (hy_N)')'$, $y_i = (y_{i1} \cdots y_{iT})'$,

$$(y_{(-1)}\, x) = \begin{pmatrix} h(y_{1,(-1)} & x_1) \\ \vdots & \vdots \\ h(y_{N,(-1)} & x_N) \end{pmatrix},$$

$y_{i(-1)} = (y_{i0} \cdots y_{i,T-1})'$ and $x_i = (x_{i1} \cdots x_{iT})'$.[19] The forms of h and Z_i above give so-called *system GMM* while replacing (1.6) by $Z_i = Z_{i\Delta}$ and (1.10) by $h = M_\Delta$ gives so-called *difference GMM*.

We obtain a one-step estimator using $a_N' a_N = (\frac{1}{N} Z'HZ)^{-1}$ where we could set $H = I$ or $H = I_N \otimes hh'$. For system GMM, the Gauss software `DPD96` implements $H = I$ while the Stata program `xtabond2` implements both forms of H. We obtain an efficient two-step estimator using $a_N' a_N = \frac{1}{N} \mathbb{V} Z'e$ and a consistent estimator of $\mathbb{V} Z'e$ is $\sum_{i=1}^N Z_i' \ddot{u}_i \ddot{u}_i' Z_i$, where $\ddot{u}_i = hy_i - h(y_{i,(-1)}\, x_i)\hat{\delta}_1$ and $\hat{\delta}_1$ is a one-step estimator.

We should ensure that the exogeneity conditions for y_{it} implicit in (1.5) are consistent with the form of y_{it} in (1.3). In the absence of the explanatory variables x_{it} (i.e. imposing $\beta = 0$), it would be sufficient to assume serially uncorrelated idiosyncratic errors $\mathbb{E} u_{is} u_{it} = 0$ for any $s \neq t$ and predetermined initial conditions $\mathbb{E} y_{i0} u_{it} = 0$ for all i and $t > 0$ to get $\mathbb{E} y_{i,t-2} \Delta u_{it} = 0$ for all (i, t) in (1.5).[20] In addition, it would be sufficient to assume $\mathbb{E} \Delta y_{i1} c_i = 0$ on the initial conditions to imply $\mathbb{E} \Delta y_{i,t-1}(c_i + u_{it}) = 0$ for all (i, t) in (1.5).[21] The assumption $\mathbb{E} u_{is} u_{it} = 0$ for $s \neq t$ above implies negative first-lag serial correlation $\mathbb{E} \Delta u_{i,t-1} \Delta u_{it} < 0$ and zero second-lag serial correlation $\mathbb{E} \Delta u_{i,t-2} \Delta u_{it} = 0$ in the differenced idiosyncratic errors, for all (i, t). The latter two conditions can be tested using the asymptotic normality of

$$\sum_{i=1}^{N} \sum_{t=2+\delta}^{T} \Delta \hat{u}_{i,t-1-\delta} \Delta \hat{u}_{it} \tag{1.14}$$

under the null hypothesis that $\mathbb{E} u_{is} u_{it} = 0$ for $s \neq t$, where $\Delta \hat{u}_{it} = \Delta y_{it} - \Delta(y_{i,t-1}\, x_{it}')\hat{\gamma}$ are the differenced residuals and $\delta \in \{0, 1\}$ [Arellano and Bond, 1991].[22] With k parameters in (1.3) to be estimated, only k equations in (1.5) are necessary to identify these parameters. If p is the number of columns of Z, then the validity of the $p - k$ extra equations in (1.5) can be tested

[19] $\hat{\gamma}$ is a weighted least squares estimator in the sense that it coincides with ordinary least squares projection of $a_N Z' y$ onto the columns of $a_N Z'(y_{(-1)}\, x)$.

[20] This is because $y_{i,t-2} = \alpha^{t-2} y_{i0} + c_i(1 + \alpha + \alpha^2 + \cdots + \alpha^{t-3}) + \sum_{k=0}^{t-3} \alpha^k u_{i,t-2-k}$.

[21] This is because $\Delta y_{i,t-1} = \alpha^{t-2} \Delta y_{i1} + \sum_{k=0}^{t-2} \alpha^k \Delta u_{i,t-1-k}$ and $\mathbb{E} \Delta y_{i,t-1}(c_i + u_{it}) = \alpha^{t-2} \mathbb{E} \Delta y_{i1} c_i$.

[22] The parameter δ determines the lag $1 + \delta$ at which serial correlation in the differenced idiosyncratic errors Δu_{it} is tested. The T appearing in (1.14) could also depend on i to allow for unbalanced panels, although this is more notation than content.

using the Sargan–Hansen statistic

$$\hat{u}'Z\left(\sum_{i=1}^{N} Z_i'\ddot{u}_i\ddot{u}_i'Z_i\right)^{-1} Z'\hat{u} \tag{1.15}$$

where $\hat{u} = y - (y_{(-1)}\; x)\hat{\gamma}$, and the Sargan–Hansen statistic has an asymptotic (in N) chi-squared distribution with $p - k$ degrees of freedom.

Columns (1) and (5) of Table 1.2 are estimation results produced by `DPD96` for Gauss that exactly replicate Columns (3) and (4) in Table 2 of Beck et al. [2000] using two-step standard errors, with $H = I$ in the first step, that are known to be subject to finite sample bias.[23,24] Columns (3) and (7) in Table 1.2 are the closest that one can get to these estimation results using `xtabond2` for Stata, because there are differences between the implementations of system GMM in `DPD96` and `xtabond2`.[25] To confirm that the only source of difference between Columns (1) and (3) is the software implementation, Columns (2) and (4) show that the analogous difference GMM estimation results agree between `DPD96` and `xtabond2`. Similarly, the fact that Columns (6) and (8) coincide shows that the only source of difference between Columns (5) and (7) is due to differences between the `DPD96` and `xtabond2` software implementations of system GMM. Beck et al. [2000] prefer the system GMM estimator to the difference GMM estimator because lagged levels of the explanatory variables can be weak instruments for the equations in first-differences when the explanatory variables are persistent [Blundell and Bond, 1998]. Explanatory variables that are stocks rather than flows, like years of schooling attained and the ratio of the stock of credit to output, can be expected to be relatively persistent through time and therefore their levels could be weak instruments for the equations in first-differences.

The estimation results in Columns (1), (3) and (5)–(8) of Table 1.2 find a positive relationship between growth and credit, although the statistical significance of the coeffcients in Table 1.2 cannot be taken literally because the two-step standard errors, as reported in that table, are known to be subject to downward finite sample bias.[26,27] The positive coefficients on credit appearing in Columns (1) and (5) are the basis for the finding in Beck et al. [2000] that there is a positive effect of credit on growth.[28] The estimation results in Columns (2) and (4) find a

[23]The finite sample bias in the two-step standard errors is explained in Windmeijer [2005].

[24]These exact replications rely on the code and data used by Beck et al. [2000]. I thank Thorsten Beck for providing these data and code.

[25]The estimation results in Columns (3) and (7) use $H = I$ to match the estimation procedure that produces Columns (1) and (5).

[26]See Footnote 23.

[27]I use a conventional 5% significance level in this paper.

[28]Beck et al. [2000] arrive at their conclusion that financial development exerts a positive effect on growth based on similar positive relationships estimated between growth and two other indicators of financial development. The two other indicators of financial development that they use are the ratio of the value of liquid liabilities of the financial system to output and the ratio of the value of commercial bank domestic assets to the sum of the values of commercial bank and central bank domestic assets. The value of liquid liabilities is the same as ‘M3’ and consists of currency held outside the banking system plus demand and interest-bearing liabilities of financial

Table 1.2: Estimation results based on `DPD96` for Gauss that exactly replicate the estimation results in Table 2 Fits 3–4 of Beck et al. [2000] and the closest possible estimation results based on `xtabond2` for Stata. The instrument set is defined formally in (1.5)*, but for completeness I re-state these instruments in words here. The difference GMM estimation results for Columns (2) and (4) use the first lag of the levels of initial income and school, and the second lag of the levels of credit, government, trade, inflation and the black market premium, as instruments for the differenced equations. In addition to the instruments used for the estimation results in Columns (2) and (4), the estimation results in Columns (1) and (3) also use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government, trade, inflation and the black market premium, as instruments for the levels equations. The difference GMM estimation results for Columns (6) and (8) use lags 1–3 of initial income and school and lags 2–4 of credit as instruments for the differenced equations. In addition to the instruments used for the estimation results in Columns (6) and (8), the estimation results for Columns (5) and (7) also use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, as instruments for the levels equations.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	dpd96		xtabond2		dpd96		xtabond2	
	sys	dif	sys	dif	sys	dif	sys	dif
initial income	-0.50***	-7.43***	-0.19	-7.43***	-1.30***	-5.60***	-1.53***	-5.60***
	(0.10)	(1.80)	(0.12)	(1.80)	(0.28)	(1.09)	(0.31)	(1.09)
school	0.95***	-11.61***	0.27	-11.61***	2.67***	-7.42***	3.02***	-7.42***
	(0.20)	(2.30)	(0.24)	(2.30)	(0.71)	(1.19)	(0.75)	(1.19)
credit	1.44***	-0.95*	1.10***	-0.95*	2.40***	1.47***	2.41***	1.47***
	(0.09)	(0.57)	(0.12)	(0.57)	(0.17)	(0.55)	(0.18)	(0.55)
government	-1.45***	-3.25***	-1.79***	-3.25***				
	(0.29)	(1.08)	(0.27)	(1.08)				
trade	1.31***	-0.03	1.21***	-0.03				
	(0.18)	(1.31)	(0.18)	(1.31)				
inflation	0.18	1.10	0.29	1.10				
	(0.25)	(2.12)	(0.30)	(2.12)				
black market prem.	-1.19***	-3.40***	-1.30***	-3.40***				
	(0.07)	(0.81)	(0.12)	(0.81)				
observations	365	288	288	288	365	288	288	288
countries	77	77	77	77	77	77	77	77
instruments			76	40			61	45
Hansen p-val.	0.506	0.364	0.478	0.364	0.183	0.078	0.201	0.0781
AR(1) stat.	-3.318	-2.240	-3.099	-2.062	-3.192	-3.028	-3.052	-2.920
AR(2) stat.	-0.249	0.035	-0.164	0.032	-0.649	-0.050	-0.660	-0.051

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

negative coefficient on credit. Due to downward finite sample bias on the standard errors noted above, the marginal statistical significance of this coefficient cannot be taken literally. The negative coefficient on credit in Column (2) and (4) could reflect misspecification of explanatory variables or weakly informative instrumental variables. Indeed, the coefficient -7.43 on the initial income variable in Column (2) is below the -6.46 obtained by applying the within groups estimator to estimate the coefficients in Column (2). The within groups estimator itself is known to be subject to downward bias for the coefficient α in the model (1.4) with T fixed,

intermediaries [King and Levine, 1993].

while in this model the within groups estimator is biased for β in unknown directions.[29]

Columns (1) and (5) of Table 1.2 show statistically significant negative coefficients on the lagged dependent variable.[30] In order to facilitate a Solow–Swan model interpretation of the coefficient on the lagged dependent variable in Table 1.2, re-write the data generating process (1.3) at the annual frequency as

$$100\frac{y_{i,\tau+5} - y_{i\tau}}{5} = 20(\alpha - 1)y_{i\tau} + 20\beta' x_{i,\tau+5} + 20(c_i + u_{i,\tau+5}) \tag{1.16}$$

where τ is time measured in years and the dependent variable is the percentage average annual growth rate used by Beck et al. [2000]. Durlauf et al. [2005] show that (1.16) arises as a first order approximation to a prediction equation from the Solow–Swan growth model. In particular,

$$\alpha - 1 = -(1 - e^{-5\lambda}) \tag{1.17}$$

where λ is a parameter describing the rate of conditional convergence of y to its steady state balanced growth path in the Solow–Swan growth model. The parameter λ is a differentiable transformation $\lambda(\cdot)$ of $\alpha > 0$, so the delta method can be applied to find an asymptotic normal distribution for $\lambda(\hat{\alpha})$ where the difference or system GMM estimator $\hat{\alpha}$ for α has an asympotic normal distribution. The estimated coefficient -0.5 and standard error 0.1 on the lagged dependent variable in Column (1) of Table 1.2 imply an estimate 0.51% for λ with standard error 0.102%.[31] The estimated coefficient -1.3 and standard error 0.28 on the lagged dependent variable in Column (5) of Table 1.2 imply an estimate 1.34% for λ with standard error 0.3%. Both of these estimates of λ, which are implied by the estimation results in Table 2 Fits 3-4 in Beck et al. [2000], give slower rates of (conditional) convergence than the estimates of between 2.1% and 2.4% shown in Table 1 of Bond et al. [2001].

Column (1) of Table 1.3 repeats Column (3) of the preceding table. Column (2) applies the Windmeijer [2005] correction for finite sample bias to the standard errors. The bias-correction renders the government expenditure share only marginally significant, the trade share insignificant; the correction also increases the standard error on the lagged dependent variable and schooling enough to make these variables insignificant even at the point estimates suggested by the `DPD96` estimation in Column (1) of the previous table (i.e. in Beck et al. [2000] Table 2 Fit 4). Column (3) of Table 1.3 exponentiates the authors' log-log inflation variable to obtain a more standard log inflation variable from their data. Column (4) applies the first step moment

[29] Nickell [1981] explains the downward bias of the within groups estimator for α in the model (1.4) with T fixed.

[30] Note again that the statistical significance should be viewed with caution because the two-step standard errors are subject to downward finite sample bias, as explained in Footnote 23.

[31] $\hat{\lambda} = -\log(\hat{\alpha})/5 = -\log(-0.5/20 + 1)/5 = 0.51\%$ and the standard deviation of $\hat{\lambda}$ is asymptotically-approximately equal to $|1/(5\hat{\alpha})|(0.1/20) = 0.102\%$.

weight matrix $H = I_N \otimes hh'$.[32] The use of this weight matrix recovers the strong significance of the coefficient on credit from the marginal significance in Column (3) and attenuates the coefficient on the black market premium enough to lose its statistical significance. Column (5) restricts the sample of (4) to those country–5-years for which I have been able to reconstruct the variables appearing in that column based on current versions of the authors' original data sources.

Column (6) of Table 1.3 replaces all variables with modern versions reconstructed from the authors' original data sources. In particular, I replace the authors' measures of output on the left and right hand side and expenditure shares by their equivalents in constant price 2000 US dollars from the WDI. In addition, I replace the authors' schooling variable by its equivalent from Barro and Lee [2013]. The ratio of credit to GDP is replaced by its equivalent constructed from the IFS, the black market premium is replaced by a variable that matches the original exactly and the inflation variable is replaced by the log-transformed average effective annual inflation rate from the IFS. Credit remains statistically significant when the authors' data are updated to modern equivalents. An analogous table that updates the authors' Table 2 Fit 3 is in Appendix 1.A.

1.3.2 Adjusting for purchasing power

1.3.2.1 The adjustment method

In the analysis below I follow Barro [1991], Mankiw et al. [1992] and Caselli et al. [1996] in using constant price international dollars for the left-hand side income growth rates and right-hand side initial income levels.[33] Ideally I would then measure expenditure shares in income, like government consumption, trade openness and investment, as shares in constant price international dollar GDP. However, such shares are not well defined for the chain-linked Fisher GDP provided in PWT versions 8.0 and 8.1. The expenditure shares in constant price international dollar GDP in versions 7.1 and earlier of the PWT are shares in fixed-weight

[32] This form of the moment weight matrix can be obtained as the variance of the difference-and-levels-transformed error components under the conditions that the idiosyncratic error term is uncorrelated across countries and time and there are no country-specific components c_i in the error terms.

[33] Nuxoll [1994] suggests that the left-hand side income growth rates be computed in constant price local currency and the right-hand side initial income level be computed in current or constant price international (PPP-adjusted) dollars. If these recommendations are followed, it is not clear what is being measured. The original motivation of Mankiw et al. [1992] was to study the validity of the Solow–Swan model as a 'dynamic' model describing the evolution of real income per worker, derived from the transition to a steady state growth path. Mankiw et al. [1992] show how the Solow–Swan model implies (1.4) to a first order approximation, which is equivalent to the form (1.3). The continuing motivation of empirical growth work is to discover the determinants of the level and growth rate of real income per worker (or per person). Using one set of units for the dependent variable and a different set of units for the lagged dependent variable invalidates interpretations of growth regressions as dynamic models for real income per worker. Consequently, for such a model the concepts of steady state and transition path are not well defined and it would be difficult to learn about the determinants of the levels of real income per worker (or per person).

Table 1.3: Estimation results reflecting the effects of correcting for finite-sample bias in the estimation results of Beck et al. [2000] Table 2 Fit 4 and of updating the variables used by those authors to versions based on current versions of the authors' original data sources. The estimations in this table use the first lag of the levels of initial income and schooling, and the second lag of the levels of credit, government, trade, inflation and the black market premium, as instruments for the differenced equations. The estimations in this table also use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government, trade, inflation and the black market premium, as instruments for the levels equations.

	(1)	(2)	(3)	(4)	(5)	(6)
	t2f4	w	w pi	w pi h3	w h3 pi-	usd
initial income	-0.19	-0.19	-0.13	-0.43	-0.42	-0.58
	(0.12)	(0.44)	(0.49)	(0.40)	(0.45)	(0.56)
school	0.27	0.27	0.03	1.02	0.27	0.33
	(0.24)	(1.23)	(1.36)	(1.38)	(1.03)	(0.30)
credit	1.10***	1.10**	1.12*	1.34***	1.18**	1.22**
	(0.12)	(0.52)	(0.63)	(0.45)	(0.59)	(0.51)
government	-1.79***	-1.79*	-1.76	-1.62*	-1.07	-0.43
	(0.27)	(1.05)	(1.13)	(0.95)	(1.42)	(0.97)
trade	1.21***	1.21	1.40	0.92	1.80**	0.86
	(0.18)	(1.27)	(1.37)	(1.08)	(0.90)	(0.61)
inflation	0.00	0.00	0.00	-0.00	-0.00	-0.00
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
black market prem.	-1.30***	-1.30***	-1.41***	-0.83	-0.82	-0.92*
	(0.12)	(0.39)	(0.48)	(0.54)	(0.79)	(0.48)
observations	288	288	288	288	235	235
countries	77	77	77	77	65	65
instruments	76	76	76	76	76	76
Hansen p-val.	0.478	0.478	0.458	0.358	0.717	0.722
AR(1) stat.	-3.099	-3.060	-3.038	-3.034	-3.259	-3.009
AR(2) stat.	-0.164	-0.161	-0.109	-0.291	-1.156	-1.213

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Laspeyres GDP. In order to use the latest version of the PWT, one is therefore left with the difficult question of how to control for expenditure shares when the ideal measures are not available.

One option is to use expenditure shares in constant price USD (or constant national prices), which have the advantage of being comparable within each country over time. Across countries, to the extent that differences in relative prices of government services to GDP between countries are time invariant, we can expect such differences to be absorbed in the country-specific fixed effects (and similarly for trade openness or investment expenditure share categories). A second approach would be to use the expenditure shares in current price international dollar GDP that PWT 8.0 and 8.1 provide. The time-specific dummy variables absorb time-incomparability in these shares that is common across countries. A final approach would be to try to adjust expenditure shares in current price international dollar GDP for price variation in the numerator and/or denominator of the expenditure share in the numeraire currency, which implicitly treats the chain-linked expenditure shares as fixed-weight expenditure shares.

I try all three of the above options for measuring expenditure shares. Fortunately the broad conclusions of this analysis do not depend on these choices. I refer the reader at various points to the appendices for the results of robustness checks. In the main results presented below, I use expenditure shares in constant price USD income.

1.3.2.2 The effect of adjustment

Column (1) in Table 1.4 presents the constant price USD specification from the previous table. Column (2) restricts the sample of the estimation underlying the results in Column (1) to those country–5-year observations for which the PWT version 8.1 provides constant price 2000 US dollar income per capita, which estimation results are therefore comparable to those presented in Column (3). Column (3) in Table 1.4 replaces the left-hand side growth rate and the right-hand side initial output by constant price USD output from the national accounts data underlying PWT8.1. Column (4) replaces left-hand side growth and the right-hand side initial output by constant price international dollar output `rgdpo` from the PWT8.1 and replaces output in the denominator of the ratio of credit to output by the measure of output in current price local currency from the national accounts data underlying the PWT8.1. Credit remains statistically significant in the specification of Beck et al. [2000] that I modify to adjust for differences in purchasing power across countries. Appendix 1.A shows that this conclusion is robust across other versions of the PWT and in the smaller (Table 2 Fit 3) specification of Beck et al. [2000].

1.3.3 Omitted variables and irrelevant variables

It is surprising that variables like the expenditure share of investment in output and population growth do not appear in the growth specifications in King and Levine [1993] and Beck et al. [2000], given the empirical relevance that these variables possess in Mankiw et al. [1992] and their strong links to the original Solow–Swan growth model. Barro [1991], Barro and Lee [1994] and Caselli et al. [1996] find other important determinants of steady state growth that are absent from the specifications presented in Beck et al. [2000] and Arcand et al. [2012], for example. In this section, I check whether there is evidence for the inclusion of these variables in the growth specification adjusted for purchasing power parity differences across countries. I then eliminate statistically insignificant variables to restrict attention to a parsimonious specification that can be used in the following section to explore claims in the more recent empirical finance–growth literature.

Column (1) of Table 1.5 presents the growth specification adjusted for purchasing power

Table 1.4: Estimation results showing the effect of correcting the left hand side real per capita output growth rate and right hand side initial real output per capita in the specification in Table 1.3 for differences in purchasing power of currencies across countries at every point in time. The estimations in this table use the first lag of the levels of initial income and schooling, and the second lag of the levels of credit, government, trade, inflation and the black market premium, as instruments for the differenced equations. The estimations in this table also use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government, trade, inflation and the black market premium, as instruments for the levels equations. The source and currency of initial income and credit in the instrument set varies across columns to match the source and currency of initial income and credit in the explanatory variables.

	(1)	(2)	(3)	(4)
	usd	usd-	81qkd	81okd
initial income (USD WDI)	-0.58	-0.75		
	(0.56)	(0.59)		
school	0.33	0.44	0.42	0.42
	(0.30)	(0.32)	(0.31)	(0.44)
credit	1.22**	1.59***	1.10*	
	(0.51)	(0.60)	(0.64)	
government	-0.43	-0.91	-0.92	-0.72
	(0.97)	(1.15)	(1.14)	(1.65)
trade	0.86	0.65	0.41	0.00
	(0.61)	(0.63)	(0.89)	(0.66)
inflation	-0.00	-0.01	-0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
black market prem.	-0.92*	-0.39	-0.74	0.06
	(0.48)	(0.61)	(0.51)	(1.14)
initial income (USD PWT)			-0.64	
			(0.54)	
initial income (PPP PWT)				-0.96
				(0.86)
credit (PWT)				1.43**
				(0.64)
observations	235	231	231	232
countries	65	63	63	62
instruments	76	76	76	76
Hansen p-val.	0.722	0.745	0.624	0.720
AR(1) stat.	-3.009	-2.842	-2.862	-3.499
AR(2) stat.	-1.213	-1.210	-0.161	1.127

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

differences across countries, as seen in the previous table. Column (2) restricts the sample according to joint non-missingness of additional explanatory variables to be included in this table. With this specification I test for the inclusion of five other variables in the specification of Caselli et al. [1996]: the investment expenditure share in constant price USD, the Solow-type population growth rate, growth in the terms of trade,[34] life expectancy and numbers of revolutions. Of these variables, only the investment expenditure share in output, the population growth rate and the growth rate in the terms of trade have significant evidence for individual inclusion in the model. The exclusion of these three variables from the specifications in Beck et al. [2000] is therefore a potential source of parameter inconsistency.[35] Appendix 1.C.1 shows that this significance of investment does not depend on the particular variable used to measure investment. I present the estimated models with the individual inclusion of the first two of these variables in Columns (3) and (4). The extended model with the joint inclusion of the investment expenditure share in output and population growth is in Column (5). When either the investment expenditure share, population growth or both are included in the specification, the coefficient on the lagged dependent variable becomes statistically significant. That is, I find evidence for convergence conditional on the Solow–Swan growth determinants. In addition, we have some evidence that credit is not statistically significant in the presence of the investment expenditure share in output.[36,37] In Appendix 1.C.1, I show that credit is not statistically significant in the presence of investment regardless of how investment is measured. Appendix 1.C.2 shows that any significance of credit in the presence of investment is also weak across

[34]In this exercise, I use the terms of trade growth variable as provided in the dataset accompanying Beck et al. [2000] and Levine et al. [2000]. Levine et al. [2000] employ almost identical specifications to Beck et al. [2000]—I explain the differences in what follows—and include this terms of trade growth variable as an extra control variable in investigations of the sensitivity of the estimated coefficients on their variables proxying for financial development. I use this terms of trade growth variable provided by Beck et al. [2000] because I did not find it possible to construct a terms of trade variable from available sources with sufficient coverage. The differences between the estimation methods in Beck et al. [2000] and Levine et al. [2000] are as follows. Levine et al. [2000] use a variable measuring the number of years of secondary schooling, while Beck et al. [2000] use a variable measuring the number of years of primary and secondary schooling. Levine et al. [2000] use the second lag of the level of (secondary) schooling as an instrument for the equations in differences, while Beck et al. [2000] use the first lag of the level of (primary and secondary) schooling as an instrument for the equations in differences. Levine et al. [2000] use a slightly smaller sample than Beck et al. [2000].

[35]The investment expenditure share in output, the population growth rate and the growth rate in terms of trade are therefore "omitted variables" in the extended specification of Beck et al. [2000].

[36]Note by comparing Columns (3) and (4) of Table 1.5 that it is really the presence of investment, not population growth, which eliminates the statistical significance of credit.

[37]In Columns (3) and (4), the p-value of the Sargan–Hansen statistic is close to 1. It is well known that a test using the Sargan–Hansen statistic and asymptotic critical values has low power to reject the null hypothesis of valid overidentifying restrictions when the number of instruments is large relative to the number of observations. We may be concerned with large numbers of instruments that the system GMM estimator (1.13) is biased in the direction of

$$((y_{(-1)}\ x)'(y_{(-1)}\ x))^{-1}(y_{(-1)}\ x)'y,$$

which is OLS applied to the levels- and difference-transformed equations, because large numbers of instruments relative to the sample size N would overfit the regressors being instrumented in the first stage projection. Nevertheless, Bowsher [2002] Table 1 provides Monte Carlo results showing that the finite sample bias of the system GMM estimator is not worse for (large) numbers of instruments that limit the null rejection frequency of the asymptotic Sargan–Hansen test to less than 0.0005 (for a nominal 5% test size) than for (small) numbers of instruments that achieve approximately the nominal 5% test size. This evidence suggests that the finite sample properties of the Sargan–Hansen test based on asymptotic critical values deteriorate at numbers of instruments smaller than those at which the system GMM estimator is subject to finite sample overfitting bias. We therefore cannot infer from a Sargan–Hansen test asymptotic p-value close to 1 that the coefficient estimates are likely to be subject to finite sample overfitting bias.

different measures of output from other versions of the PWT. Credit survives marginally using the PWT8.1 expenditure-based constant price international dollar output, but not using the PWT8.0 version of this variable.[38]

The specification in Column (5) of Table 1.5 contains many irrelevant variables. I try sequentially eliminating the insignificant variables trade openness, government expenditure share, inflation and black market premium in that order. The elimination of individual variables at no point causes other variables in this set to become statistically significant. A Wald test on the estimation in Column (5) for the joint exclusion of the share of government consumption and trade in output, the inflation rate and the black market premium does not reject the null hypothesis of zero coefficients, with a p-value of 0.9.[39] I present the estimates of the resulting specification in Column (6). Although the government expenditure share and inflation are absent from the set of explanatory variables, I retain them in the instrument set because they are informative for identifying the coefficient on the lagged dependent variable and their validity is not rejected by a test of overidentifying restrictions. In Table 1.5, it is reassuring that the coefficient on the lagged dependent variable in Column (6) lies between the coefficients on the lagged dependent variable when I estimate the model by ordinary least squares (Column (7)) and within groups (Column (8)).[40] A version for the restricted (Table 2 Fit 3) specification of Beck et al. [2000] appears in Appendix 1.D.

1.4 Stylised facts in finance–growth

In Section 1.3 I obtain a parsimonious specification for examining the (log-linear) relationship between growth and credit that does not suffer from problems of omitted variables or incomparability across countries. In this section I use this specification to explore the (log-linear) relationship between growth and credit over time (Subsection 1.4.1) and the potential non-linear relationship between growth and credit (Subsection 1.4.2). In Section 1.3 I also find that

[38]The growth rate in terms of trade is not significant in the model with the investment expenditure share and population growth included, so I exclude the growth rate in terms of trade from the joint model. Years of life expectancy and numbers of revolutions are not statistically significant in the presence of the other explanatory variables in Column (1) of Table 1.5. To save space, I choose not to present the estimation results for specifications with terms of trade, years of life expectancy or numbers of revolutions in Table 1.5.

[39]A similar stepwise elimination procedure is able to eliminate the same four variables from a specification that does not include the investment expenditure share and population growth to begin with.

[40]The coefficient on the lagged dependent variable $y_{i,t-1}$ obtained by the OLS estimator converges in probability to a number larger than the coefficient α on the lagged dependent variable in the data generating process (1.4) due to the positive covariance between the lagged dependent variable $y_{i,t-1}$ and the fixed effect component c_i in the error term. The lagged dependent variable on the right hand side of the data generating process (1.4) for y_{it}, in every time period t apart from the first ($t = 1$), itself depends positively on the fixed effect error component c_i, which drives this positive covariance. The coefficient on the lagged dependent variable $y_{i,t-1}$ obtained by the WG estimator converges in probability to a number smaller than the coefficient α on the lagged dependent variable in the data generating process (1.4) as the number of countries N diverges and with the number of time periods T held fixed [Nickell, 1981].

Table 1.5: Estimation results showing the effect on the specification in Table 1.4 of including potentially omitted explanatory variables and of excluding irrelevant explanatory variables from the latter specification. The estimations in Columns (1) and (2) use the first lag of the level of initial income and schooling, and the second lag of the level of credit, government, trade, inflation and the black market premium, as instruments for the differenced equations. The estimations in Columns (1) and (2) also use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government, trade, inflation and the black market premium, as instruments for the levels equations. The instrument set for the estimation results in Column (3) contains the same instruments as the specifications in Columns (1) and (2) and also adds the second lag of the level of investment as an instrument for the differenced equations and the first lag of the difference of investment as an instrument for the levels equations. The instrument set for the estimation results in Column (4) contains the same instruments as the specifications in Columns (1) and (2) and also adds the second lag of the level of population growth as an instrument for the differenced equations and the first lag of the difference of population growth as an instrument for the levels equations. The instrument set for the estimation results in Column (5) is the union of the instrument sets used in Columns (3) and (4). The estimation in Column (6) uses the first lag of the level of initial income and school, and the second lag of the level of credit, government, inflation, investment and population growth, as instruments for the differenced equations. In addition, the estimation in Column (6) also uses the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government, inflation, investment and population growth, as instruments for the levels equations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	81okd	81okd-	+inv	+pop	81okd +inv +pop	-bmp -infl(x) -gov(x) -trade	ols	wg
initial income	-0.96	-1.38	-1.98**	-1.92*	-1.89*	-2.41**	-0.81*	-7.47***
	(0.86)	(0.95)	(0.96)	(1.04)	(1.06)	(1.06)	(0.46)	(1.24)
school	0.42	0.54	1.00**	0.01	0.35	0.81	0.22*	0.09
	(0.44)	(0.46)	(0.47)	(0.54)	(0.52)	(0.54)	(0.13)	(0.54)
credit	1.43**	1.92***	0.87	1.77**	0.72	0.55	0.26	1.04*
	(0.64)	(0.64)	(0.62)	(0.73)	(0.77)	(0.65)	(0.30)	(0.58)
government	-0.72	0.58	-0.60	0.54	-0.41			
	(1.65)	(1.70)	(1.60)	(1.69)	(1.45)			
trade	0.00	0.43	-0.23	0.86	0.31			
	(0.66)	(1.03)	(0.88)	(1.19)	(0.91)			
inflation	0.01	0.02	0.01	0.02	0.01			
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)			
black market prem.	0.06	-0.37	-0.40	-0.72	-0.39			
	(1.14)	(1.27)	(0.97)	(1.19)	(0.86)			
investment			5.62***		4.34***	3.99***	3.75***	4.54***
			(1.39)		(1.33)	(1.40)	(0.78)	(1.55)
population growth				-11.36***	-10.45*	-9.78*	-4.37**	-12.20***
				(4.10)	(5.83)	(5.29)	(1.77)	(3.91)
observations	232	229	229	229	229	229	229	229
countries	62	60	60	60	60	60		60
instruments	76	76	86	86	96	76		
Hansen p-val.	0.720	0.759	0.989	0.933	0.998	0.938		
AR(1) stat.	-3.499	-3.470	-3.294	-3.470	-3.354	-3.328		
AR(2) stat.	1.127	1	1.096	0.816	0.991	1.207		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.6: Estimation results of the specification from Table 1.5 on different sample periods. The estimations in Columns (1), (3) and (4) use the first lag of the level of initial income and school, and the second lag of the level of credit, government, inflation, investment and population growth, as instruments for the differenced equations. In addition, the estimations in Columns (1), (3) and (4) also use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government, inflation, investment and population growth, as instruments for the levels equations. The estimation in Column (2) uses the second lag of the level of initial income and school, and the third lag of the level of credit, government, inflation, investment and population growth, as instruments for the differenced equations. In addition, the estimation in Column (2) also uses the first lag of the difference of initial income and school, and the second lag of the difference of credit, government, inflation, investment and population growth, as instruments for the levels equations. The estimations in Columns (5) to (8) use the first lag of the level of initial income and school, and the second lag of the level of credit, government and inflation, as instruments for the differenced equations. In addition, the estimations in Columns (5) to (8) use the contemporaneous difference of initial income and school, and the first lag of the difference of credit, government and inflation, as instruments for the levels equations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	81okd +inv +pop	77 6095	77 6085	77 6010	81okd	77 6095	77 6085	77 6010
initial income	-2.41**	-1.52*	-1.78*	-1.10	-2.39**	-1.98***	-1.90*	-1.58*
	(1.06)	(0.88)	(0.96)	(0.76)	(1.21)	(0.77)	(1.00)	(0.85)
school	0.81	1.10***	0.39	0.98***	1.35**	1.02***	0.56*	0.83***
	(0.54)	(0.35)	(0.43)	(0.30)	(0.60)	(0.29)	(0.32)	(0.26)
credit	0.55	0.48	1.46	-0.55	1.67**	1.32**	2.09***	0.39
	(0.65)	(0.86)	(1.00)	(0.55)	(0.79)	(0.59)	(0.56)	(0.51)
investment	3.99***	4.27**	2.20	4.06***				
	(1.40)	(1.87)	(2.28)	(1.24)				
population growth	-9.78*	1.35	-6.70	-0.27				
	(5.29)	(1.52)	(4.77)	(2.15)				
observations	229	339	218	528	229	452	312	661
countries	60	61	59	67	60	70	69	70
instruments	76	92	62	137	56	74	52	107
Hansen p-val.	0.938	0.990	0.614	1	0.276	0.513	0.494	0.996
AR(1) stat.	-3.328	-3.468	-2.360	-4.348	-3.517	-4.189	-3.204	-4.775
AR(2) stat.	1.207	2.306	0.869	0.664	1.273	1.259	-0.000946	-0.338

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

the association between growth and credit appears to be proxying for the well known association between growth and investment. This finding leads us to ask whether there is evidence for a relationship between investment and credit, which I explore in Subsection 1.4.3.2.

1.4.1 Sample periods

Column (1) in Table 1.6 repeats the estimation results of the parsimonious model from Section 1.3. This estimation is based on a 229-observation sample, which is the largest subset of the 288-observation sample on which Beck et al. [2000] estimate their specifications for which I have been able to construct the variables in Column (5) of Table 1.5 in Section 1.3.3. However, I have an additional 110 country–5-year observations in my dataset on the variables in Column (1) of Table 1.6 within the original 77 country, 1960–1995 sample of Beck et al. [2000]. In Column (2) of Table 1.6 I present estimation results when adding these 110 observations to the estimation sample for the same specification as in Column (1). I also lag the instrument set for the specification in Column (2) by one 5-year period so that the instrument set is not invalidated by first order serial correlation in the idiosyncratic errors, which is evidenced by statistically significant second order correlation in the differenced residuals shown in the estimation results in that Column. Adding these 110 observations and lagging the instrument set one period loses individual statistical significance on all the coefficients in the specification. In Columns (3) and (4) I present estimation results when the sample used to estimate Column (2) is adjusted to the 1960–1985 and 1960–2010 periods respectively. Only initial income retains some statistical significance in the 1960–1985 sample. Initial income loses its marginal statistical significance in the Column (4) 1960–2010 sample relative to the Column (3) 1960–1985 sample estimation results. I find that credit is statistically insignificant in the presence of investment in all three sample periods across Columns (2)-(4).

In Column (5) of Table 1.6 I present the estimation results for a specification that excludes the investment expenditure share and population growth from the explanatory variables and instrument set of the Column (1) specification. Column (5) also arises when I apply stepwise model selection to eliminate the trade and government consumption expenditure shares in output, inflation and the black market premium, which are also jointly insignificant, from the extended (Table 2 Fit 4) specification in Beck et al. [2000]. In Column (6) of Table 1.6 I add 223 country–5-year observations that are available in my dataset within the 77 countries and 1960–1995 time periods considered by Beck et al. [2000] (but not used in their 288-observation estimation sample). In Column (7) I restrict the time period of the sample in Column (6) to 1960–1985 and in Column (8) I expand the time period of the sample in Column (6) to 1960–2010. From the estimation results in Column (3), either credit or investment can individually be excluded from the specification, but individually excluding investment leads the coefficient on credit to become statistically significant in Column (7) and separate estimation results (not shown) indicate that individually excluding credit leads the coefficient on investment to become statistically significant (p-value 0.01). The fact that investment and credit are statistically significant when included on their own, but neither is individually statistically significant when they are included together, suggests that both may be important determinants of growth in the 1960–1985 sample and there is not enough information to distinguish between them in this sample. Columns (6)-(8) confirm a finding from Arcand et al. [2012] (Table 4 Fits 1-4) that the (positive log-linear) relationship between growth and credit has been attenuating through time. While the relationship between growth and credit has been attenuating, the estimation results in Columns (3) and (4) show a strengthening relationship between growth and investment.

1.4.2 Non-linearity

In this subsection I use the parsimonious specification found in Section 1.3 to investigate whether there is evidence for a non-linear relationship between growth and credit in the 77-country, 1960–2010 sample from Table 1.6 of Section 1.4.1. Arcand et al. [2012] estimate non-monotonic functional forms for the relationship between growth and credit. I begin this subsection by explaining their work and then relate my findings to theirs.

Arcand et al. [2012] find evidence for a concave quadratic relationship between growth and credit that does not seem to be driven by endogenous explanatory variables or country-specific time-invariant factors. They do so by estimating the parameters in equation (1.3) with system GMM, which is explained in Section 1.3.1, on non-overlapping 5-year aggregated data on a sample of 133 countries over the 1960–2010 period. Their specification of x in (1.3) follows Beck et al. [2000], but excludes the black market premium. Arcand et al. [2012] consider a log-linear functional form for credit like Beck et al. [2000] and also investigate a non-monotonic functional form for credit by replacing the log of the ratio of credit to output in x by the level and the squared level of the ratio of credit to output.

There are at least five reasons to be skeptical about the concave relationship between growth and credit found in Arcand et al. [2012]. First, the authors' choice of a quadratic specification does not help them to distinguish between the positive log-linear relationship found in Beck et al. [2000] and a non-monotonic relationship between growth and credit. In fact, estimating a concave quadratic relationship is exactly what one would expect if the relationship between growth and credit were monotonic, increasing and logarithmic.[41] Second, the choice of Arcand et al. [2012] to measure output in constant price US dollars means that output is not comparable

[41] Note that $\log(x) = \log(x_0) + \frac{1}{x_0}(x - x_0) - \frac{1}{2x_1^2}(x - x_0)^2$ for all $x, x_0 > 0$ and some $x_1 \in (0, x_0)$. If X, Y are random variables with $Y = \log(X)$, then a linear projection of Y onto X and X^2 gives coefficients α_1 on X and α_2 on X^2 where

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \left[\mathbb{V}\begin{pmatrix} X \\ X^2 \end{pmatrix}\right]^{-1} \operatorname{cov}\left[\begin{pmatrix} X \\ X^2 \end{pmatrix}, \log(X)\right], \tag{1.18}$$

whence

$$\alpha_1 = \frac{1}{x_0} + \frac{x_0}{x_1^2} \tag{1.19}$$

$$\alpha_2 = -\frac{1}{2x_1^2} \tag{1.20}$$

and therefore $\alpha_1 > 0$ and $\alpha_2 < 0$ for any distribution of X as long as $X > 0$ and all covariances are well defined.

across countries due to purchasing power differences between them. A third reason to be doubtful is that the authors have combined a large heterogenous group of countries, including oil exporters for which the growth determinants could be expected to be quite different, and the statistical significance of the concave quadratic relationship they find between growth and credit disappears on subsamples defined by important subsets of countries. In particular, the statistical significance of the quadratic term in the authors' Table 5 Fit 1 disappears if we restrict their estimation sample to the 98 country non-oil sample of countries analysed by Mankiw et al. [1992]. Similarly, the estimation results in Appendix 1.E show that the statistical significance of the quadratic term in the authors' Table 5 Fit 1 disappears when I again estimate with the authors' original data and methods, but restrict their estimation sample to those observations for which I am able to construct an investment variable from the WDI. A fourth and related problem with Arcand et al. [2012] is that the authors aggregate from the annual to the 5-year frequency by picking out the first observation for the explanatory variables in each 5-year window, for which I am not aware of a precedent in the empirical growth literature.[42] Fifth, a within groups estimator applied to the specification and sample in Arcand et al. [2012] Table 5 Fit 4 on the authors' original data gives an estimate (-0.19) of the coefficient on the lagged dependent variable that is larger than the -0.73 estimated by the authors using system GMM. This disparity suggests model misspecification in Arcand et al. [2012] because the within groups estimator of the coefficient on the lagged dependent variable in (1.3) is biased down in samples with large numbers of countries and few time periods, which in turn owes to the negative correlation between $-\frac{1}{T-1}y_{it-1}$ and u_{it-1} and between y_{it-1} and $-\frac{1}{T-1}u_{it-1}$ in (1.3) after the within groups transformation has been applied [Nickell, 1981]. A consequence of this misspecification in the authors' Table 5 Fit 4 is that the coefficients on explanatory variables other than the lagged dependent variable, including the coefficient on credit, are biased in large samples in unknown directions.

In what follows I specifically test for departure from the log-linear functional form for the relationship between growth and credit that is estimated in the earlier work of Beck et al. [2000], by specifying a functional form that nests both the log-linear functional form and potential non-monotonic relationships between growth and credit. I use constant price international dollar output that adjusts for differences in purchasing power across countries. I follow Beck et al. [2000] in the sample of countries used, which I explain in Appendix 1.F, and in variable timing and variable aggregation, which I explain in Section 1.2.3. I find in Section 1.3.3 that the

[42] That is, the authors use the timing assumption illustrated by the representative element $y_{i,1995} - y_{i,1990} = (\alpha - 1)y_{i,1991} + \beta' x_{i,1991} + \cdots$ in their specifications. The lagged dependent variable is thus measured in years $1961, 1966, \ldots, 2006$ rather than $1960, 1965, \ldots, 2005$ and the explanatory variables in x are not averaged in each 5-year window. It is difficult to measure the effect of this timing assumption using the authors' data because their dataset does not contain the necessary measurements for the year 1960. If I eliminate the 1960–1965 period from their estimation sample and then correct the authors' timing of the lagged dependent variable and properly average the explanatory variables over each 5-year window, I eliminate the statistical significance of the coeffcients on the linear and quadratic credit terms in the authors' Table 5 Fit 4 specification.

estimated coefficient on the lagged dependent variable in the specification with a log-linear functional form lies between the coefficients on the lagged dependent variable estimated by ordinary least squares and within groups, which points away from the misspecification that casts doubt on the results in Arcand et al. [2012].

In Column (1) of Table 1.7 I repeat the estimation results for the parsimonious specification with investment and population growth explanatory variables over the 1960–2010 period seen in Table 1.6, but I replace the logarithm of the ratio in equation (1.1) of credit to output by the level of the ratio in equation (1.1) of credit to output and the square of this ratio. Such a non-linear functional form for credit is found in Arcand et al. [2012] to have strong explanatory power for growth. The estimation results in Column (1) show marginal evidence for a concave quadratic relationship between growth and credit, which mirrors the finding from Arcand et al. [2012] Table 5 Fit 4.[43] I estimate a turning point in the quadratic relationship of 0.6, which is within one standard error of the turning point of 0.9 obtained by Arcand et al. [2012] for the 1960–2010 sample.

The estimation results in Table 1.7 Column (2) mirror the result in Arcand et al. [2012] Table 4 Fit 4 that there is no evidence for the log-linear relationship between growth and credit in the 1960–2010 sample.[44] Column (3) of Table 1.7 presents estimation results for a specification that adds an indicator variable for observations where the ratio of credit to output exceeds 75% and an interaction term between this indicator variable and the logarithm of the ratio of credit to output. The threshold 75% is the median of the ratio of credit to output in the dataset; Column (4) presents estimation results for an analogous specification to that in Column (3) with an alternative threshold of 90%, which is the turning point estimated in Arcand et al. [2012] Table 5 Fit 4. The specification estimated in Columns (3) and (4) is a complicated piecewise log-linear functional form, but it has two advantages. First, it nests the log-linear functional form in Column (2) for the relationship between growth and credit investigated by King and Levine [1993] and Beck et al. [2000]. Second, it does not restrict the 'effect' of credit at low levels of credit to be opposite in sign to the 'effect' of credit at high levels of credit.[45] In other

[43] The estimation results in Column (1) of Table 1.7 differ from the estimation results in Arcand et al. [2012] Table 5 Fit 4 because in Column (1) of Table 1.7 I have used measures of output adjusted for purchasing power differences across countries, a country sample determined by the largest subsample of the 77 countries of Beck et al. [2000] for which I have been able to obtain data, a parsimonious specification as selected in Section 1.3.3 and a smaller instrument set following Beck et al. [2000]. In Appendix 1.E I provide exact replication results of the estimation results in Arcand et al. [2012] using the authors' original data and methods.

[44] Note that the ratio of credit to output in Column (1) is unitless, equalling 1 when credit equals output, to match Arcand et al. [2012], while the ratio of credit to output in Column (2) inside the logarithm is in percentage point units, equalling 100 when credit equals output, to match Beck et al. [2000].

[45] The quadratic relationship in Column (1) between growth and credit studied by Arcand et al. [2012], by contrast, restricts the 'effect' of credit at low levels of credit to be opposite in sign to the 'effect' of credit at high levels of credit. Note if the coefficient on the squared ratio of credit to output is not specified to be exactly zero then the quadratic specification only allows U-shaped or inverse U-shaped relationships to be estimated from the data. Now it is nevertheless possible that over a range of ratios of credit to output of practical interest, say 0 to 3, the estimated quadratic relationship between growth and credit may be monotonic. However, such a range of

words, the piecewise log-linear functional form for the relationship between growth and credit specified here allows non-monotonic relationships to be estimated from the data.

The specifications whose estimation results I present in Columns (5)–(8) add the (log) expenditure share of investment in output and the Solow–Swan population growth rate to the specifications whose estimation results I present in Columns (1)–(4). The marginal significance of the quadratic credit terms in Column (1) suggests that the quadratic relationship between growth and credit may not be robust. Indeed, when we add the investment expenditure share in output and the population growth rate to the specification in Column (5), the coefficients on the linear and quadratic terms in the ratio of credit to output lose their individual significance.[46] Interestingly, a Wald test finds some evidence (p-value 0.0397) for the linear and quadratic credit terms jointly, which suggests a potential explanatory role for credit even in the presence of investment, but the quadratic functional form suggested by Arcand et al. [2012] certainly cannot be viewed as a stylised fact. Appendix 1.E shows that the inclusion of the investment expenditure share makes linear and quadratic terms in the ratio of credit to output (individually) statistically insignificant also using the data and methods that exactly replicate Arcand et al. [2012].

In Column (6) of Table 1.7 I repeat the estimation results for the parsimonious specification with investment and population growth explanatory variables over the 1960–2010 period seen in Table 1.6 of Section 1.4.1. Once again we have evidence that the credit term is proxying for investment in the sense that the point estimate of the coefficient on the logarithm of the ratio of credit to output declines from a positive value in Column (2) to a negative value in Column (6) with the inclusion of investment and population growth.

The relationship between growth and credit estimated in Column (7) is decreasing below and above the 75% credit threshold, in the presence of investment and population growth. However, a Wald test of the joint exclusion of all three credit variables in Column (7) finds some evidence (p-value 0.0374) that the credit terms cannot be excluded from the specification in Column (7), despite the fact that none of these three coefficients are individually statistically significant in Column (7). This finding suggests that there may be a functional form for credit

credit to output of practical interest is somewhat arbitrary and there is no guarantee that the estimates of the quadratic relationship between growth and credit over that range will be montonic. The point is that Arcand et al. [2012] produce estimates of a non-monotonic (quadratic) relationship between growth and credit over a range of practical interest and we do not know from their work whether such a non-monotonic relationship over this range of practical interest is due to their choice of a quadratic specification, which guarantees (in the above sense) non-monotonicity over the entire real line, or whether the relationship between growth and credit over a range of practical interest is in fact non-monotonic in the data generating process. Another related concern with the quadratic specification in Arcand et al. [2012] is that it is a necessarily symmetric relationship around its turning point; we may wish to allow for asymmetries.

[46] The point estimate of the coefficient on the (linear) ratio of credit to output switches from positive in Column (1) to negative in Column (5), which shifts the implied turning point from a positive level of the ratio of credit to output to a negative level of this ratio. The ratio of credit to output can of course never be negative, so the relationship between growth and credit being estimated in Column (5) is monotonic over a range of ratios of credit to output of practical interest.

with explanatory power for growth even in the presence of investment and population growth. Indeed, the estimation results in Column (8) find evidence that the slope of the relationship between growth and the ratio of credit to output is more negative above a 90% ratio of credit to output than below, in the presence of investment and population growth. Using this piecewise log-linear functional form for the relationship between growth and credit with a turning point at 90%, I find a negative relationship between growth and credit at high levels of credit in the presence of investment, which is not evident from the quadratic functional form in Column (5).

The positive log-linear relationship between growth and credit found in King and Levine [1993] and Beck et al. [2000] therefore seems to be proxying for the well studied relationship between growth and investment. I do find evidence in Columns (3) and (4) for a hump-shaped relationship between growth and credit, but the increasing component of this relationship at low levels of credit is proxying for the well studied increasing relationship between growth and investment. The component of the hump-shaped relationship between growth and credit that relates to the decreasing relationship between growth and credit at high levels of credit does not work through investment. In Section 1.4.3 I investigate the shapes of the relationships between growth and investment and between investment and credit in more detail.

1.4.3 Growth, investment and credit

In Sections 1.4.1 and 1.4.2 I show that the inclusion of investment eliminates the positive log-linear relationship between growth and credit, but does not eliminate the negative log-linear relationship between growth and credit at levels of credit above 90% of output in the 1960–2010 sample period. These findings raise the natural question of whether there is a positive log-linear relationship between investment and credit, and the extent to which the relationships between growth and investment and between investment and credit could explain the concave relationship between growth and credit that prevails when we do not control for investment. In particular, we may ask if there is any evidence for departures from log-linearity in the relationships between growth and investment and between investment and credit. In this section I show evidence for positive log-linear relationships between growth and investment (Section 1.4.3.1) and between investment and credit (Section 1.4.3.2).

1.4.3.1 Growth and investment

In Table 1.8, I check whether there is a role for a non-linear term in investment in explaining growth using a piecewise log-linear form and the selected model with population growth from Section 1.3 for the 77 countries in the 1960–2010 sample period. I choose thresholds for the median (21%) and 75$^{\text{th}}$ percentile (25%) of the distribution of the expenditure share of investment in output across the whole dataset. In both Columns (1) and (2) of Table 1.8, the log-linear term in the ratio of investment to output is statistically significant, while the interaction terms are not significant. The Wald tests reported in the last row of Table 1.8 do not find evidence for departure from log-linearity in the relationship between growth and investment. So the relationship between growth and investment appears to be monotonic, increasing and logarithmic.

1.4.3.2 Investment as the dependent variable

In the specifications underlying the estimation results in Table 1.9, I put the (log) average annual investment-to-GDP ratio in constant price USD from WDI on the left-hand side. The dependent variable is therefore the same variable that displaces credit in the growth equations presented above. On the right-hand side I retain all the explanatory variables other than the investment expenditure share in output from the parsimonious model for growth selected in Section 1.3. I keep the same constant price international dollar output from PWT8.1 and country–5-year estimation samples introduced in Section 1.4.1 above. The last row of the table shows the p-values from Wald tests for the exclusion of all credit variables in each specification. Note from the description in the caption to Table 1.9 that I have lagged the instruments in the specifications of Columns (1), (3), (5) and (6) by one period relative to the instrument sets in Columns (2) and (4) so that the presence of first-order serial correlation in the idiosyncratic errors (i.e. second-order serial correlation in the differenced idiosyncratic errors) in Columns (1), (3), (5) and (6) does not invalidate the instrument sets used to estimate the specifications in those Columns.

The estimation results in Columns (1) and (2) of Table 1.9 show that the logarithm of the ratio of credit to output enters the specification for the expenditure share of investment in output with a statistically significant positive coefficient in both the 1960–2010 and 1960–1995 samples. Beck et al. [2000] find a similar positive log-linear relationship between investment and credit in their Table 4. The positive coefficients on credit I find in Columns (1) and (2) suggest that the positive log-linear relationship between growth and credit found in Beck et al. [2000] is indeed proxying for the well known positive log-linear relationship between growth and investment. This conclusion follows because I also find in Section 1.4.1 that the inclusion of investment substantially decreases the point estimate of the coefficient on credit and in the 1960–1995 sample eliminates its statistical significance. To the extent that the instruments are valid, credit seems to exert a positive effect on investment, which in turn exerts a positive effect on growth, and there does not seem to be any other positive effect of credit on growth.

Columns (3) and (4) show estimation results for a quadratic specification between investment and credit. The joint and marginal tests in Column (3) do not show evidence for a quadratic relationship between investment and credit in the 1960–2010 sample, which is consistent with the marginal evidence in Section 1.4.2 that a quadratic form for credit has some role for explaining growth in the 1960–2010 sample even in the presence of investment and population growth.

Column (4) shows marginal ($p < 0.1$) evidence for a departure from linearity in the relationship between investment and credit, which is consistent with the monotonic, increasing log-linear relationship found in Column (2).[47]

Columns (5) and (6) show evidence for a positive log-linear relationship between investment and credit. Not presented in Table 1.9, Wald tests for the exclusion of the indicator and interaction credit terms only in Columns (5) and (6) show no evidence (p-values of 0.2002 and 0.1720 respectively) for departure from log-linearity in the relationship between investment and credit. Therefore the relationship between investment and credit appears to be monotonic, increasing and logarithmic.

Nevertheless, Section 1.4.2 shows that the downward-sloping relationship between growth and credit at levels of credit above 90% of output remains even in the presence of a log-linear investment term. Combined with the strong support for a log-linear relationship between investment and credit in Table 1.9, this suggests two channels through which credit affects growth. To the extent that the instrument sets are valid, there seems to be a positive log-linear effect of credit on investment, which in turn exerts a positive log-linear effect on growth. In addition, there seems to be a negative log-linear effect of credit on growth at levels of credit above 90% of output, which does not seem to work through investment.

1.5 Conclusion

The finance–growth literature has found a strong positive relationship between growth and credit that cannot be explained by unobserved time-invariant factors affecting growth. Recent contributions to the literature have also found evidence for a non-monotonic concave relationship between growth and credit.

To date, these stylised facts have been questionable because they have relied on measures of output that did not adjust for differences in purchasing power across countries. Using data that adjust for differences in purchasing power across coutries, I find that the strong positive association between credit and growth indeed prevails. There is also some evidence for a non-monotonic concave relationship between growth and credit, although only using a piecewise log-linear functional form, not using a quadratic functional form.

The empirical finance–growth literature has also not investigated the role for typical Solow-type growth determinants in its specifications. Of the Barro–Lee growth determinants studied by Caselli et al. [1996], I find that the Solow–Swan model growth determinants—the expenditure share of investment in output and population growth—are important omitted variables in canonical finance–growth specifications. This finding leads us to ask whether the increasing log-linear and non-monotonic quadratic relationships between growth and credit documented

[47] Footnote 41 explains that an estimated concave quadratic relationship is expected if the true relationship is monotonic, increasing and logarithmic.

in Beck et al. [2000] and Arcand et al. [2012] respectively could be explained by the usual endogeneity concerns that arise through omitted variables.

Indeed, based on output data that adjust for differences in purchasing power across countries and a sample of countries investigated in the earlier literature over the 1960–2010 period, evidence for a log-linear relationship between growth and credit disappears when I control for investment. In fact, the point estimates of coefficients on a log-linear credit term become negative in the 1960–2010 period. For the quadratic specification considered by Arcand et al. [2012], the coefficients on the linear and squared credit terms become individually insignificant when I control for investment, although the two credit coefficients remain jointly significant (p-value 0.0397).

I specify and estimate a piecewise log-linear functional form for the relationship between growth and credit that has the advantages over the quadratic functional form of nesting the log-linear functional form from earlier work and allowing for non-monotonic relationships to be estimated from the data. The inclusion of Solow–Swan model growth determinants in a piecewise log-linear specification of the relationship between growth and credit shows evidence for a departure from log-linearity above a 90% threshold for the ratio of credit to output. In particular, it shows that growth and credit are negatively associated at levels of credit above 90% of output, even in the presence of investment, which suggests that higher levels of credit imply a lower steady state level of output, but not by reducing investment.

The fact that controlling for investment eliminates the positive log-linear relationship between growth and credit (at any level of credit), combined with the strong positive log-linear relationships between growth and investment and between investment and credit, leads me to conclude that any positive effect of credit on growth (at any level of credit) would seem to operate through increased investment. Conversely, the fact that controlling for investment does not eliminate the negative log-linear relationship between growth and credit at high ($>$ 90% of output) levels of credit, combined with the lack of evidence for departures from log-linearity in the relationships between growth and investment and between investment and credit, leads me to conclude that any negative effect of credit on growth at high ($>$ 90% of output) levels of credit would seem to operate independently of investment.

1.A Fixing and updating BLL (2000) T2F3

In the table below I present estimation results that correct the standard errors of the restricted specification (Table 2 Fit 3) in Beck et al. [2000] for finite sample bias and update the authors' estimation results using data constructed from modern versions of the authors' original sources. This section is analogous to Section 1.3.1. Column (1) below starts from the closest possible estimation results obtainable in Stata to the authors' original estimation results that is presented in the replication exercise in Section 1.3.1. Unlike in the full (Table 2 Fit 4) specification, the estimations in Column (1) use the population orthogonality restriction assumptions

$$\begin{aligned} \mathbb{E}(\pi'_{i,t-s}\,\eta'_{i,t-1-s})\Delta u_{it} &= 0 \quad \text{for all } (i,t) \text{ and } s = 1,2,3 \\ \mathbb{E}(\Delta\pi'_{i,t-s}\,\Delta\eta'_{i,t-1-s})(c_i + u_{it}) &= 0 \quad \text{for all } (i,t) \text{ and } s = 0,1,2, \end{aligned} \tag{1.21}$$

to identify the parameters in (1.3), where η in this context contains only the credit variable because other variables from the full (Table 2 Fit 4) specification are dropped. In other words, Beck et al. [2000] use lags 1 to 3 of initial income ($y_{i,t-1}$) and schooling and lags 2 to 4 of credit as instruments for the explanatory variables in the first-differenced equations; they use lags 0 to 2 of the first-differences of initial income and schooling and lags 1 to 3 of the first-difference of credit as instruments for the explanatory variables in the levels equations. In their estimations, Beck et al. [2000] implicitly exclude some of these available instruments from their sample. Column (2) of the table adds these instruments back into the estimation sample, which attenuates the point estimates of the coefficients on initial income and schooling by a third. Column (3) implements a finite sample correction to the two-step standard error estimates due to Windmeijer [2005], which renders the coefficients on initial income and schooling statistically insignificant. Column (4) recovers some evidence for the lagged dependent variable using the initial moment weight matrix explain in Section 1.3.1. Column (5) replaces the authors' original variables by variables constructed from modern versions of the authors' original data sources according to Section 1.2, which reduces the sample by 22 country–5-year observations, but does not change the qualitative characteristics of the bias-corrected estimation results based on the authors' original data. Column (6) restricts the instrument set (1.21) to the same instrument lags as used for the unrestricted (Table 2 Fit 4) specification (1.5) and shows that the statistical significance of the coefficient on credit is maintained.

	(1)	(2)	(3)	(4)	(5)	(6)
	t2f3	zu	w	w h(3)	new	t2f4z
initial income (BLL00)	-1.53***	-1.07***	-1.07	-1.36*		
	(0.31)	(0.33)	(0.91)	(0.83)		
school (BLL00)	3.02***	1.99***	1.99	1.83		
	(0.75)	(0.71)	(2.29)	(2.04)		
credit (BLL00)	2.41***	2.15***	2.15***	2.58***		
	(0.18)	(0.21)	(0.65)	(0.55)		
initial income (WDI USD)					-0.75*	-0.36
					(0.45)	(0.63)
school (BL13)					0.37	0.15
					(0.25)	(0.29)
credit (IFS)					2.06***	1.93***
					(0.52)	(0.69)
observations	288	288	288	288	266	266
countries	77	77	77	77	73	73
instruments	61	61	61	61	65	36
Hansen p-val.	0.201	0.138	0.138	0.178	0.434	0.227
AR(1) stat.	-3.052	-3.118	-2.996	-3.042	-3.418	-3.403
AR(2) stat.	-0.660	-0.626	-0.625	-0.567	-0.734	-0.411

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.B Effect of credit under a PPP adjustment PWT5.6-8.1

In this section, I extend the analysis of Section 1.3.2 to previous versions of the PWT and other measures of constant price international dollar output from PWT8.1 in the full (Table 2 Fit 4) and restricted (Table 2 Fit 3) specifications of Beck et al. [2000]. This section shows that the (log-linear) relationship between growth and credit is robust to various ways of measuring constant price international dollar output.

In the table below, Column (1) starts from the bias-corrected estimation of the full (Table 2 Fit 4) specification of Beck et al. [2000] based on modern data, which is presented in Section 1.3.1. The column labels indicate versions 5.6, 6.1, 7.1, 8.0 and 8.1 of the PWT. I maintain the shares of government consumption and trade in output in constant price 2000 US dollars from the WDI throughout, which is explained in Section 1.3.2. In each column (2)-(8) of the table, I replace output on the left- and right-hand sides of the specification in Column (1) by the indicated PWT constant price international dollar version and I replace the denominator of the ratio of credit to output by current price local currency output from the national accounts that accompany the same indicated PWT version.[48,49] For versions 5.6, 6.1 and 7.1 of the PWT, I use the fixed-weight Laspeyres versions of constant price international dollar output. Columns (5) and (6) of the table use output-based versions of constant price international dollar output that allow for differences in the relative prices of exports and imports across countries, while Columns (7) and (8) of the table use expenditure-based versions of constant price international dollar output that do not allow for such differences and are hence more comparable with earlier versions of the PWT.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	wdi usd	56lkd	61lkd	71lkd	80okd	81okd	80ekd	81ekd
initial income	-0.58	-1.07	-1.07	-0.81	-1.59*	-0.96	-1.77*	-1.62*
	(0.56)	(1.05)	(0.77)	(0.58)	(0.96)	(0.86)	(0.95)	(0.85)
school	0.33	0.07	0.30	0.29	0.65*	0.42	0.66	0.60
	(0.30)	(0.34)	(0.29)	(0.28)	(0.38)	(0.44)	(0.51)	(0.48)
credit	1.22**	1.70**	1.15**	1.23***	1.84***	1.43**	1.82**	1.90***
	(0.51)	(0.73)	(0.49)	(0.43)	(0.68)	(0.64)	(0.74)	(0.64)
government	-0.43	0.20	-0.52	-0.63	-1.21	-0.72	-0.62	-0.93
	(0.97)	(0.96)	(1.14)	(1.38)	(1.94)	(1.65)	(1.98)	(1.88)
trade	0.86	-0.34	0.34	0.62	-0.48	0.00	-0.00	-0.00
	(0.61)	(0.92)	(0.81)	(0.69)	(1.06)	(0.66)	(0.96)	(1.11)
inflation	-0.00	-0.01	-0.00	-0.00	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
black market prem.	-0.92*	-0.94	-1.25***	-1.56***	0.19	0.06	0.09	-0.01
	(0.48)	(0.72)	(0.48)	(0.49)	(0.86)	(1.14)	(1.11)	(1.14)
observations	235	172	228	231	232	232	232	232
countries	65	58	61	63	62	62	62	62
instruments	76	61	76	76	76	76	76	76
Hansen p-val.	0.722	0.378	0.850	0.709	0.621	0.720	0.811	0.777
AR(1) stat.	-3.009	-2.742	-3.126	-2.774	-3.693	-3.499	-3.458	-3.461
AR(2) stat.	-1.213	-0.262	-0.583	-0.0771	0.873	1.127	0.216	0.158

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The following table shows the effect of the same purchasing power adjustments to output (and measurement of output in the denominator of the ratio of credit to output) as discussed

[48]I leave unaffected the national currency consumer price index adjustment recommended by Beck et al. [2000] for the numerator of the ratio, which is designed to match the credit stock numerator with an output flow denominator.

[49]Version 5.6 of the PWT does not come with an accompanying set of national accounts, so I leave the denominator of the ratio of credit to output in this case unaltered.

above, for the restricted (Table 2 Fit 3) specification of Beck et al. [2000].

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	new	56l	61l	71l	80o	81o	80e	81e
initial income	-0.75*	-1.84	-0.84	-1.08	-0.15	-0.13	-1.04	-1.03
	(0.45)	(1.33)	(0.92)	(0.79)	(0.86)	(0.85)	(0.74)	(0.75)
school	0.37	0.32	0.05	0.22	0.13	0.07	0.30	0.29
	(0.25)	(0.42)	(0.23)	(0.23)	(0.27)	(0.25)	(0.22)	(0.23)
credit	2.06***	2.60***	1.99***	2.19***	1.34**	1.40**	1.96***	1.97***
	(0.52)	(0.66)	(0.44)	(0.52)	(0.60)	(0.59)	(0.55)	(0.49)
observations	266	206	272	275	270	270	270	270
countries	73	67	71	73	70	70	70	70
instruments	65	53	66	66	66	66	66	66
Hansen p-val.	0.434	0.401	0.414	0.276	0.332	0.430	0.290	0.262
AR(1) stat.	-3.418	-3.179	-3.777	-3.506	-3.776	-3.742	-3.364	-3.441
AR(2) stat.	-0.734	-0.105	0.117	-0.602	0.850	0.953	0.390	0.434

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.C Credit and investment together as explanatory variables

In this section, I extend the conclusions about the effect of including investment in a specification with credit, which are presented in Section 1.3.3, to alternative measures of investment and alternative measures of output. The results show that investment displaces the statistically significant coefficient on credit with some regularity across different choices for the measurement of growth and investment.

1.C.1 Alternative measures of investment

In the following table, Column (1) presents the full (Table 2 Fit 4) specification of Beck et al. [2000] using output-based constant price international dollar output from the PWT8.1 that is introduced in Section 1.3.2. Columns (2)-(7) each introduce a measure of the investment expenditure share in output to the specification in Column (1). Column (2) introduces a constant price 2000 US dollar share of total investment expenditure in output from the WDI to the specification, which is presented in Section 1.3.3. In Column (3), investment is measured as *fixed* investment expenditure from the WDI. In Column (4) the investment expenditure share is measured in current price international dollars as a share of current price international dollar chain-linked output, according to the PWT8.1. In Column (5), the investment expenditure share is measured in current price local currency from the WDI.

Columns (6) and (7) attempt to adjust the current price international dollar investment expenditure share provided by PWT8.1 for inflation in the numeraire currency to produce estimates of the constant 2005 price international dollar investment expenditure share in output. These conversions are not rigorous because they implicitly treat chain-linked output as if it were fixed-weight output; nevertheless they are approximations that might contain useful illustrative information. In the underlying country–year data, the approximation in Column (6) adjusts the numerator of the current price international dollar investment expenditure share in output for inflation in the numeraire currency between each year and 2005. The approximation in Column (7) adjusts both the numerator and the denominator of the current price international dollar investment expenditure share in output for inflation in the numeraire currency. If we temporarily let $t = 1960, 1961, \ldots, 1995$ denote years then the two formulae are

$$\texttt{csh\_i}_{it} \frac{\texttt{cgdpo}_{it}}{\texttt{rgdpo}_{it}} \frac{\texttt{pl\_i}_{\mathrm{US},2005}}{\texttt{pl\_i}_{\mathrm{US},t}} \tag{1.22}$$

$$\texttt{csh\_i}_{it}\texttt{pl\_gdpo}_{\mathrm{US},t} \frac{\texttt{pl\_i}_{\mathrm{US},2005}}{\texttt{pl\_i}_{\mathrm{US},t}} \tag{1.23}$$

for Columns (6) and (7) respectively, where i denotes the country and text in `fixed width font` denotes variable names from PWT8.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	81o	kd tot	kd fix	cid	clc	rid(n)	rid(nd)
initial income	-0.96	-1.98**	-1.53	-1.69*	-0.87	-1.28	-1.65*
	(0.86)	(0.96)	(0.95)	(0.87)	(0.91)	(0.91)	(0.91)
school	0.42	1.00**	0.84**	0.39	0.45	0.28	0.51
	(0.44)	(0.47)	(0.39)	(0.33)	(0.39)	(0.40)	(0.37)
credit	1.43**	0.87	0.77	0.78	0.27	0.83	0.83
	(0.64)	(0.62)	(0.72)	(0.81)	(0.62)	(0.69)	(0.78)
government	-0.72	-0.60	-1.22	-0.39	-1.05	-0.24	-0.42
	(1.65)	(1.60)	(1.62)	(1.33)	(1.53)	(1.90)	(1.75)
trade	0.00	-0.23	-0.53	-0.49	0.16	0.06	-0.13
	(0.66)	(0.88)	(0.96)	(0.70)	(0.73)	(0.85)	(0.83)
inflation	0.01	0.01	0.02	0.00	0.00	0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
black market prem.	0.06	-0.40	1.18	0.30	-0.28	0.01	-0.01
	(1.14)	(0.97)	(1.09)	(1.10)	(1.21)	(1.18)	(1.22)
investment		5.62***	5.02***	2.75**	5.62***	2.13*	2.43*
		(1.39)	(1.76)	(1.13)	(1.67)	(1.26)	(1.25)
observations	232	229	226	232	232	232	232
countries	62	60	61	62	62	62	62
instruments	76	86	86	86	86	86	86
Hansen p-val.	0.720	0.989	0.943	0.985	0.973	0.971	0.977
AR(1) stat.	-3.499	-3.294	-3.194	-3.319	-3.294	-3.353	-3.340
AR(2) stat.	1.127	1.096	0.865	1.047	1.425	1.086	1.031

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.C.2 Alternative measures of output

In this section, I expand on the results in Section 1.3.3 and Appendix 1.B by including the investment expenditure share in constant 2000 price USD in the bias-corrected version of the full (Table 2 Fit 4) specification of Beck et al. [2000] based on modern data. I try alternative measures of output adjusted for purchasing power differences across countries available from the PWT. Column (1) in the table below adds investment to the bias-corrected specification with modern data and output in constant 2000 price USD from the WDI; I present estimation results for this specification without the investment explanatory variable in Section 1.3.1. Columns (2)-(8) replace output on the left-hand side, right-hand side and in the denominator of the ratio of credit to output by analogues from different versions in the PWT. The labels of

Columns (2)-(8) match those in Appendix 1.B, so the reader can refer to that section for details of the output variables in the specifications in each column.

The investment expenditure share in constant 2000 price USD displaces the ratio of credit to output with output measured according to fixed-weight Laspeyres or chain-linked Fischer constant price international dollars from the PWT7.1-8.0. With PWT8.1 constant price international dollar output, a marginally significant coefficient on credit survives under the expenditure-based output measure, but not under the output-based output measure. By comparing to the appropriate column in the table in Appendix 1.B, we see that the point estimate of the coefficient on credit declines by about one third, or equivalently one standard deviation, with the inclusion of an investment explanatory variable; even though the coefficient on credit is marginally significant in Column (8) in the presence of investment, about one third of any 'effect' on growth previously attributable to credit seems to be proxying for the well known 'effect' of investment on growth. We see some evidence for an independent role for credit when using PWT6.1 fixed-weight Laspeyres constant price international dollar output. There does not seem to be enough data in the Beck et al. [2000] sample on PWT5.6 variables to infer much about that specification.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	wdi usd	56lkd	61lkd	71lkd	80okd	81okd	80ekd	81ekd
initial income	-0.90*	-0.45	-1.31**	-0.69	-1.86**	-1.98**	-1.94*	-2.04**
	(0.51)	(1.17)	(0.66)	(0.64)	(0.91)	(0.96)	(1.02)	(0.91)
school	0.67*	0.05	0.53*	0.46*	0.80*	1.00**	0.98*	0.92*
	(0.37)	(0.56)	(0.28)	(0.26)	(0.44)	(0.47)	(0.54)	(0.49)
credit	0.64	1.27	0.97**	0.34	0.92	0.87	1.02	1.21*
	(0.52)	(0.99)	(0.44)	(0.43)	(0.75)	(0.62)	(0.84)	(0.73)
government	-0.79	0.07	-1.03	-0.56	0.05	-0.60	-0.32	-0.42
	(1.29)	(0.99)	(1.20)	(1.15)	(1.59)	(1.60)	(1.61)	(1.49)
trade	-0.26	0.07	-0.28	0.19	-0.72	-0.23	-0.02	-0.53
	(0.64)	(0.67)	(0.59)	(1.16)	(0.92)	(0.88)	(0.89)	(0.78)
inflation	-0.00	-0.01**	-0.00	-0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
black market prem.	0.21	-0.72	-0.01	-0.72	-0.50	-0.40	-0.92	-0.78
	(0.81)	(1.09)	(0.54)	(0.84)	(0.96)	(0.97)	(1.19)	(1.03)
investment	3.78***	0.51	2.96***	2.95**	5.28***	5.62***	4.18***	4.42***
	(1.09)	(1.55)	(1.07)	(1.15)	(1.43)	(1.39)	(1.47)	(1.48)
observations	232	170	225	228	229	229	229	229
countries	63	56	59	61	60	60	60	60
instruments	86	69	86	86	86	86	86	86
Hansen p-val.	0.933	0.862	0.967	0.970	0.994	0.989	0.992	0.988
AR(1) stat.	-2.647	-2.568	-2.722	-2.505	-3.357	-3.294	-3.226	-3.222
AR(2) stat.	-0.914	-0.299	-0.373	0.243	0.454	1.096	0.274	0.279

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

1.D Check BLL00T2F3 for omitted variables

In this Section I show the evidence for including an investment expenditure share explanatory variable in the bias-corrected specification from Section 1.B that uses PWT8.1 output-based constant price international dollar output. In the following table, Column (1) presents the specification from Section 1.B. Columns (2) and (3) present estimation results for specifications that include a constant 2000 price USD investment expenditure share and a Solow-type population growth rate respectively in the Column (1) specification. Not shown here, the coefficient on the lagged dependent variable in Column (2) lies between the corresponding coefficients estimated by ordinary least squares and within groups.

	(1)	(2)	(3)
	81o	+inv	+pop
initial income	-0.13	-1.25	-0.31
	(0.85)	(1.14)	(0.78)
school	0.07	0.65	0.16
	(0.25)	(0.59)	(0.26)
credit	1.40**	1.00	1.72***
	(0.59)	(0.71)	(0.54)
investment		4.66**	
		(2.05)	
population growth			0.87
			(1.57)
observations	270	236	270
countries	70	61	70
instruments	66	83	86
Hansen p-val.	0.430	0.958	0.923
AR(1) stat.	-3.742	-3.382	-3.786
AR(2) stat.	0.953	1.404	0.904

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

1.E Including investment in ABP (2012) specifications

In this section, I illustrate the fragility of the non-linear relationship between growth and credit found by Arcand et al. [2012]. Columns (1) and (4) exactly replicate the estimation results of Arcand et al. [2012] (Table 5 Fits 1 and 4), which correspond to sample periods 1960–1995 and 1960–2010 respectively, using the authors' original data and methods. Columns (2) and (5) restrict the estimation samples to those country–5-year observations for which I have data on the constant 2000 price USD investment expenditure share in output. These sample restrictions reveal the sensitivity of the quadratic relationship that Arcand et al. [2012] find in the 1960–1995 period to their choice of countries. Columns (3) and (6) show the effect on the estimation results of including an investment explanatory variable. In particular, the coefficient on the squared ratio of credit to output in the 1960–2010 period becomes insignificant in the presence of investment.

	(1)	(2)	(3)	(4)	(5)	(6)
	t5f1	t5f1-	t5f1+inv	t5f4	t5f4-	t5f4+inv
initial income	-0.71*	-0.60	-0.31	-0.73**	-0.63*	-0.28
	(0.38)	(0.45)	(0.36)	(0.31)	(0.38)	(0.33)
school	0.98	1.03	1.30*	2.27***	1.62***	1.72***
	(0.76)	(0.92)	(0.78)	(0.61)	(0.62)	(0.62)
credit	8.72***	5.47**	1.94	3.63**	3.16	0.61
	(2.78)	(2.58)	(2.83)	(1.73)	(1.98)	(1.72)
(credit)^2	-3.03*	-1.77	-0.87	-2.02***	-1.89**	-0.93
	(1.64)	(1.79)	(2.00)	(0.73)	(0.79)	(0.75)
government	-2.76***	-2.28***	-1.89**	-1.46**	-0.51	-1.26**
	(0.65)	(0.84)	(0.93)	(0.74)	(0.82)	(0.50)
trade	1.78***	1.55**	0.34	1.09**	1.06**	-0.10
	(0.59)	(0.68)	(0.41)	(0.51)	(0.48)	(0.40)
inflation	0.01	0.18	0.08	-0.27	0.10	-0.07
	(0.22)	(0.18)	(0.16)	(0.21)	(0.19)	(0.17)
investment			3.46***			3.52***
			(0.93)			(0.68)
observations	549	393	393	917	700	700
countries	107	85	85	133	114	114
instruments	147	147	167	318	318	362
Hansen p-val.	0.999	1.000	1	1	1	1
AR(1) stat.	-3.753	-3.072	-3.056	-5.386	-4.537	-4.080
AR(2) stat.	-0.363	-0.0469	0.132	-1.607	-0.686	-0.944
joint p-val.	0.000589	0.0242	0.689	0.00682	0.00626	0.0350

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.F Sample lists

The following table provides a list of the country–5-year observations in each of the samples listed in Table 1.1:

iso3	BLL00 T2F4	WDI USD	PWT8.1 USD	PWT8.1 PPPo	PWT8.1 PPPo	iso3	BLL00 T2F4	WDI USD	PWT8.1 USD	PWT8.1 PPPo	PWT8.1 PPPo
ARG	7-9				4-13	JAM	6-10				
AUS	6-10	6-10	6-10	6-10	4-13	JPN	6-10	6-10	6-10	6-10	6-13
AUT	6-10	6-10	6-10	6-10	6-13	KEN	7-10	7-10	7-10	7-10	4-13
BEL	6-10	6-10	6-10	6-10	6-13	KOR	8-10	8-10	8-10	8-10	5-13
BOL	6-10	6-10	6-10	6-10	4-13	LKA	6-10	9-10	9-10	9-10	4-13
BRA	6-10	8-10	8-10	8-10	8-13	LSO	10-10	10-10	10-10	10-10	6-13
CAF	10-10				12-13	MEX	6-10	6-10	6-10	6-10	4-13
CAN	6-10	6-10	6-10	6-10	6-13	MLT	9-9	9-9	9-9	9-9	6-13
CHE	6-10	8-10	8-10	8-10	8-13	MUS	10-10	10-10	10-10	10-10	7-13
CHL	6-10	6-10	6-10	6-10	4-13	MWI	10-10				13-13
CMR	8-10	8-10	8-10	8-10	5-13	MYS	6-10	6-10	6-10	6-10	4-13
COD	8-9					NER	7-9				
COG	9-10	9-10	9-10	9-10	9-13	NIC	9-9	9-9			
COL	7-10	7-10	7-10	7-10	4-13	NLD	6-10	6-10	6-10	6-10	6-13
CRI	6-10	6-10	6-10	6-10	4-13	NOR	6-10	6-10	6-10	6-10	6-13
CYP	9-10	9-10	9-10	9-10	7-13	NZL	6-10	8-10	8-10	8-10	8-13
DEU	6-10	6-10	6-10	6-10	6-13	PAK	6-10	6-10	6-10	6-10	4-13
DNK	6-10	6-10	6-10	6-10	6-13	PAN	6-9	8-9	8-9	8-9	8-13
DOM	6-10	6-10	6-10	6-10	4-12	PER	6-10	6-10	6-10	6-10	4-13
DZA	8-10	8-10				PHL	6-10	6-10	6-10	6-10	4-13
ECU	7-10	7-10	7-10	7-10	4-13	PNG	9-10				
EGY	9-10	9-10	9-10	9-10	5-13	PRT	6-9	7-9	7-9	6-9	6-13
ESP	6-10	6-10	6-10	6-10	6-13	PRY	6-10	10-10	10-10	10-10	10-13
FIN	6-10	6-10	6-10	6-10	6-13	RWA	8-10	8-9	8-9	8-9	5-12
FRA	6-10	6-10	6-10	6-10	6-13	SDN	8-9	8-9	8-9	8-9	6-13
GBR	6-10	6-10	6-10	6-10	6-13	SEN	8-10	8-10	8-10	8-10	5-13
GHA	7-10					SLE	9-10	9-10	9-10	9-10	12-13
GMB	9-9	9-9	9-9	9-9	13-13	SLV	7-10	10-10	10-10	7-10	5-13
GRC	6-10	6-10	6-10	6-10	6-13	SWE	6-10	6-10	6-10	6-10	6-13
GTM	6-10	6-10	6-10	6-10	4-13	SYR	8-10	8-10	8-10		
GUY	9-9					TGO	8-10	8-10	8-10	8-10	5-13
HND	7-10	7-10	7-10	7-10	4-13	THA	7-10	7-10	7-10	7-10	5-13
HTI	6-9					TTO	9-10	9-10	9-10	9-10	4-12
IDN	10-10	10-10	10-10	10-10	8-13	URY	6-10	6-10	6-10	6-10	5-13
IND	6-10	6-10	6-10	6-10	4-13	USA	6-10	6-10	6-10	6-10	4-13
IRL	6-10	6-10	6-10	6-10	6-13	VEN	8-10	8-10	8-10	8-10	4-13
IRN	9-10	9-10	9-10	9-10	5-12	ZAF	7-10	7-10	7-10	7-10	5-13
ISR	6-10				11-13	ZWE	9-10				13-13
ITA	6-10	6-10	6-10	6-10	6-13	total	288	235	231	232	528

In order to summarise the country–5-year observations included in each sample, I rely on the following numbering scheme:

period	code
(1945,1950]	1
(1950,1955]	2
(1955,1960]	3
(1960,1965]	4
(1965,1970]	5
(1970,1975]	6
(1975,1980]	7
(1980,1985]	8
(1985,1990]	9
(1990,1995]	10
(1995,2000]	11
(2000,2005]	12
(2005,2010]	13

Table 1.7: Estimation results of the specification selected in Table 1.5 on the 1960–2010 period, modified with non-linear terms in the ratio of credit to ouput. Column (6) repeats the estimation results in Column (4) of Table 1.6. The estimation results in Column (7) use the first lag of initial income and school, the second lag of all credit terms in that specification, and the second lag of government, inflation, investment and population growth, as instruments for the differenced equations. In addition, the estimation results in Column (7) use the contemporaneous difference of initial income and school, the first difference of all credit terms in that specification, and the first difference of government, inflation, investment and population growth, as instruments for the levels equations. The estimation results in Column (8) use the analogous instrument set to the estimation results in Column (7), but replace all credit terms that appear in the specification in Column (7) by the credit terms that appear in the specification in Column (8). The estimation results in Column (5) use the analogous instrument set to the estimation results in Column (7), but replace all credit terms that appear in the specification in Column (7) by the credit terms that appear in the specification in Column (5). The instrument sets for the estimation results in Columns (1)–(4) repeat the instrument sets for the estimation results in Columns (5)–(8), but exclude terms in investment and population growth.

	(1) quad	(2) t2f3z	(3) 75	(4) 90	(5) quad	(6) inv pop	(7) 75	(8) 90
initial income	-0.41	-1.10	-0.56	-0.56	-0.93	-1.10	-0.68	-0.85
	(0.75)	(0.76)	(0.71)	(0.71)	(0.71)	(0.76)	(0.90)	(0.77)
school	0.50*	0.72**	0.43	0.54**	0.71**	0.98***	0.67**	0.78***
	(0.28)	(0.30)	(0.31)	(0.27)	(0.30)	(0.30)	(0.29)	(0.30)
credit-to-GDP	1.89				-1.14			
	(2.09)				(1.75)			
(credit-to-GDP)^2	-1.58*				-0.27			
	(0.83)				(0.80)			
log(credit-to-GDP)		0.54	1.48**	1.15		-0.55	-0.18	-0.07
		(0.59)	(0.68)	(0.71)		(0.55)	(0.64)	(0.52)
credit-to-GDP > 75%			19.68**				13.14	
			(9.20)				(9.00)	
(credit-to-GDP > 75%) × log(credit-to-GDP)			-4.69**				-3.01	
			(2.02)				(1.91)	
credit-to-GDP > 90%				20.41**				17.89*
				(9.77)				(9.69)
(credit-to-GDP > 90%) × log(credit-to-GDP)				-4.76**				-3.95**
				(2.08)				(2.00)
investment					3.75***	4.06***	3.29**	3.39***
					(1.15)	(1.24)	(1.33)	(1.13)
population growth					-0.38	-0.27	0.08	-0.05
					(2.18)	(2.15)	(2.30)	(2.22)
observations	528	528	528	528	528	528	528	528
countries	67	67	67	67	67	67	67	67
instruments	127	107	134	131	156	137	164	161
Hansen p-val.	1.000	0.994	1.000	1.000	1	1	1	1
AR(1) stat.	-4.424	-4.402	-4.411	-4.385	-4.475	-4.348	-4.391	-4.401
AR(2) stat.	0.310	0.0813	-0.383	-0.216	0.613	0.664	0.525	0.562
joint p-val.	0.0103		0.0430	0.0326	0.0397		0.0374	0.0510
turning point	0.598				-2.081			
turning point s.e.	0.384				9.128			

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.8: Estimation results for a specification that includes potential non-linear explanatory variables in the expenditure share of investment in output. The estimation results in Column (1) use the first lag of the level of initial income and school, the second lag of the level of all terms in investment, and the second lag of the level of credit, government, inflation and population growth, as instruments for the differenced equations. The estimation results in Column (1) also use the contemporaneous difference of initial income and school, the first lag of the difference of all terms in investment, and the first lag of the difference of credit, government, inflation and population growth, as instruments for the levels equations. The estimation results in Column (2) use the analogous instrument set to the estimation results in Column (1), but replace all terms in investment in the instrument set from Column (1) with such terms in investment from Column (2). The last row of this table gives the p-values associated with Wald tests where in each column the null hypothesis is that the coefficients on all the investment terms in that column, other than the log-linear investment term, are jointly equal to zero.

	(1) 21%	(2) 25%
initial income	-1.08	-1.07
	(0.79)	(0.77)
school	0.88***	0.90***
	(0.28)	(0.30)
credit	-0.50	-0.39
	(0.47)	(0.46)
log(investment-to-GDP)	3.34**	2.71*
	(1.58)	(1.59)
investment-to-GDP > 21%	1.90	
	(10.80)	
(investment-to-GDP > 21%) × log(investment-to-GDP)	-0.64	
	(3.36)	
investment-to-GDP > 25%		-3.86
		(13.49)
(investment-to-GDP > 25%) × log(investment-to-GDP)		1.26
		(4.13)
population growth	-1.01	-0.70
	(2.55)	(2.83)
observations	528	528
countries	67	67
instruments	170	169
Hansen p-val.	1	1
AR(1) stat.	-4.360	-4.345
AR(2) stat.	0.514	0.510
joint p-val.	0.968	0.897

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.9: Estimation results for a specification with the expenditure share of investment in output as the dependent variable. The instrument sets used to estimate the specifications in Columns (1), (3), (5) and (6) are lagged one period relative to the instrument sets used to estimate the specifications in Columns (2) and (4) because the differenced residuals display statistically significant second order serial correlation in Columns (1), (3), (5) and (6). More specifically, Column (1) uses the level of the second lag of initial income ($y_{i,t-3}$) and school, and the third lags of (log) credit, government and inflation, as instruments for the differenced equations. Column (1) also uses the difference of the first lag of initial income ($\Delta y_{i,t-2}$) and school, and the difference of the second lag of (log) credit, government and inflation, as instruments for the levels equations. Column (2) uses the level of the first lag of initial income ($y_{i,t-2}$) and school, and the second lags of (log) credit, government and inflation, as instruments for the differenced equations. Column (2) also uses the difference of contemporaneous initial income ($\Delta y_{i,t-1}$) and school, and the difference of the first lag of (log) credit, government and inflation, as instruments for the levels equations. Columns (3), (5) and (6) use the level of the second lag of initial income ($y_{i,t-3}$) and school, and the third lags of the credit variables, government and inflation, as instruments for the differenced equations. Columns (3), (5) and (6) also use the difference of the first lag of initial income ($\Delta y_{i,t-2}$) and school, and the difference of the second lags of the credit variables, government and inflation, as instruments for the levels equations. The estimation results in Column (4) use the level of the first lag of initial income ($y_{i,t-2}$) and school, and the level of the second lag of the credit terms, government and inflation, as instruments for the differenced equations. In addition, the estimation results in Column (4) use the contemporaneous difference of initial income ($\Delta y_{i,t-1}$) and school, and the first lag of the difference of the credit terms, government and inflation, as instruments for the levels equations. The last row of the table shows p-values for Wald tests where the null hypothesis is the exclusion of all explanatory variables in credit.

	(1) 60-10 log lin.	(2) 60-95 log lin.	(3) 60-10 quad.	(4) 60-95 quad.	(5) 60-10 75%	(6) 60-10 90%
initial income	0.10	0.09	0.18***	0.20**	0.08	0.10
	(0.07)	(0.09)	(0.06)	(0.08)	(0.06)	(0.07)
school	-0.04	-0.07*	-0.06**	-0.10***	-0.04*	-0.06**
	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.03)
log(credit-to-GDP)	0.11**	0.25***			0.15***	0.14***
	(0.05)	(0.06)			(0.06)	(0.05)
credit-to-GDP			0.19	0.77***		
			(0.22)	(0.28)		
(credit-to-GDP)^2			-0.08	-0.27*		
			(0.08)	(0.15)		
credit-to-GDP > 75%					0.54	
					(0.58)	
(credit-to-GDP > 75%) × log(credit-to-GDP)					-0.14	
					(0.12)	
credit-to-GDP > 90%						0.39
						(1.15)
(credit-to-GDP > 90%) × log(credit-to-GDP)						-0.11
						(0.22)
observations	528	339	528	339	528	528
countries	67	61	67	61	67	67
instruments	100	69	119	82	124	121
Hansen p-val.	0.995	0.653	1.000	0.954	1.000	1.000
AR(1) stat.	-0.849	-1.161	-0.668	-0.642	-1.102	-1.142
AR(2) stat.	-2.154	-1.193	-2.328	-0.903	-2.164	-2.214
joint p-val.			0.583	0.00881	0.0444	0.0351

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Chapter 2

Measuring nominal yield risk: Theory and an application to the loan books of US banks

I present a model for describing the dependence of financial intermediary net worth and net cashflow on the term to maturity structure of nominal interest rates. I show how the model nests the approaches in three strands of the literature. Such a model can be used to predict the change in net worth caused by a change in nominal interest rates. I show how to quantify the uncertainty in this change in net worth that arises from term structure parameter estimation and measurement error in accounting data. I apply the methods to measure the losses due to revaluations of the loan book at each of $10,759$ *US banks, and of the aggregate loan book of the banking system, between 2005 and 2010 in response to increases in the short term nominal interest rate. Point estimates of the sensitivity of the value of the loan book to changes in the term structure of nominal interest rates display some cyclicality, peaking just before the onset of the financial crisis and reaching their lowest levels at the end of the sample. Standard errors due to parameter estimation uncertainty can be substantial and therefore should not be ignored.*

Keywords: interest rate risk, term structure, systemic risk, net worth
JEL Classification: G21, G32

Short and long term nominal interest rates are highly variable over short and long time periods. Figure 2.1 plots the one-year zero-coupon nominal interest rate and the difference between the thirty-year and one-year zero-coupon nominal interest rates since 1985. Changes in short and long term nominal interest rates could reflect changes in the real interest rate or expected future inflation. These changes redistribute wealth between the parties to a contract in which future

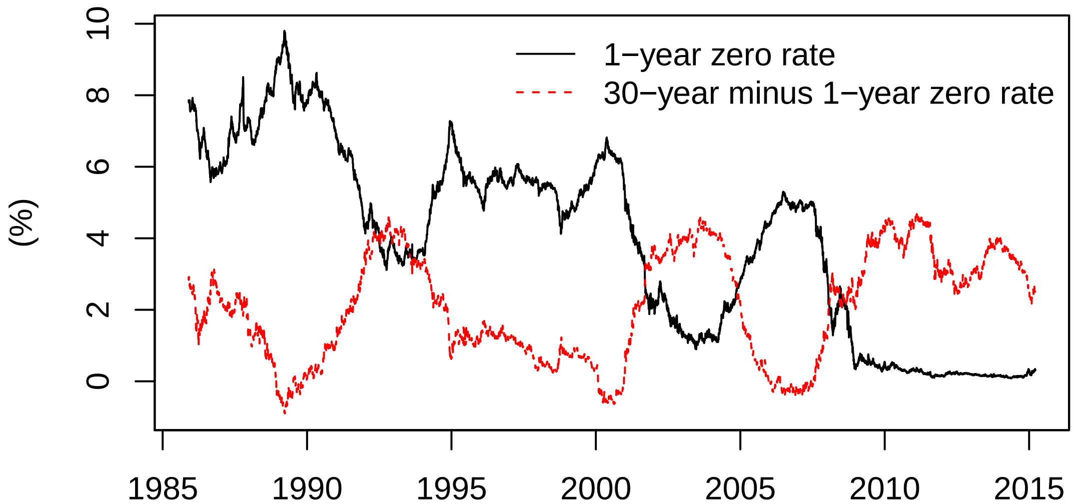


Figure 2.1: The solid black line draws the one-year zero-coupon bond nominal interest rate and the dashed red line draws the difference between the thirty-year and one-year zero-coupon bond nominal interest rates over time. I obtain data from the Federal Reserve H15 release for constant-maturity Treasury rates.

payments are fixed in nominal terms. Doepke and Schneider [2006] estimate the potential value of such transfers between old and young population cohorts and between foreign and domestic governments due to changes in the term structure of nominal interest rates and find that they can be large. For financial intermediaries in particular, a change in the term structure of nominal interest rates has implications for their solvency and future profitability.

A literature on the linkages between macroeconomics and finance, exemplified by Gertler et al. [2010], argues that the net worth of financial intermediaries amplifies aggregate fluctuations by determining the amount that financial intermediaries can borrow from households and hence the amount of credit they can supply to the productive sector. Describing the dependence of financial intermediary net worth on the term to maturity structure of nominal interest rates elucidates this particular aspect of these macroeconomic models.

Monetary policy has the power to influence short and long term interest rates through open market operations and commitments to future policy that affect agents' expectations for future inflation. Particularly relevant in this regard are episodes where central banks have actively managed the average maturity of their portfolios of securities with the intention of 'twisting' the shape of the term structure. In general, such actions can alter the net worth of borrowers and lenders and the terms of contracts between them, which in bad cases could result in some borrowers being unable to honour the terms of their contracts [Jiménez et al., 2009] or some banks contracting their credit supply [Jiménez et al., 2012]. Methods for measuring the effects of changes in the term structure of nominal interest rates on financial intermediary net worth are useful for determining the potential financial instability consequences of hypothetical monetary policy actions. At the time of writing, the potential risks from the so-called normalisation of monetary policy are being discussed.

Prudential regulation is partly concerned with measuring the exposure of financial intermediaries to changes in the term structure insofar as such changes could affect their safety and soundness. For example, it is well documented that high and unpredictable interest rates in the 1980s led to widespread failures of banks and thrift institutions in the US [Federal Deposit Insurance Corporation, 1994]. Today, the effects of changes in the term structure of nominal interest rates on the solvency and net worth of financial intermediaries might be expected to manifest themselves in restricted access to funding.

A related literature in empirical finance studies the role for the term structure of nominal interest rates in explaining the cross section of nominal returns on stocks, over and above the well known market return factor. Flannery and James [1984] find evidence that parallel increases in the term to maturity structure of nominal interest rates on government bonds lead to lower stock returns for banks. They also find that the stock returns of banks with longer average nominal contract maturity are more sensitive to interest rate changes, which they interpret as evidence that exposure to changes in the level of the term structure are priced into the stock returns of banks.[1]

There are several papers that study methods for measuring the effect of changes in nominal interest rates on the net worth of banks.[2] Houpt and Embersit [1991] propose a method for measuring the effect of changes in a scalar interest rate parameter on the net worth of a bank based on first-order approximations to the relationship between the prices of the nominal contracts it owns and this interest rate. This model has been developed further, confidentially, by the Federal Reserve System and has been referred to[3] as the Economic Value Model (EVM). A strength of the EVM is that it relies only on publicly available bank disclosures. However, the prices of contracts with embedded optionalities are known to depend non-linearly on nominal interest rates, so that first-order approximations to their sensitivity to nominal interest rates are likely to perform poorly. Wright and Houpt [1996] criticise the EVM for its performance in assessing the sensitivity of the prices of deposit contracts to changes in nominal interest rates. The prices of derivative contracts are also known to depend non-linearly on nominal interest rates. Sierra and Yeager [2004] restrict their sample to banks that do not engage in derivative transactions and find that banks the EVM identifies as most sensitive to rising rates show the greatest deterioration in solvency in the 1998-2000 period of rising rates in the US. Sierra [2009] confirms this finding and shows that the banks the EVM identifies as most sensitive to changes

[1]The average maturity of nominal contracts is perhaps the most commonly used summary statistic for the effect of changes in the level of the term structure of interest rates on the net worth of a financial intermediary. An introduction appears in Kaufman [1984].

[2]These methods apply equally to other types of financial intermediary, although in the tradition of the literature I refer to 'banks' for ease of exposition.

[3]For example, by Wright and Houpt [1996].

in nominal interest rates also experience the greatest changes in stock price at times when US nominal interest rates changed rapidly.

The Basel Committee on Banking Supervision [2004] provides a guidance document in which Annexes 3 and 4 provide specific recommendations for measuring the effect of changes in the term structure of nominal interest rates on the net worth of a financial intermediary. These annexes focus on measuring the effect of a parallel shift in the term structure of nominal interest rates. The proposal in Annex 4 follows a first-order-approximate method similar to Houpt and Embersit [1991] and the EVM.

The models described so far relied on a (possibly time-varying) scalar nominal interest rate parameter defining interest rates for all maturities, which is inconsistent with the strong variation in the term premium shown in Figure 2.1. Sher and Loiacono [2013] generalise the EVM and Basel Committee on Banking Supervision [2004] Annexes 3-4 guideline methods by describing the term structure of nominal interest rates with a functional form that allows for differences between nominal interest rates at different maturities and removing the reliance on first-order approximations. The authors apply their method to measure the sensitivity of the values of portfolios of loan contracts held by large European banks to changes in nominal interest rates.

In an original contribution, Drehmann et al. [2010] write down a model that describes the dependence of a bank's net worth on the term structure of nominal interest rates. The model is dynamic, which departs from those I discuss above and consequently allows the authors also to describe the dependence of a bank's cash flow on the term structure of nominal interest rates. The authors appeal to results by Duffie and Singleton [1999] to show that the model can also be used to capture the risk that borrowers default on their obligations to repay. Under technical conditions outlined in Duffie and Singleton [1999], the no-arbitrage price of a defaultable nominal contract depends on its likelihood and severity of default only through a linear spread component for credit riskiness, specific to that nominal contract, over the term structure for nominal interest rates associated with default-free securities. Drehmann et al. [2010] provide only an illustrative exposition of their model based on a bank that holds a single Treasury bond and an unspecified borrowing contract. The model is therefore not explicit enough for its properties to be analysed. The model also specifies that the investments of the bank change in value over time only as a result of changes in the likelihood or severity of defaults, which ignores the effects of changes over time in default-free nominal interest rates plotted in Figure 2.1. In reality, the price of a nominal contract evolves naturally over time as its residual maturity shortens even in the absence of changes in the term structure of nominal interest rates, which Drehmann et al. [2010] also ignore.[4] Drehmann et al. [2010] apply their method to measure the sensitivity of the

[4] Another minor difficulty with the model is that Drehmann et al. [2010] set it up in discrete time with time measured in quarters. They require that cashflows associated with nominal contracts occur at most at the quarterly frequency.

net worth of a fictitious bank to a change in the term structure of nominal interest rates.

Building on Doepke and Schneider [2006], Piazzesi and Schneider [2010] show how the evolution in price of a zero-coupon bond depends on the evolution in factors underlying an affine model of the term structure of nominal interest rates.[5] Assuming that porfolios of such zero-coupon bonds of different maturities can be used to replicate all the cashflows associated with the nominal contracts held by financial intermediaries, the net worth of a financial intermediary can be written as a linear combination of prices of zero-coupon bonds. Over a short time interval where no repayments are made on nominal contracts, the evolution in net worth of a financial intermediary can therefore be written as a linear combination of changes over time in the underlying affine factors describing the term structure of nominal interest rates. A disadvantage is that the change in net worth over a non-negligible time interval implied by the model excludes cashflows from nominal contract repayments and thus, without further assumptions, the model is not able to describe the net nominal cashflow over that time interval. In its current form, the model cannot therefore describe the dependence of net nominal cashflow on the term structure of nominal interest rates. Begenau et al. [2015] apply this method to public balance sheet disclosures of large US banks.

In this paper I offer an analytical model for describing the dependence of the net worth and net nominal cashflow of a financial intermediary on the term structure of nominal interest rates. The model is fully specified, allows for static and dynamic models of the term structure of nominal interest rates and describes both net worth and net cashflow consistently. In Section 2.1 I introduce the model and show how it nests the models of Drehmann et al. [2010], Piazzesi and Schneider [2010] and Sher and Loiacono [2013] as special cases.

The advantage of a fully specified model is that I can study its properties. The model can be used to predict the change in net worth that could be expected to occur, given a change in the term structure of nominal interest rates. In the special case of the model where we consider a counterfactual scenario for the term structure of nominal interest rates on a fixed date, I derive probabilistic and deterministic bounds for the change in net worth implied by the model. Probabilistic bounds could reflect uncertainty in the estimate of the change in net worth due to parameter estimation uncertainty underlying the term structure of nominal interest rates, or uncertainty in the public accounting disclosures on which the model is calibrated. Deterministic bounds reflect the range of changes in net worth that could be produced by the model given its physical form. For example, under models where the net worth of a financial intermediary is decreasing in the term to maturity structure of nominal interest rates, if the term to maturity structure is bounded above and below, then these bounds act as worst and best case scenarios for the net worth of the financial intermediary. I present such bounds in Section 2.2.

I apply the model to measure the changes in values of the loan books of US banks regulated by the Federal Deposit Insurance Corporation. I estimate the parameters associated with the

[5]Duffie and Kan [1996] introduce the affine model of the term structure of nominal interest rates.

term structure of nominal interest rates for the syndicated loan market each quarter between 2005 and 2010. I then use the model to measure the change in net worth that would result if short term nominal interest rates were to have been higher, without any change in long term nominal interest rates, on each valuation date. I pick a scenario specified by the Bank for International Settlements where the short term nominal interest rate increases 85% from its original level and the long term nominal interest rate is unaffected. The results for revaluation of the loan book with the values of other assets and liabilities held constant indicate that the average bank in an average time period loses 16% of equity under this scenario. The exposure for the average bank peaks in June 2007 before the onset of the financial crisis. I present the data on which the application is based in Section 2.3. I present the assumptions and results of the application in Section 2.4. Section 2.5 concludes.

2.1 Introducing the framework

For our purposes it does not matter whether we suppose that nominal interest rates are determined by an equilibium model or a no-arbitrage model. The following definition of a zero coupon bond nominal yield curve, or zero coupon nominal term structure of interest rates, is general enough to encompass yield curves that are flat (constant), of the exponential-polynomial family[6] or of the affine family [Duffie and Kan, 1996].

Definition 1 (zero curve)**.** *The term structure of zero coupon bond nominal interest rates is a function $z : \mathbb{R}_+ \to \mathbb{R}_+$ that maps maturities to zero rates.*

This defines a term to maturity structure of nominal interest rates on zero coupon bonds at a point in time. A family of such zero curves indexed by time is a useful extension, but it is not required for characterising the dependence of net worth on nominal interest rates at a point in time.

Example 1. *The Svensson [1994] zero curve can be written*

$$
\begin{aligned}
z(u) = \beta_0 + \beta_1 \frac{1 - e^{-u/\tau_0}}{u/\tau_0} &+ \beta_2 \left(\frac{1 - e^{-u/\tau_0}}{u/\tau_0} - e^{-u/\tau_0} \right) \\
&+ \beta_3 \left(\frac{1 - e^{-u/\tau_1}}{u/\tau_1} - e^{-u/\tau_1} \right),
\end{aligned}
\tag{2.1}
$$

where $\tau_0, \tau_1 > 0$.

Example 2. *Under a zero curve z, the price of a zero-coupon bond paying a nominal amount of 1 at time u is $e^{-uz(u)}$ and the value of nominal repayments of c per year payable continuously*

[6] Exponential-polynomial yield curves are a general class of curves that include important special cases like Nelson and Siegel [1987] and Svensson [1994].

for m years is $c\int_0^m \exp(-uz(u))du$. If $z(u) = y > 0$ for all u then the latter quantity is $c(1-e^{-my})/y$. In addition, if the price $p > 0$ of a loan contract can be represented as the value of a discounted stream of nominal cashflows of c per year payable continuously to the lender for m years, then the yield to maturity of this contract is the number y such that

$$p = c\frac{1-e^{-my}}{y} \tag{2.2}$$

which exists and is unique.

Assumption 1. *The net worth of a financial intermediary at any point in time can be represented as a sum of a countable sequence of values*

$$w = \sum_i v_i \quad \textit{for} \quad v_i \in \mathbb{R}. \tag{2.3}$$

We can interpret the index i as partitioning the assets and liabilities of a financial intermediary into representative contracts with values v_i for each i.

Example 3. *In addition to Assumption 1, let $w' = \sum_i v'_i$. Then*

$$w' - w = \sum_i v_i \frac{v'_i - v_i}{v_i} \mathbb{I}\{v_i \neq 0\} \tag{2.4}$$

decomposes the change in net worth $w' - w$ into initial allocations $v_1, v_2, \ldots$ and percentage changes $\frac{v'_i - v_i}{v_i}$.

Example 4. *In addition to Assumption 1, let $v_i = p_i q_i$ with $p_i > 0$ and let $w' = \sum_i p'_i q'_i$. Then $w' - w = \sum_i (p'_i q'_i - p_i q_i)$ and two equivalently-valued representations*

$$p'_i q'_i - p_i q_i = \begin{cases} (p'_i - p_i)q'_i + p_i(q'_i - q_i) \\ (p'_i - p_i)q_i + p'_i(q'_i - q_i) \end{cases} \tag{2.5}$$

decompose the change in net worth $w' - w$ into price and quantity changes separately. If $q'_i = q_i$ then the decompositions have the same form. Further, if we put $w = w_s > 0$, $p_i = p_i(\theta^i_s, z^i_s)$, $q_{is} = q_i = q_{it}$ and denote w', p'_i by substituting t for s then we get

$$\frac{w_t - w_s}{w_s} = \sum_i \frac{p_i(\theta^i_s, z^i_s)q_i}{w_s} \frac{p_i(\theta^i_t, z^i_t) - p_i(\theta^i_s, z^i_s)}{p_i(\theta^i_s, z^i_s)}, \tag{2.6}$$

which decomposes the percentage change in net worth between times s and $t > s$ into a leverage component, measuring the number of multiples of net worth invested in each contract i at initial time s and the percentage change in price of each contract over the interval $(s, t]$. If $q_i > 0$ for *each i then the leverage component simply measures the share of net worth invested in contract i.*

Definition 2. *(portfolio) A portfolio is defined by a nonnegative measure G in a measure space $(\Theta, \mathcal{T}, G)$, where Θ is a set and $\mathcal{T}$ is a collection of measurable subsets of Θ.*

Any $\theta \in \Theta$ could be a vector describing contract parameters like maturity, coupon rate and repayment frequency that determine the price of an individual contract. Describing a portfolio of contracts by a measure on contract parameters avoids arbitrary distinctions between portfolios of finitely many contracts and continua of contracts. It also allows a class of discounted cashflow models to be represented as integrals (to follow). An important special case is where $\Theta = \mathbb{R}$ and $\tilde{m} > 0$ is the residual time to maturity of a contract.

Assumption 2. *The value of the i^{th} collection of representative contracts can be written*

$$v_i = \int_{\mathbb{R}} p_i(\tilde{m}, z_i) dG_i(\tilde{m}) \tag{2.7}$$

for some functional p_i mapping $\tilde{m} \in \mathbb{R}$ and zero curve z_i to $\mathbb{R}_+$ and some measure G_i on subsets of $\mathbb{R}$ where $G_i(-\infty, 0] = 0$. Moreover, p_i is decreasing in z_i.

The functional p_i is then the price of a representative contract of type i with residual maturity $\tilde{m}$ at the prevailing zero curve z_i.[7] Dependence of p_i on contract parameters θ other than residual maturity $\tilde{m}$ is suppressed.

Definition 3. *If v_i satisfies (2.7) and there exists a finite, nonnegative measure H_i on subsets of $\mathbb{R}$ such that*

$$p_i(\tilde{m}, z_i) = \int_{\mathbb{R}} \exp(-u z_i(u)) dH_i(u) \tag{2.8}$$

then we say that the collection of contracts of type i can be represented as a portfolio H_i of zero coupon bonds.

If the price p_i of a contract of type i can be represented as the value of a portfolio H_i of zero coupon bonds according to (2.8) then the components of the right hand side of (2.8) constitute a discounted cashflow model for the price of the contract of type i.

One can define a measure of risk in many ways. The following definition is designed to be consistent with Heath et al. [1999] to capture a wide variety of such measures. The definition can at first appear abstract, so I provide specific examples that are of interest.

[7] The price p_i of each contract of type i depends on the zero curve z_i specific to contract type i. For example, the prices of Treasury securities may depend on the characteristics of their own zero-coupon bond curve, and the corporate bond zero curve may be higher than the Treasury zero curve to reflect a premium for default risk. The existence of a zero curve ensures that two contracts subject to the same zero curve with similar residual maturity exhibit similar prices.

Definition 4. *(risk) Given a starting zero curve z_i, the risk of the i^{th} collection of representative contracts can be written as a functional mapping the counterfactual zero curves z_i' to $\mathbb{R}$.*

Example 5. *Holding G_i constant, the (dollar) change in value of a portfolio satisfying* (2.7) *when the zero curve changes from z_i to z_i' is*

$$v_i' - v_i = \int_{\mathbb{R}} p_i(\tilde{m}, z_i') - p_i(\tilde{m}, z_i) dG_i(\tilde{m}), \tag{2.9}$$

the effective percentage change in value is $\frac{v_i'-v_i}{v_i}$ and the continuously compounded percentage change in value is $\log \frac{v_i'}{v_i}$. All three are risk measures according to Definition 4; in fact all three are monotonic transformations of one another. Combining (2.9) *with* (2.4)*, we get*

$$w' - w = \sum_{i:v_i \neq 0} v_i \left[\frac{\int_{\mathbb{R}} p_i(\tilde{m}, z_i') dG_i(\tilde{m})}{\int_{\mathbb{R}} p_i(\tilde{m}, z_i) dG_i(\tilde{m})} - 1 \right]. \tag{2.10}$$

In general, p_i depends on the parameters of the contract other than residual maturity. Nevertheless, some parameters of some contracts are not relevant for the riskiness of a contract. A fixed-rate amortising loan, for example, is characterised by a principal amount, coupon rate, repayment frequency, original contractual maturity and origination date. The origination date, original contractual maturity and current date determine the residual maturity of the contract. A portfolio of fixed-rate amortising loans would have to be specified as a measure on subsets of $\mathbb{R}^k$ where k is large enough to capture all the parameters necessary for valuation.

It turns out that not all parameters of a contract are necessary to determine its risk, in the sense of valuation sensitivity to nominal yield curve changes defined above. For example, it is easy to show that the percentage change in value of a single fixed-rate amortising loan does not depend on its coupon rate. The following result shows that the origination date of a fixed rate amortising loan can also be ignored if we know the contract's residual maturity, which simplifies the measurement and modelling of risk.

Lemma 1 (irrelevance of age of fixed-rate amortising loan)**.** *Let t_0 denote the issue date, m the original contractual maturity and c the repayment rate of a fixed-rate amortising loan contract. Let $t \in [t_0, t_0 + m)$ be the current date. Define the price of an individual fixed-rate amortising loan contract with these parameters by*

$$\int_{\mathbb{R}} \exp(-uz(u)) c \mathbb{I}\{0 \leq u \leq t_0 + m - t\} du. \tag{2.11}$$

Then the change in value and the percentage change in value of a portfolio of fixed-rate amortising loans, due to a change in the zero curve z, only depend on the underlying profile of issue dates (each t_0 in the portfolio) through its underlying profile of residual maturities (each t_0+m-t in the portfolio).

Proof. A portfolio of fixed-rate amortising loans defined by a finite measure ϕ on $\mathbb{R}^3$ has value

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}} \exp(-uz(u))c\mathbb{I}\{0 \leq u \leq t_0 + m - t\}dud\phi(t_0, m, c). \tag{2.12}$$

Since ϕ is finite (and Lebesgue measure is sigma-finite), we can reverse the integral signs and then partition the inner integral to get

$$\int_{\mathbb{R}} \exp(-uz(u))\mathbb{I}\{u \geq 0\} \int_{\mathbb{R}}\int_A c\mathbb{I}\{\tilde{m} \geq u\}d\phi(t_0, m, c)d\tilde{m}du. \tag{2.13}$$

where $A = \{(t_0, m, c) \in \mathbb{R}^3 : t_0 + m - t = \tilde{m}\}$. The indicator function can be taken out of the inner integral and the order of the outer two integrals can be reversed to get

$$\int_{\mathbb{R}^2} \exp(-uz(u))\mathbb{I}\{0 \leq u \leq \tilde{m}\}dug(\tilde{m})d\tilde{m} \tag{2.14}$$

which is the value of a portfolio described by measure of fixed-rate amortising loans with density function

$$g(\tilde{m}) = \int_A cd\phi(t_0, m, c) \tag{2.15}$$

with respect to Lebesgue measure and $\tilde{m} = t_0 + m - t$ denotes residual time to maturity. The portfolio ϕ can therefore be summarised by the density g without loss of generality. □

Lemma 1 says that we do not need to know both the residual time to maturity and the age, or original contractual maturity, of a fixed-rate amortising loan for the purposes of calculating its risk; only the residual maturity matters. This simplifies the measurement of risk because balance sheets often provide information on the residual times to maturity of fixed-rate amortising loans without any information on the ages, or original contractual maturities, of the loans. In general the value of a fixed-rate amortising loan at some future time will depend on the origination date and the coupon rate at which it was issued. However, the profile of residual maturities at a point in time is sufficient to characterise the sensitivity of the value of a fixed-rate amortising loan to changes in its underlying zero-coupon bond yield curve. This observation simplifies the modelling of risk by reducing the dimension of models of risk of fixed-rate amortising loans.

When projecting net worth over a time interval, we need to specify how intermediaries treat the proceeds from contracts that mature within that time interval. Drehmann et al. [2010] assume that the financial intermediary rolls over each maturing contract for a new contract of equivalent contractual maturity. The following condition formalises their assumption. It applies in the special case of the above framework for a portfolio of zero coupon bonds, the measure describing which is absolutely continuous with respect to Lebesgue measure and hence can be written in terms of a density function. The condition that follows is sufficient to show that the framework here generalises the risk measures proposed by Drehmann et al. [2010], Piazzesi and

Schneider [2010] and Sher and Loiacono [2013].

Assumption 3 (passive portfolio evolution)**.** *Let $a : \mathbb{R}^3 \to \mathbb{R}_+$ be a density function describing the number of zero coupon bonds in a portfolio at each $(\tilde{m}, m, t)$ where $\tilde{m}$ is residual maturity, m is original contractual maturity and t is the current date. Suppose for small $h > 0$ that*

$$a(\tilde{m}, m, t+h) = a(\tilde{m}+h, m, t) + \nu \mathbb{I}\{\tilde{m} = m\} \frac{a(h, m, t)}{p(m, t+h)} \tag{2.16}$$

where $p(m, t+h)$ denotes the price at time $t+h$ of a zero coupon bond with residual maturity m, $\nu \in [0, 1]$ and $a(0, m, t) \equiv 0$.

Assumption 3 has a natural interpretation. The left-hand side of (2.16) represents the 'number' of contracts with residual maturity $\tilde{m}$ and original contractual maturity m at time $t+h$. The first term on the right-hand side of (2.16) represents the number of contracts at (earlier) time t with (longer) residual maturity $\tilde{m}+h$ and original contractual maturity m. This term captures the behaviour that the residual maturity of any individual contract decreases with time (at rate 1 per unit of time) and the original contractual maturity of any contract does not vary with time. For $\nu = 1$, the second term on the right-hand side of (2.16) captures the behaviour that each immediately maturing zero coupon bond (those with $0 < \tilde{m} \leq h$ close to zero) is used to buy $p(m, t+h)^{-1}$ zero coupon bonds of maturity m, the original contractual maturity. For $\nu < 1$, a fraction $1-\nu$ of maturing contracts are lost from the model, for example by being paid immediately to the owners of the intermediary, and a fraction ν of maturing contracts are re-invested in new zero coupon bonds of their original contractual maturity.

Lemma 2. *Let net worth at time s be written $w_s = \sum_i v_{i,s}$ and suppose we have the representation*

$$v_{i,s} = \int_{\mathbb{R}^2} p(\tilde{m}, s) a(\tilde{m}, m, s) d\tilde{m} dm \tag{2.17}$$

for some density function a specific to i and satisfying (2.16)*. Then the change in net worth over the time interval $(s, t]$ can be written as $w_t - w_s = \sum_i (v_{i,t} - v_{i,s})$ where*

$$\begin{aligned} v_{i,t} - v_{i,s} = &\int_{\mathbb{R}} [p(\tilde{m}, t) - p(\tilde{m}, s)] \psi(\tilde{m}, s) d\tilde{m} + \\ &\int_{\mathbb{R}^2} \int_s^t \frac{\partial}{\partial \tilde{m}} a(\tilde{m}, m, \tau) d\tau p(\tilde{m}, t) d\tilde{m} dm + \\ \nu &\int_{\mathbb{R}} \int_s^t p(m, \tau)^{-1} \frac{\partial}{\partial \tilde{m}} a(0, m, \tau) d\tau p(m, t) dm \end{aligned} \tag{2.18}$$

where $\psi(\tilde{m}, s) \equiv \int_{\mathbb{R}} a(\tilde{m}, m, s) dm$ is the residual maturity profile specific to i at time s. Moreover, when we put $\nu \equiv 0$ this reduces to

$$v_{i,t} - v_{i,s} = \int_{\mathbb{R}^2} [p(\tilde{m} - t + s, t) - p(\tilde{m}, s) a(\tilde{m}, m, s) d\tilde{m} dm. \tag{2.19}$$

Proof. From (2.16),

$$\begin{aligned}\frac{1}{h}(a(\tilde{m}, m, t+h) - a(\tilde{m}, m, t)) \to \frac{\partial}{\partial \tilde{m}} a(\tilde{m}, m, t) + \\ \mathbb{I}\{\tilde{m} = m\} p(m, t)^{-1} \frac{\partial}{\partial \tilde{m}} a(0, m, t)\end{aligned} \tag{2.20}$$

as $h \to 0^+$. Hence for $s \leq t$,

$$\begin{aligned}a(\tilde{m}, m, t) = a(\tilde{m}, m, s) + \int_s^t \frac{\partial}{\partial \tilde{m}} a(\tilde{m}, m, \tau) d\tau + \\ \nu \mathbb{I}\{\tilde{m} = m\} \int_s^t p(m, \tau)^{-1} \frac{\partial}{\partial \tilde{m}} a(0, m, \tau) d\tau.\end{aligned} \tag{2.21}$$

We can use equation (2.21) to replace $a(\tilde{m}, m, t)$ in

$$v_{i,t} = \int_{\mathbb{R}^2} p(\tilde{m}, t) a(\tilde{m}, m, t) d\tilde{m} dm, \tag{2.22}$$

and recall the definition of $v_{i,s}$ to get the result in equation (2.18). In the special case where $\nu \equiv 0$ we can use $a(\tilde{m}, m, t+h) = a(\tilde{m} + h, m, t)$ from (2.16) to get

$$v_{i,t} = \int_{\mathbb{R}^2} a(\tilde{m} + t - s, m, s) p(\tilde{m}, t) d\tilde{m} dm. \tag{2.23}$$

Making a change of variable $\tilde{m} \mapsto \tilde{m} - t + s$ and subtracting $v_{i,s}$ gives the result in equation (2.19). □

The first term on the right-hand side of (2.18) gives the change in portfolio value if the yield curve were different on the valuation date, without any passage of time, which is the method used by Sher and Loiacono [2013].[8] The sum of the first two terms on the right-hand side of (2.18), which can be obtained from (2.18) by putting $\nu \equiv 0$, gives the change in portfolio value if time were allowed to reduce the residual maturities of the contracts in the time-s portfolio without producing any new contracts. In the case $\nu \equiv 0$, all maturing contracts are presumed to be lost (or paid out to the owners of the company) and therefore the portfolio value declines through time even if $p(\tilde{m}, t) = p(\tilde{m}, s)$ at each $\tilde{m}$. This is the method proposed by Piazzesi and Schneider [2010]. The third term on the right-hand side of (2.18) with $\nu \equiv 1$ measures the contribution to the change in value of portfolio i between dates s and t due purely to the change in the numbers of contracts at each maturity due to the renewal of contracts that mature over the interval (s, t). The whole right-hand side of (2.18) with $\nu \equiv 1$ gives the continuous portfolio analog of the discrete risk measure that is implicit in Drehmann et al. [2010].

[8] I refer the reader to the final expression in Section 3.2 of Sher and Loiacono [2013].

2.2 Bounds on risk

In this section I provide two sets of bounds for risk measures in the counterfactual scenario where the term structure of nominal interest rates is presumed to be different on the valuation date, which is the type of risk measure consistent with the first term on the right-hand side of (2.18). The first set of bounds provides standard errors due to uncertainty both in the estimated parameters underlying the term structure of nominal interest rates and in the accounting data on which the risk measure is based. These can be useful in testing whether risk has increased or decreased over time beyond what could be expected from typical sampling variation. The second set of bounds describes the range of values that any such risk measure in this model class can take.

In the following proposition, I consider the (percentage) change in net worth due to a change in the term structure of nominal interest rates from z_i to z_i' for each i. In the notation of Section 2.1, this change in net worth is

$$\frac{w' - w}{|w|^\delta} \tag{2.24}$$

where $\delta = 0$ gives the dollar nominal change and $\delta = 1$ gives the percentage change. I suppose that the scenario z_i' for each i is given exogenously, for example as part of a Monte Carlo exercise,[9] and the functional form of each z_i depends on parameters α. In turn, I assume α is consistently estimated by some $\hat{\alpha}_l$ with an asymptotic normal distribution in the sample size l.[10] I assume that the v_i characterising w through (2.3) are random, to accommodate sampling uncertainty in the accounting valuations. We could imagine a controlled experiment where the same nominal contract is given to the accounting departments at several financial intermediaries. Each accounting department determines a slightly different value for this contract, which gives a sense in which the accounting valuations are random. In the proposition below, I only specify a mean and variance for these accounting valuations because a full distribution would be difficult to determine in applications.[11] If the accounting valuations are thought to be unbiased for some 'true' valuation, then these 'true' valuations would give the means of the accounting valuations.[12] The variance of the accounting valuations describes with how much uncertainty the accounting valuations are determined. The proposition provides standard errors for the measure of change ($\delta = 0$) or percentage change ($\delta = 1$) in net worth due to parameter estimation uncertainty and

[9] Extension to the case where z_i' depends on z_i is straightforward.

[10] The proposition does not require an asymptotic normal distribution. It requires only that $\mathbb{E}(\hat{\alpha}_l - \alpha) \in o(1/\sqrt{l})$ and $\mathbb{V}\hat{\alpha}_l \in o(1/l)$. The asymptotic normality is included because it is likely to be satisfied in some applications, including Section 2.4. Similarly, the square-root rate convergence can be replaced by other rates.

[11] More precise statements would be possible if more information were known about the distributions of v_i. For example, if the full distributions of v_i were known, then it would be possible to characterise completely the asymptotic distribution of $(w' - w)/|w|^\delta$, although in general only numerically. It would then be possible to obtain explicit confidence intervals for $(w' - w)/|w|^\delta$ as opposed to just the first two moments of its distribution.

[12] For example, we might like to write the accounting valuations as $v_i = v_i^* + e_i$ where v_i^* is deterministic and e_i is a mean-zero random variable.

accounting valuation uncertainty jointly.[13]

Proposition 1. *Define*

$$w = \sum_{i=1}^{n} v_i \qquad v_i = \int_{\mathbb{R}} e^{-uz_i(u,\alpha)} dH_i(u) \tag{2.25}$$

$$w' = \sum_{i=1}^{n} v_i' \qquad v_i' = \int_{\mathbb{R}} e^{-uz_i'(u)} dH_i(u) \tag{2.26}$$

for $n \in \mathbb{N}$, random vector $v = (v_1, \ldots, v_n)^T$ and deterministic zero-coupon bond yield curves z, z' where z is allowed to depend on a vector of finitely many parameters α. Suppose that $\sqrt{l}(\hat{\alpha}_l - \alpha) \xrightarrow{d} N(0, \Sigma)$ as $l \to \infty$ and v is independent of $\hat{\alpha}_l$ for each l. Denote $\mathbb{V}\frac{v}{|w|^\delta} = \Omega_\delta$ and $\mathbb{E}\frac{v}{|w|^\delta} = \mu_\delta$ for $\delta \in \{0, 1\}$. Then

$$\mathbb{E}\frac{w' - w}{|w|^\delta} = f(\alpha)^T \mu_\delta + o(1/\sqrt{l}) \tag{2.27}$$

and

$$\begin{aligned} \mathbb{V}\frac{w' - w}{|w|^\delta} = f(\alpha)^T \Omega_\delta f(\alpha) + \operatorname{tr}\left(\frac{1}{l}\frac{\partial f}{\partial \alpha^T}\Sigma\frac{\partial f}{\partial \alpha}\Omega_\delta\right) + \\ \mu_\delta^T \frac{1}{l}\frac{\partial f}{\partial \alpha^T}\Sigma\frac{\partial f}{\partial \alpha}\mu_\delta + o(1/l) \end{aligned} \tag{2.28}$$

where the approximation improves as l increases and $f(\alpha) \equiv (f_1(\alpha), \ldots, f_n(\alpha))^T$,

$$f_i(\alpha) \equiv \frac{\int e^{-uz_i'(u)} d\tilde{H}_i(u)}{\int e^{-uz_i(u,\alpha)} d\tilde{H}_i(u)} - 1, \tag{2.29}$$

and $\tilde{H}_i(u) \equiv \left(\int dH_i(u)\right)^{-1} H_i(u)$.

Proof. Note that we can write

$$\begin{aligned} v_i' &= \int_{\mathbb{R}} e^{-uz_i'(u)} d\tilde{H}_i(u) \int_{\mathbb{R}} dH_i(u) && (2.30) \\ &= \frac{\int_{\mathbb{R}} e^{-uz_i'(u)} d\tilde{H}_i(u)}{\int_{\mathbb{R}} e^{-uz_i(u,\alpha)} d\tilde{H}_i(u)} v_i && (2.31) \end{aligned}$$

for $\tilde{H}_i$ as defined in the Proposition because H_i is a finite measure; $\tilde{H}_i$ is deterministic because

[13] Plugging zeros into one or another of the covariance matrices can be justified by taking limits as either covariance matrix approaches zero. In this way, standard errors in the presence of only parameter estimation uncertainty or only accounting valuation uncertainty could be obtained as special cases.

H_i in (2.25) only depends on v_i through $\int_{\mathbb{R}} dH_i(u)$. Therefore

$$\frac{w' - w}{|w|^\delta} = f(\hat{\alpha}_l)^T \frac{v}{|w|^\delta}. \tag{2.32}$$

Denote $f(\hat{\alpha}_l)$ by f. Then using the law of total variance and measurability of $|w|^\delta$ with respect to the information generated by v we have

$$\begin{aligned} \mathbb{V}\frac{w' - w}{|w|^\delta} &= \mathbb{V}f^T \cdot \frac{v}{|w|^\delta} &(2.33)\\ &= \mathbb{V}\left(\mathbb{E}(f^T|v)\frac{v}{|w|^\delta}\right) + \mathbb{E}\left(\frac{v^T}{|w|^\delta}\mathbb{V}(f^T|v)\frac{v}{|w|^\delta}\right) &(2.34)\end{aligned}$$

The conditional expectation and variance can be replaced by an unconditional expectation and variance respectively using the independence between v and $\hat{\alpha}_l$. Now use the delta method to deduce

$$\sqrt{l}(f(\hat{\alpha}_l) - f(\alpha)) \xrightarrow{d} N\left(0, \frac{\partial f}{\partial \alpha^T}\Sigma\frac{\partial f}{\partial \alpha}\right) \tag{2.35}$$

as $l \to \infty$ if the derivatives exist. In particular $\mathbb{E}f(\hat{\alpha}_l) = f(\alpha) + o(1/\sqrt{l})$, so the first term of (2.34) becomes

$$\mathbb{E}f^T\Omega_k\mathbb{E}f = f(\alpha)^T\Omega_\delta f(\alpha) + o(1/l). \tag{2.36}$$

Expand the quadratic form in the second term of (2.34) to get

$$\mathbb{E}\left(\frac{v^T}{|w|^\delta}(\mathbb{V}f)\frac{v}{|w|^\delta}\right) = \operatorname{tr}\left((\mathbb{V}f)\Omega_\delta\right) + \mu_\delta^T(\mathbb{V}f)\mu_\delta \tag{2.37}$$

and then use asymptotic property $\mathbb{V}f(\hat{\alpha}_l) = \frac{1}{l}\frac{\partial f}{\partial \alpha^T}\Sigma\frac{\partial f}{\partial \alpha} + o(1/l)$ to get the result.[14] $\square$

The following lemma, showing boundedness of the Svensson zero curve, is useful in the subsequent discussion on deterministic bounds for the risk measures of type consistent with the first term on the right-hand side of (2.18).[15]

Lemma 3. *Let z be the zero curve defined by* (2.1) *with $\tau_0, \tau_1 > 0$ and define*

$$\begin{aligned} \underline{z} &= \beta_0 + \beta_1\mathbb{I}\{\beta_1 < 0\} + (0.298426\ldots)(\beta_2\mathbb{I}\{\beta_2 < 0\} + \beta_3\mathbb{I}\{\beta_3 < 0\}) \\ \bar{z} &= \beta_0 + \beta_1\mathbb{I}\{\beta_1 > 0\} + (0.298426\ldots)(\beta_2\mathbb{I}\{\beta_2 > 0\} + \beta_3\mathbb{I}\{\beta_3 > 0\}). \end{aligned}$$

[14] Note that

$$\frac{1}{l}\left[\mathbb{V}\sqrt{l}(f(\hat{\alpha}_l) - f(\alpha)) - \frac{\partial f}{\partial \alpha^T}\Sigma\frac{\partial f}{\partial \alpha}\right] \in o(1/l).$$

[15] These bounds can also be used to show that $\partial f/\partial a$ exists in Proposition 1 when zero curves are described of the Svensson type.

Then $\inf_{u\geq 0} z(u) = \underline{z}$ *and* $\sup_{u\geq 0} z(u) = \bar{z}$. *In particular, this means that* $\underline{z} \leq z(u) \leq \bar{z}$ *for all* $u > 0$ *with each equality being achieved for some* u.

The Svensson zero curve is bounded above and below by finite constants that depend on its β parameters, but not its τ parameters. Note that the Svensson curve permits monotonically increasing zero curves as long as they are bounded above.[16]

Two important absolute bounds on risk measures are easy to derive. First, if the measure G admits to a density $g \in L^2$ with respect to Lebesgue measure, then by the Cauchy-Schwartz inequality

$$\left| \int_\Theta p(z', \theta) - p(z, \theta) dG(\theta) \right| \leq ||p(z', \theta) - p(z, \theta)||_2 ||g||_2 \tag{2.38}$$

for any zero-coupon yield curves z, z' where $|| \cdot ||_2$ denotes the L^2-norm over $\theta \in \Theta$. Second, if $z(u) \in [\underline{z}, \bar{z}]$ and $z'(u) \in [\underline{z}', \bar{z}']$ for two zero-coupon yield curves z, z' and some constants $\underline{z}, \bar{z}, \underline{z}', \bar{z}'$ and if $p(z, \theta)$ is decreasing in z,[17] then

$$\int_\Theta p(\bar{z}', \theta) - p(\underline{z}, \theta) dG(\theta) \leq \int_\Theta p(z', \theta) - p(z, \theta) dG(\theta) \leq \int_\Theta p(\underline{z}', \theta) - p(\bar{z}, \theta) dG(\theta). \tag{2.39}$$

In the case of the Svensson yield curve, the bounds in (2.39) can be easier to compute than the risk measure itself. The bounds in Lemma 3 are tight in the sense that they are actually reached by the Svensson zero curve. When the starting value $\int p(z, \theta) dG$ in (2.39) is known, for example if it is observed, then its bounds based on $\underline{z}, \bar{z}$ in (2.39) can be replaced by its known value. The risk bounds can be applied to a particular collection of representative contracts or to the aggregate solvency position of an intermediary.

2.3 Data

2.3.1 Loan yields

To match the balance sheet data in Section 2.3.2, I would ideally use yields on loans traded in the secondary market between banks that were originally negotiated between a relationship bank and a borrower. However, an organised secondary market for such loans does not exist. One of the few sources of data on the pricing of loan contracts are the loan yields at origination for syndicated loans, provided by Dealogic.

Syndicated loans are loans negotiated between a group of banks and an individual borrower. Each lending bank has a claim on the borrower, called a *tranche*, although there is a single loan agreement. Syndicated loans tend to be large, because the risks of the contract are spread among

[16] Boundedness from below is clearly an attractive feature of a description of the zero curve, because we would not want zero-coupon nominal interest rates to become arbitrarily negative at large maturities.

[17] $p(z, \theta)$ is a decreasing functional of z if for any zero-coupon yield curves z, z' we have $z(u) \geq z'(u) \forall u \Rightarrow p(z, \theta) \leq p(z', \theta)$.

banks in the group. Since the terms of syndicated loans are negotiated with the borrower, they are similar to the loans negotiated between a relationship bank and a borrower that appear on the balance sheets in Section 2.3.2. However, syndicated loans would be made to organisations rather than individuals and a secondary market also exists for syndicated loans, so that these contracts exhibit features similar to corporate bonds [Gadanecz, 2004].

I use nominal yields to maturity on fixed rate syndicated loans without any embedded optionality that were originated in the United States between January 2005 and June 2010 from Dealogic. The data provide maturity and yields to maturity on the day of origination of each tranche of each syndicated loan. I retain only those observations that have a (Fitch) credit rating of A- or higher, in order to focus on a sample of relatively homogeneously priced contracts, underlying which we could plausibly expect to lie a single zero curve.[18]

2.3.2 Bank balance sheets

I use data on individual bank holdings of loan contracts between 2005 and 2010, provided by the Federal Deposit Insurance Corporation (FDIC) through its quarterly Call Report filings. The FDIC defines loans as debt receivables whose terms are negotiated with the borrower at the date of issue [FDIC, 2014].

Loans in the Call Reports are valued according to the purpose for which they are held by the bank [FASB, 1993]. Loans that are intended to be sold and likely to be sold within one year are classified as *held for sale* [FASB, 2001, paragraph 30]. Loans in this category are valued in the Call Reports at the lower of *amortised cost* or *fair value* [FDIC, 2014]. Appendix 2.A defines amortised cost and fair value. Held for sale loans make up 2.3% of the value of all loans on the aggregate banking system balance sheet in the second quarter of 2010. Loans that the bank has the ability and intent to hold for the foreseeable future or until maturity are classified as *held for investment* [FASB, 2010, paragraph 6]. The reporting bank has the option[19] to report the value of any of these loans at fair value. Loans in this category that were previously transferred from the held for sale category are valued at the lower of amortised cost and fair value [FDIC, 2014]. Federal Deposit Insurance Corporation [2014] does not specify how other loans in this category are to be valued. Held for investment loans make up the remaining 97.7% of the value of all loans on the aggregate banking system balance sheet in the second quarter of 2010.

The largest proportion (59.7%) of the aggregate banking system loan book in the second quarter of 2010 was allocated to loans secured by real estate. This proportion includes loans to households and firms, but the split is difficult to determine without further assumptions.

[18] About half the observations do not have a credit rating. Of those that do have a credit rating, very few have credit ratings below A-. Within the observations that are rated A- or higher, most observations have a rating of AAA. Some experimentation suggests I do not have sufficiently many observations to narrow the sample based on credit rating further.

[19] Financial Accounting Standards Board [2007] explains this fair value option.

Table 2.1: For the aggregate banking system, the value v_j of all assets of type j relative to the value of equity w at four representative year-ends in the sample. The middle horizontal line separates assets from liabilities.

contract type i	1998	2002	2006	2010
cash & ff sold	0.8	1.0	1.1	1.1
securities & trading	3.2	3.1	2.7	2.6
loans	7.1	6.3	5.8	4.8
property	0.2	0.1	0.1	0.1
other assets	0.5	0.7	0.8	0.8
deposits & ff purchased	7.9	7.9	7.4	6.7
trading & notes	0.7	0.7	0.5	0.4
other liabilities	1.4	1.4	1.4	1.1

The next largest proportion, 19.7%, represents loans for commercial or industrial purposes that are not secured by real estate (and hence do not fall into the first category above). Consumer loans for credit cards and revolving credit plans make up 10.8% of the value of the loan book of the aggregate banking system in the second quarter of 2010.[20] Because syndicated loans are made to firms, governments and international organisations, they are more similar to the commercial loans in the Call Reports than loans to individuals. However, due to difficulty in obtaining maturity information separately for loans to individuals and firms in the Call Reports and pricing information for loans to inviduals, in the following application I treat the prices of syndicated loans as representative of the prices of all loans with maturity information on the balance sheets of banks in the Call Reports. This caveat should be borne in mind when interpreting the results of the application.

In Table 2.1 I present a summary of the balance sheet of the banking system, obtained by adding up the balance sheets of individual banks at every point in time.[21] Using the notation of Asssumption 1, each cell in Table 2.1 represents $|v_i|/w$ at a quarter-end and the values v_i/w in each column add up to $1 = \sum_i v_i/w$, apart from an accounting residual. Table 2.1 shows that loan contracts and deposit contracts are the largest items on the balance sheet. The average maturity of loan contracts can be large, while the sensitivity of the value of deposit contracts to changes in nominal interest rates is small [O'Brien, 2000], which motivates my focus of this application on the nominal interest rate sensitivity of the values of loan contracts.

The Call Report data do not separate the value of loans into values of fixed rate loans and floating rate loans. The nominal repayments owed by the borrower on a floating rate loan contract are contingent upon the level of nominal interest rates. Typically however, the repayments are contingent on past nominal interest rates rather than current interest rates so

[20]The remaining aggregate loan book in the second quarter of 2010 was allocated to other financial intermediaries, governments, farms and leases.

[21]The Call Report balance sheets are reported at the individual bank level, so that if two intermediaries are part of the same holding company there is no double-counting when adding their assets together.

Table 2.2: For the aggregate banking system, the percentage of the value of the loan book allocated to each maturity bucket at four representative year-ends in the sample.

maturity bucket k (years)	1998	2002	2006	2010
3/12 or less	43.3	43.1	45.2	45.8
between 3/12 and 1	14.6	10.0	8.1	8.2
between 1 and 3	13.2	14.5	13.7	15.0
between 3 and 5	11.1	11.6	11.1	9.6
between 5 and 15	11.8	12.0	11.7	10.3
over 15	6.1	8.8	10.3	11.0

that the nominal value of the next repayment is always known at the time of a repayment. The price of a floating rate loan resets to the outstanding face value of the loan at each repricing date. Therefore, on other dates the price of a floating rate loan behaves like the price of a fixed rate amortising loan maturing on its next repricing date.

Table 2.2 shows the maturity profile of loan contracts on the same four representative year-end dates as in Table 2.1. The Call Reports used to produce Table 2.2 provide values of loan contracts by time until next repricing date, rather than by residual maturity, which is the sense in which the maturities in Table 2.2 should be interpreted. In 2010, 45.8% of the value of the loan book represented contracts with a time until next repricing date of at most three months, which includes floating rate loans with time until next repricing date of at most three months and fixed rate loans with residual maturity of at most three months. Since their prices depend on nominal interest rates in the same way, we do not need to differentiate between the proportions of floating and fixed rate loans in this maturity bucket (or other maturity buckets). In 2010, a non-negligible 11% of the value of the aggregate banking system loan book was allocated to loans maturing in more than 15 years' time. These long maturities also motivate a consideration of the sensitivity of the values of loan contracts to changes in nominal interest rates.

2.4 Application

2.4.1 Estimating yield curves for loan contracts

I group the loan yields to maturity by month and estimate a single Svensson [1994] zero-coupon bond yield curve in each month. I restrict attention to the final month in each quarter beginning with March 2005 and ending with June 2010 to match the quarterly balance sheet data. Given n observations $\{(y_j, m_j) : j = 1, \ldots, n\}$ on loan yields to maturity and maturity, I find the Svensson parameters $\alpha = (\beta_0 \;\; \beta_1 \;\; \beta_2 \;\; \beta_3 \;\; \tau_0 \;\; \tau_1)^T$ and the nuisance parameter $\sigma^2 > 0$ that maximise the quasi-log-likelihood function[22]

$$l(\alpha,\sigma^2) = \sum_{j=1}^{n} l_j(\alpha,\sigma^2) \tag{2.40}$$

$$l_j(\alpha,\sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}e_j^2 \tag{2.41}$$

associated with the n pricing errors

$$e_j = \frac{1-e^{y_j m_j}}{y_j} - \int_0^{m_j} e^{-uz(u,\alpha)}du. \tag{2.42}$$

I find the Svensson parameters each month by applying an unconstrained Nelder–Mead optimiser to the transformed problem

$$\alpha \mapsto \begin{pmatrix} \sqrt{\beta_0} & \beta_1 & \beta_2 & \beta_3 & \log(\tau_0) & \log(\tau_1) \end{pmatrix}^T, \tag{2.43}$$

which constrains the long rate β_0 and the scale parameters τ_0, τ_1 to be positive.[23] This maximum likelihood estimator $\hat{\theta}$ for $\theta = \begin{pmatrix} \alpha & \sigma^2 \end{pmatrix}^T$ has the asymptotic distribution

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, A^{-1}BA^{-1}) \tag{2.44}$$

where for any j

$$A = -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\theta^T} l_j(\alpha,\sigma^2) \tag{2.45}$$

$$B = \mathbb{E}\frac{\partial}{\partial\alpha} l_j(\alpha,\sigma^2)\frac{\partial}{\partial\alpha^T} l_j(\alpha,\sigma^2). \tag{2.46}$$

The delta method can be used to find the (pointwise) asymptotic distribution of $z(u,\hat{\alpha})$ for each u, and $A^{-1}BA^{-1}$ can be consistently estimated using the sample analogues of A and B.[24] I follow this procedure to plot the zero-coupon yield curve for the second quarter-end of 2010 with its asymptotic 95% pointwise confidence intervals in Figure 2.2.

---

[22] Svensson [1994] and Gürkaynak et al. [2007] estimate the parameters α by maximum likelihood for US Treasury securities under the assumption that the errors e_j are normally distributed. Svensson [1994] does not clarify whether his standard errors are robust to heteroskedasticity and Gürkaynak et al. [2007] do not compute any standard errors. I am not aware of a paper that estimates a Svensson [1994] nominal term to maturity structure for interest rates underlying loan contracts.

[23] I choose the Nelder–Mead algorithm because it does not require gradient information and is therefore easier to implement. I limit the procedure to a maximum of 2000 iterations. Neither Svensson [1994] nor Gürkaynak et al. [2007] explain the search algorithm they employ to find the parameters α, so benchmarking is difficult.

[24] The sample analogues of A and B replace the expectation operator by the sample average. Further details on the quasi-log-likelihood asymptotic distribution and covariance estimator used here can be found in White [1982]. I implicitly assume some regularity conditions to appeal to the results in that paper.

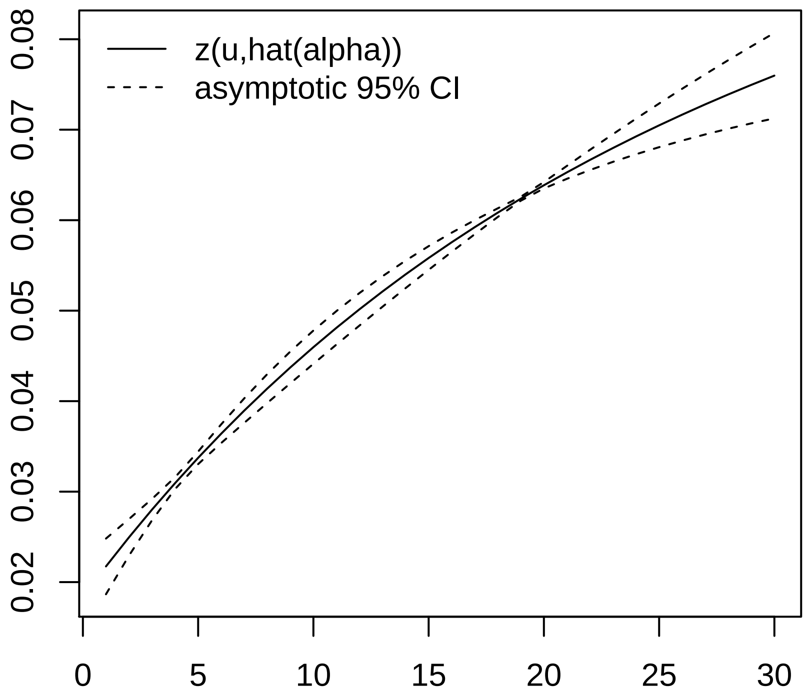


Figure 2.2: The solid line shows the zero coupon bond yield curve $z(u, \hat{\alpha})$ plotted against maturity u in years according to the parameter estimates $\hat{\alpha}$ for the second quarter of 2010. On the vertical axis, the number 0.05 means 5%. The dashed lines show 95% asymptotic pointwise confidence bands based on the delta method and heteroskedasticity-robust covariance matrix as described in the text.

2.4.2 Specifying an illustrative yield curve shock scenario

I follow the Bank for International Settlements [2015] Section II.2.2 in specifying an illustrative interest rate risk scenario. In particular, I focus on their scenario which specifies an 85% increase in the short rate with no change in the long rate, using the map for Svensson parameters defined by $\beta_1 \mapsto \beta_1 + 0.85|\beta_0 + \beta_1|$, where no other Svensson parameters change. Note that this scenario would not be well defined if a flat zero-coupon bond yield curve were used (i.e. if $\beta_1 = \beta_2 = \beta_3 = 0$).

I apply the above scenario map to each of the monthly Svensson parameter estimates to obtain 'shocked' yield curves. I present the estimated and 'shocked' zero-coupon bond yield curve for the second quarter of 2010 in Figure 2.3.

2.4.3 Individual bank exposure to loan contract revaluation

There are at least two difficulties in applying the methods above to the accounting data on loan contracts described in Section 2.3.2. First, for those contracts that are valued at amortised cost, we need to know the age of each contract, the yields prevailing at the date the contract was issued and the particular amortisation schedule (whether effective interest or another method)

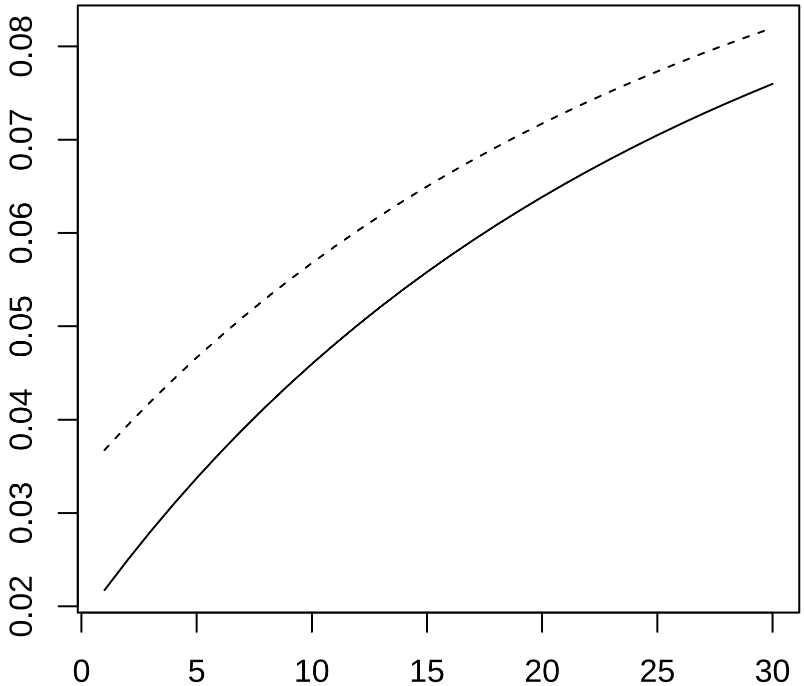


Figure 2.3: Scenario for shift in zero-coupon bond yield curves in the second quarter of 2010. The solid line shows the estimated zero-coupon bond yield curve $z(u, \hat{\alpha})$ against maturity u in years for the second quarter of 2010 and the dashed line shows the 'shocked' zero-coupon bond yield curve scenario. On the vertical axis, the number 0.05 *means* 5%.

chosen by the bank. Second, in the available for sale category, we need to distinguish between contracts that are valued at fair value and those that are valued at amortised cost. The balance sheets do not provide answers to either of these questions. For the purposes of an accurate risk assessment, one could introduce assumptions and investigate the robustness of conclusions to these assumptions. The most important special case of such assumptions is the case where the accounting valuations coincide with fair value. I work under this assumption for the purposes of illustrating the methods above.

In the notation of Section 2.1, we have the equity of the bank w is the sum of the values of finitely many representative assets and liabilities $w = v_1 + \cdots + v_n$. I restrict attention to the interest rate risk of the loan book only, which is defined by six maturity buckets with endpoints $(m_{k-1}, m_k]$.[25] I define the value of the loans in the k^{th} maturity bucket to be

$$v_k = \int_{m_{k-1}}^{m_k} p(\tilde{m}, z) dG_k(\tilde{m}) \tag{2.47}$$

where

$$p(\tilde{m}, z) = \int_0^{\tilde{m}} \exp(-uz(u, \alpha)) du \tag{2.48}$$

[25]That is, $m_k \in \{0, 1/4, 1, 3, 5, 15, \infty\}$.

Table 2.3: Revaluation of loan contracts by maturity bucket in the second quarter of 2010.

	0-0.25	0.25-1	1-3	3-5	5-15	15-Inf
old price	0.12	0.62	1.95	3.79	8.18	12.50
new price	0.12	0.62	1.92	3.68	7.72	11.52
price adjustment factor	1.00	0.99	0.98	0.97	0.94	0.92

and the measure G_k is absolutely continuous with respect to Lebesgue measure with density

$$g_k(\tilde{m}) = \begin{cases} c_k(m_k - m_{k-1})^{-1} & \text{if } m_k < \infty \\ (c_k/5)\exp(-(\tilde{m} - m_{k-1})/5) & \text{if } m_k = \infty \end{cases} \tag{2.49}$$

where c_k is a normalisation constant chosen to satisfy (2.47). In Table 2.3 I present the revaluation of each representative loan–maturity bucket in response to the yield curve shock specified in Section 2.4.2 above in the second quarter of 2010. The revaluation applies to $\frac{v_k}{c_k}$ from (2.47), i.e. per dollar of holdings in each loan–maturity bucket. The revaluation corresponds to the first term in (2.18), which holds the portfolio quantities G_k constant and only adjusts the zero-coupon yield curve z at a point in time. Table 2.3 shows that the value of loans maturing within three months is, to two decimal places, not sensitive to the change in zero curves drawn in Figure 2.3, while the value of loans with more than 15 years' residual maturity falls in value by 8%. This pattern of increasing revaluation sensitivity with maturity occurs despite the fact that the scenario increases short-term nominal interest rates $z(0+)$ while leaving long-term nominal interest rates $z(\infty)$ unaffected. In Figure 2.3 we see that the zero-coupon bond nominal interest rate at thirty years' maturity shifts appreciably under this scenario, i.e. $z'(30) \gg z(30)$, even though $z'(\infty) = z(\infty)$, because the estimated rates $1/\tau_0, 1/\tau_1$ at which the short-term nominal interest rate $z(0+)$ transitions to the long-term nominal interest rate $z(\infty)$ are small.

I then apply the representative loan constract revaluations from Table 2.3 to the balance sheets described in Section 2.3.2. Figure 2.4 shows the resulting losses for all banks in the second quarter of 2010. In Figure 2.4, each black circle gives

$$\sum_{k=1}^{6} \frac{v_k' - v_k}{w} \tag{2.50}$$

for an individual bank on the vertical axis. For the purposes of this application, I consider only the values of loan contracts and therefore the expression for w in (2.3) implies that w should be interpreted here as the total value of loan contracts rather than net worth. The black circles in Figure 2.4 are the result of applying the per-dollar revaluations in Table 2.3 to the loan books of each individual bank in the second quarter of 2010, as a fraction of the value of the bank's total loan book. To each black circle in Figure 2.4 there is a pair of grey circles equidistant above and below the black circle that gives two-standard deviation bounds computed according to formula (2.28), which represent uncertainty due to α-estimation and accounting measurement error. The heteroskedasticity-robust covariance matrix estimator from the log-likelihood maximisation problem in Section 2.4.1 gives a consistent estimate for the covariance matrix of $\hat{\alpha}$ in formula (2.28). I calibrate the measurement error Ω_1 in formula (2.28) to the data presented in Chart 1 of Haldane [2011].[26] Chart 1 of Haldane [2011] shows a range of possible valuations of approximately 10% around the 'true' value. By setting the distribution of $\log(v_k)$ to be normal with mean $\log(\mathbb{E}v_k) - \frac{\sigma^2}{2}$ and variance $\sigma^2 = \log(1.1)/2.58$, we ensure that 99% of observations on v_k lie no more than 10% above the median of v_k. An independence assumption on the v_k motivates setting

$$\Omega_0 = \operatorname{diag}\left((e^{\sigma^2} - 1)(\mathbb{E}v_k)^2\right) \tag{2.51}$$

and due to lack of analytical tractability I set $\Omega_1 = \left(\frac{1}{\mathbb{E}w}\right)^2 \Omega_0$.

Three interesting conclusions can be drawn from Figure 2.4. First, the range of the vertical axis shows that some banks' loan books devalue by as much as 7.5% of their original value as a result of this scenario, in which the short rate increases by 85% of its original level and long rates are unaffected. To illustrate the magnitude of such a loss, consider that the median bank at the end of the sample period has a ratio of assets to equity (leverage) of 10 and allocates 70% of the value of its assets on its balance sheet to loan assets. Holding the values of other assets and liabilities constant, a fall in the value of the loan book by 7.5% for such a bank would result in a $7.5\% \times 10 \times 70\% = 52.5\%$ deterioration in the solvency position.[27] These losses are therefore large enough to pose the question of whether loan book revaluations caused by movements in interest rates could lead to funding liquidity difficulties for the riskiest banks, which in turn could cause contractions in credit supply to the real economy.

A second feature of Figure 2.4 is the strong relationship across banks between the size of the loss and the weighted average maturity of the loan book. If the goal of a regulator is to rank institutions by their riskiness, for example to focus on the riskiest institutions, then the weighted average maturity will be a highly informative summary statistic. However, the weighted average maturity measure fails to identify the riskiest institutions to the extent that (i) their assets and liabilities are priced with substantial convexity with respect to the level of the underlying zero curves, (ii) the relevant zero curve adjustment scenarios are large and (iii) the relevant zero curve adjustment scenarios are not parallel. This conclusion is very similar to the conclusion obtained by Sher and Loiacono [2013] in the Eurozone context.

[26] The same figure appears as Chart 3.26 in Bank of England [2011]. For further details and a reproduction of this chart, see Appendix 2.C.

[27] This illustrative calculation abstracts from the effects of revaluations of other assets, liabilities and off-balance sheet contracts that would also impact the solvency position. The application in this section could be extended, entirely within the theoretical framework of Section 2.1, to accommodate these important considerations and in principle arrive at an overall impact on the solvency position.

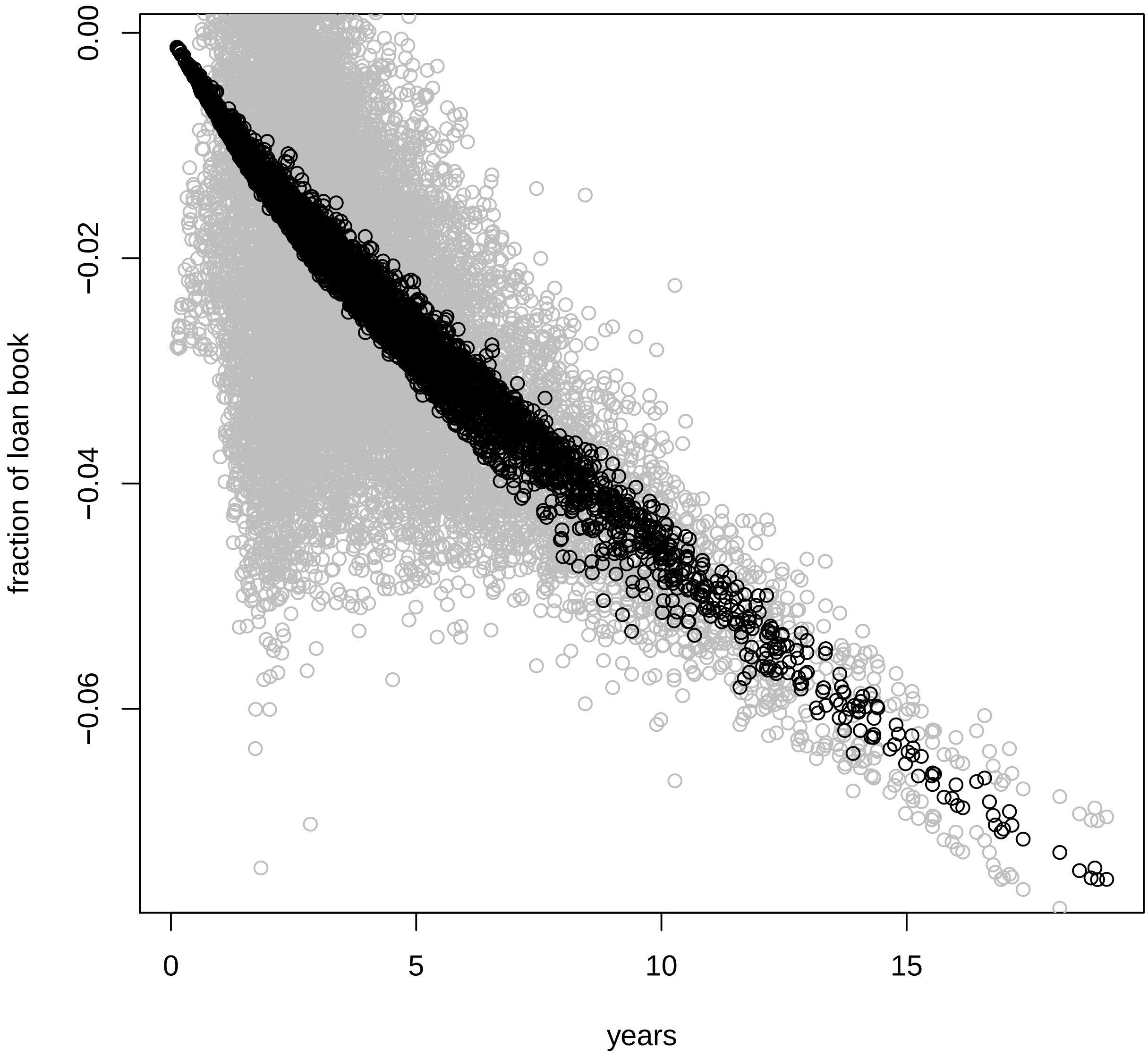


Figure 2.4: Revaluation of loan book in the second quarter of 2010 for each invididual bank as a fraction of the total value of its loan book, with negative values indicating losses (vertical axis) against weighted average maturity of the loan book in years (horizontal axis) are represented by the open black circles. Each black circle is associated with a pair of grey circles directly above and below it that indicate (pointwise) two-standard error bounds for the losses as percentages of the total value of the loan book.

A third and related conclusion to draw from Figure 2.4 is that the standard errors around such estimates of risk are substantial. In fact, the standard errors that capture parameter estimation and accouting valuation error are wide enough to prevent effective discrimination between the risk measures of different institutions. To allude to statistical hypothesis testing, how do we know that one institution is riskier than another if sampling variation alone could account for differences between their risk estimates? Figure 2.4 applies to the second quarter of 2010, for which the tight confidence bands in Figure 2.2 show that the Svensson parameters of the zero curve for loan contracts are estimated relatively precisely. In other quarters where the zero curve parameters are estimated less precisely, the standard errors associated with the risk estimates in Figure 2.4 are wider. Figure 2.4 also relies on parameters for accounting valuation uncertainty based on survey results for "high grade" assets, which are lower than the parameters for accounting valuation uncertainty based on survey results for other asset categories present in Chart 1 of Haldane [2011]. For these reasons, even the relatively wide standard error bounds in Figure 2.4 could understate the actual extent of estimation uncertainty in these risk measures.

2.4.4 Exposure of system of banks to loan contract revaluation

A particular motivation for studying the exposure of financial intermediaries' net worth to changes in the term structure of interest rates is the potential for fluctuations in such net worth to affect the macroeconomy through the supply of credit. At every date, by aggregating the values of assets and liabilities across banks I obtain a time series of balance sheets, including loan asset values by maturity bucket. The framework of Section 2.1 is equally applicable to this fictitious aggregate bank.

The estimated zero curves for loan contracts (Section 2.4.1) and the revaluations (Table 2.3) in response to the scenario of the increase in the short term interest rate (Section 2.4.2) are unchanged in this context. The form of the density g_k of G_k given in (2.49) and the heteroskedasticity-robust covariance matrix estimator of the asymptotic covariance matrix $A^{-1}BA^{-1}$ given in (2.45) and (2.46) are unchanged. The point estimates of the losses due to the increase in short term interest rates on loan contracts therefore follow immediately.

For the standard errors, I need to calibrate the accounting measurement error matrix Ω_1 in Proposition 1. The choice of this covariance matrix depends on whether I aggregate across banks by summation or by averaging. I choose to average the values of loan holdings across banks, which produces a time series of balance sheets for a representative 'average' bank with the convenient property that the covariance matrix of accounting valuations is order $1/N$ where N is the number of banks on a given date. Under the additional assumptions that the accounting valuations are uncorrelated across banks and lognormally distributed as in Section 2.4.3, I obtain

$$\Omega_1 = (e^{\sigma^2} - 1)\mathrm{diag}\left[\frac{1}{N}\sum_{i=1}^{N}(\mathbb{E}v_{ik})^2/(\mathbb{E}w_i)^2\right] \tag{2.52}$$

where v_{ik} denotes the value of loan contracts held by bank $i = 1, \ldots, N$ in maturity bucket $k = 1, \ldots, 6$.[28] Uncorrelated measurement errors in the accounting valuations across banks can therefore be made very small by averaging over all banks in each quarter. With 8000 banks in

[28] I have again used the approximation $\mathbb{V}v_i/w_i \approx (\mathbb{E}w_i)^{-2}\mathbb{V}v_i$ for analytical tractability, where $v_i = (v_{i1} \cdots v_{i6})$.

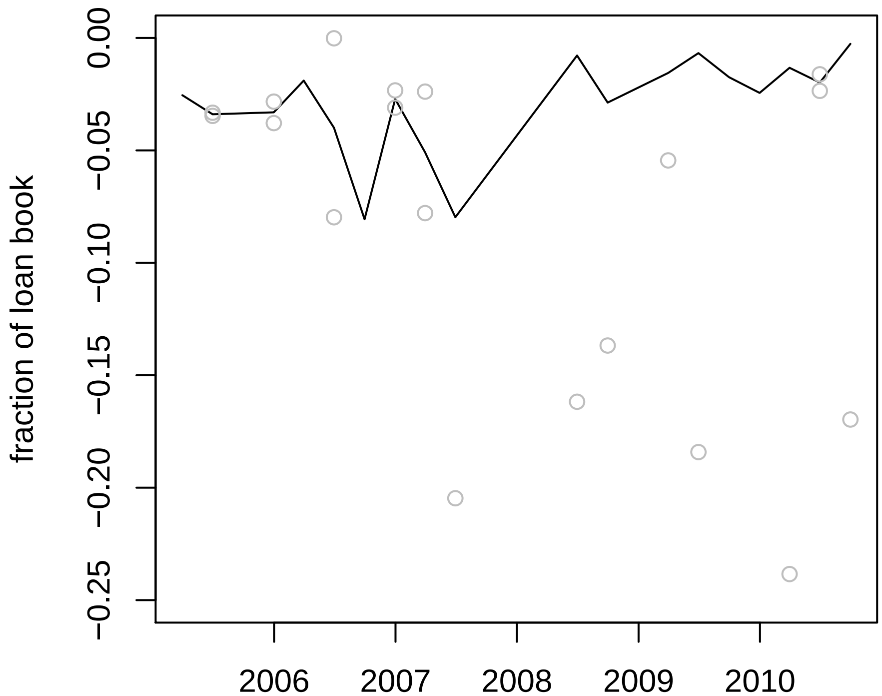


Figure 2.5: Losses for a fictitious average bank as a fraction of the value of its loan book against valuation date, drawn as a solid black line over time. Grey circles indicate two-standard error bounds specific to each date.

a typical quarter, we can expect the remaining measurement error to be negligible, in which case the standard error bounds have a natural interpretation as 95% confidence intervals from an asymptotic normal distribution.

I plot the resulting changes in value for the representative ‘average’ bank over time as the solid black line in Figure 2.5. The median loss over time in that figure is 2.5% of the loan book, which the same illustrative calculation as in Section 2.4.3 shows is large relative to the median equity in the sample.

The point estimates for riskiness in Figure 2.5 show some evidence for cyclicality. The increase in riskiness between 2005 and 2007 is particularly encouraging because the risk measure somewhat anticipates the financial crisis. Time variation in the risk measure reflects time variation in the shape and level of the zero curve,[29] the absolute magnitude and rate of decay with maturity of the zero curve scenario, and the maturity profile of the average bank’s loan book.

Two-standard error confidence intervals for each date appear as grey circles in Figure 2.5. These intervals are volatile over time, reflecting the numbers of observations on loan yields to maturity and the fit of the Svensson yield curve model to observed prices at each date, and some intervals do not appear on the plot because they fall outside the range of the vertical

[29] Higher zero curves generally imply a lower sensitivity to the same absolute parallel increase in the zero curve.

axis. By comparing the point estimate of risk at any date with the confidence interval at its preceding date, we can see that increases or decreases in the level of aggregate risk are difficult to distinguish from sampling uncertainty due to the estimation of yield curve parameters.

2.5 Conclusion

Gürkaynak et al. [2005] argue that macroeconomic models often abstract from time variation in agents' expectations for future inflation despite the subtantial movements in long term nominal interest rates over time. In Section 2.1 I show how to write down a simple framework that flexibly describes the relationship between the term to maturity structure of nominal interest rates and financial intermediary net worth. The framework might prove to be more flexible and general than currently necessary in general equilibrium models, but simplifications can easily be made. For example, β_2 and β_3 could be set to zero in the Svensson model (2.1) to obtain a monotonic zero curve, or the portfolio H_i in (2.7) could be set to $H_i(u) = h_1\mathbb{I}\{m_1 \leq u < m_2\} + h_2\mathbb{I}\{m_2 \leq u\}$ to describe a situation where the intermediary holds quantity h_1 of type i contracts with maturity m_1, quantity h_2 of type i contracts with maturity $m_2 > m_1$ and zero contracts of type i with any other maturity.[30] These simplified versions might be useful in current general equilibrium models.

The framework is as specific as possible while nesting the models of Piazzesi and Schneider [2010], Drehmann et al. [2010] and Sher and Loiacono [2013]. Combining these models facilitates their comparison. Making their definitions and assumptions explicit allows their properties to be studied.

One of the useful features of this framework is that it explicitly predicts the change in net worth caused by a change in the term structure of nominal interest rates. The model allows for counterfactual scenarios where net worth under two different term to maturity structures of nominal interest rates can be compared on the same valuation date, as studied by Sher and Loiacono [2013], or where net worth evolves with nominal interest rates between two dates, as studied by Drehmann et al. [2010]. As a first step toward obtaining properties of this class of models, I find the standard errors that are implicit in the first of these types of change in net worth. These standard errors reflect the parameter estimation uncertainty involved in estimating the term to maturity structure of nominal interest rates and the valuation uncertainty in the accounting data. The uncertainty in such changes in net worth has so far been ignored in the literature.

I illustrate the application of the framework by applying the Sher and Loiacono [2013] formula for instantaneous changes in net worth, along with the formula (2.28) for standard errors, to predict the change in value of the loan books that would occur if short-term nominal

[30] If contracts of type i are assets, then this form for H_i describes a 'two-asset' situation where the intermediary holds quantities h_1 and h_2 of one short-term and one long-term asset respectively.

interest rates on loan contracts were to increase without any change in such long-term rates, which is one of the scenarios given in Bank for International Settlements [2015].

The application reveals that losses under this scenario for nominal interest rates can be large (as large as about 52.5% of equity for some banks in June 2010). These large changes motivate further empirical work to find whether such losses could result in contractions in credit supply or stock price losses. The standard errors are also large, which motivates incorporating them into future work. In the cross section of banks at a point in time, weighted average maturity of the loan book and the ratio of the value of the loan book to the value of equity ('leverage') are important determinants of this measure of riskiness but are not sufficient for identifying the riskiest banks.

By averaging the loan asset holdings across banks, we almost eliminate any accounting valuation uncertainty and hence reduce the standard errors associated with the change in net worth. Nevertheless, parameter estimation uncertainty from term structure estimation remains substantial. For the aggregate banking system, the effect of changes in the term structure of nominal interest rates on the value of the loan book is somewhat cyclical and rises before the recent financial crisis.

This framework provided here is a first step toward a methodology for measuring the riskiness of financial intermediaries. I expect the framework here to provide sufficient flexibility and generality for most purposes, but the properties of the framework should be studied more extensively in future work. For example, for financial intermediaries like banks that tend to borrow short term and lend long term, prudential regulators are aware that increases in nominal interest rates tend to harm their net worth immediately and then benefit their future net cashflow, because new loans are issued at higher interest rates. It would be useful to establish this property in the framework of this paper.

The application provided here focuses on the effect of changes in nominal interest rates on the values of the loan books of US banks. It would be useful to extend this application to other assets and liabilities and thereby measure the effect of changes in nominal interest rates on the net worth of US banks. The main difficulty I see in doing so is the proper specification of a joint scenario for shifts in the nominal term structures associated with different contract types. Bank for International Settlements [2015] provides guidance on scenario construction for term structures of nominal interest rates associated with an individual contract type, but not for the co-movement of several term structures. A related difficult challenge is to specify consistent scenarios for changes in the nominal prices of assets like stocks or property that are not claims on fixed nominal cashflows.

The changes in net worth predicted by such a fuller application of the framework in this paper would be useful for at least three important questions in economics. First, such changes could be used in an empirical assessment of the effects of monetary policy on the credit supply and risk-taking behaviour of banks. In the spirit of Kashyap and Stein [2000], if it turns out that banks with greater sensitivity of net worth to changes in nominal interest rates experience slower growth rates of their loan books following monetary policy tightenings, then this would provide evidence for an effect of monetary policy on credit supply (as opposed to credit demand, which we would not expect to vary across banks with the sensitivity of their net worth to changes in nominal interest rates). Second, predictions of changes in net worth due to changes in nominal interest rates could be used to assess whether this source of risk is priced into these banks' publicly traded stocks. In the spirit of Flannery and James [1984], we can regard changes in nominal interest rates as an important factor determining the cross section of banks' stock returns if we can find evidence that the time series sensitivities of banks' stock returns to changes in interest rates are positively associated in the cross section with the changes in net worth predicted by the theoretical framework in this paper. Third, measures of changes in net worth of the aggregate banking system due to changes in nominal interest rates produced by the theoretical framework in this paper could be used as an input in an exercise attempting to predict aggregate financial instability.

2.A Amortised cost and fair value

Amortised cost is defined by the FASB Glossary as a combination of the value at which an asset is initially acquired and some amortisation adjustment to reduce the value of the asset to zero over its contractual life. Amortisations for loan contracts should be calculated using the *net interest method* [FASB, 1986, paragraph 18]. The net interest method values the contract by discounting its future repayments at the yield to maturity at which that contract was first issued [FASB, 1985, paragraph 235].[31] Therefore, to the extent that interest rates on the date that a bank prepares its Call Report are similar to the interest rates prevailing at a date on which one of its loan contracts was issued, the fair value of that loan contract would be similar to the amortised cost value of that loan contract.

The fair value of a loan asset is an estimate of the amount that would be realised by selling the asset in an orderly transaction between market participants. Financial Accounting Standards Board [2004] specifies that fair value should be based on observable liquid market prices for the specific contract if possible, otherwise on adjusted market prices for similar contracts if possible, otherwise on a model and assumptions about future nominal cashflows under the contract. While I am not aware of guidance for estimating the fair value of loan contracts, the absence of a liquid secondary market for loan contracts and this general fair value guidance would suggest that the Call Report valuations are prepared using prices of syndicated loans or mortgage-backed securities as the closest available alternatives.

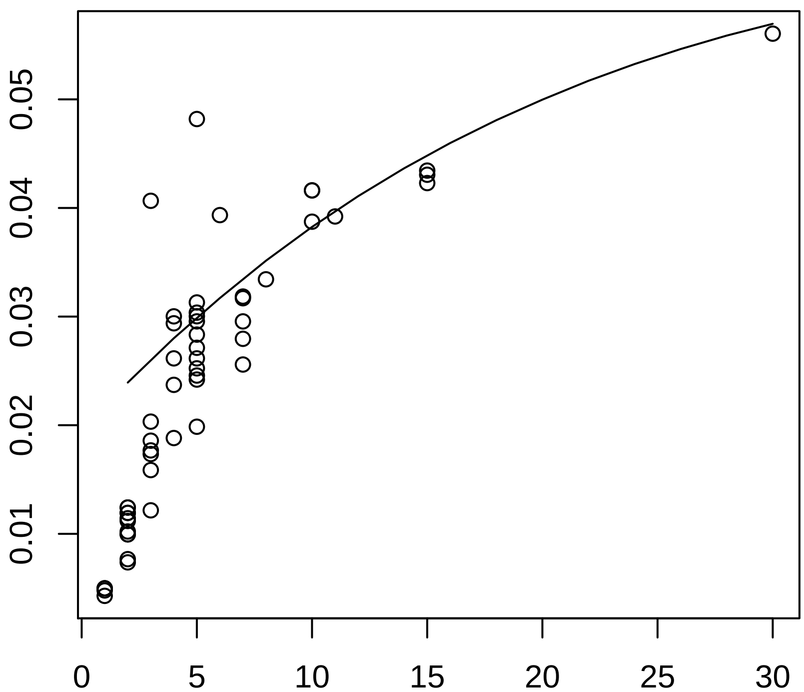
[31] The initial yield to maturity, or internal rate of return, of a fixed rate amortising loan contract is the interest rate that exactly discounts all expected future repayments to the face value of the loan.

2.B Svensson model fit in the second quarter of 2010

In the following figure, I show the data points $\{(m_j, y_j) : j = 1, \ldots, n\}$ for the second quarter of 2010 subsample. Each point j is drawn as a black circle with m_j on the horizontal axis and y_j on the vertical axis. In addition, I draw the yield to maturity function $\hat{y}(m)$ defined implicitly by

$$\frac{1 - e^{\hat{y}(m)m}}{\hat{y}(m)} = \int_0^m e^{-uz(u,\hat{\alpha})} du \tag{2.53}$$

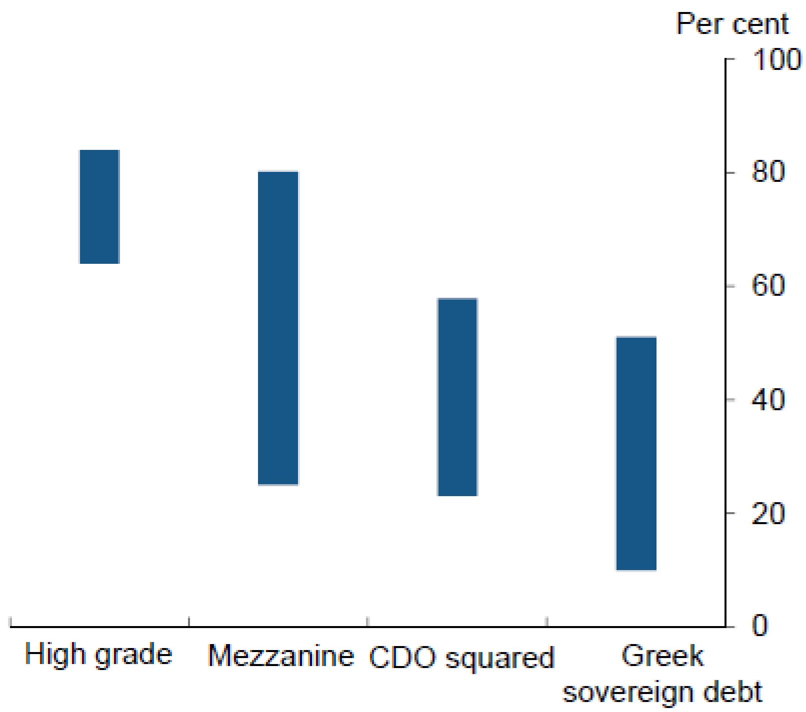
which represents the loan contract yields to maturity implied by the Svensson parameters $\hat{\alpha}$ estimated in the second quarter of 2010.



The observed yields to maturity on loan contracts are somewhat volatile and do not lie on a single curve, despite the fact that all the observations pertain to loans with similar Fitch credit ratings (rated 'A' or above). Heterogeneity is likely to be due to variation in the term structure of interest rates across days within the quarter and to variation in credit quality within the group of A-rated loans, but data limitations prevent my considering a time interval narrower than one quarter or a narrower range of credit ratings. The estimator $\hat{\alpha}$ minimises the sum of squared pricing errors, rather than yield errors, so there is no reason to expect $\hat{y}(m)$ to fit the observed yields closely. Nevertheless, the yields to maturity $\hat{y}(m)$ implied by the model are plausible.

2.C Survey measures of accounting valuation uncertainty

The following figure, entitled "ranges in reported valuations of structured credit products and sovereign bonds", appears in Haldane [2011] and Bank of England [2011]:



Haldane [2011] says that this figure reflects the range of prices at which assets of similar riskiness have recently been reported on the balance sheets of banks. Bank of England [2011] page 42 states that the figure shows the variation in banks' valuations of Greek sovereign debt.

Haldane [2011] and Bank of England [2011] provide the following notes to this figure:

- Sources: Citigroup, company reports, SEC filings and Bank of England calculations.
- Implied or reported marks on selected structured credit products by five banks at end-2007. The range of implied marks is not based on a like-for-like comparison of individual exposures, which might differ in their precise characteristics. So the chart should only be interpreted as an illustrative indicator of valuation uncertainty.
- Impairment charges on available-for-sale holdings of Greek sovereign debt by 24 European banks as of 2011 Q2.

Based on this information I interpret the vertical bars in the above figure as ranges of prices of bond-like contracts per unit of principal. For example, the high grade category shows of the banks surveyed that they price roughly the same highly rated government bond between 0.7 and 0.8 per unit of principal, or a range of valuations of $0.8 - 0.7 = 10\%$.

Chapter 3

Measuring dependence through the entropy rate

I consider a measure of serial dependence in stochastic processes based on the compressibility of samples under modern compression algorithms. I relate the measure, which depends on the entropy rate of the stochastic process, to traditional measures of dependence, including those based on entropy like Hong and White [2004]. I show that the measure of dependence can be applied to random vectors with continuous state spaces, to cross-sections and to detect Markov lag order for time series. I present numerical evidence on the finite and large sample distributions of the measure under serial independence. These distributions can be used to choose critical values for tests of dependence. I compute the size and power of such a test for the special cases of a Markov chain and AR(1) process based on finite sample critical values. The test based on the entropy rate has more power to detect dependence in the Markov chain considered than a standard test based on correlation measures.

Keywords: entropy rate, universal code, dependence, hypothesis test

3.1 Introduction

Deciding whether time series or cross-sectional data exhibit some form of dependence is important for specifying an appropriate model for the data, and for verifying that the unmodelled component of this data does not exhibit any discernable structure. For example, before specifying a model for some time series data, it would be useful to know that any stationary, ergodic model for the data can be limited to Markov processes of lag order at most 2; after specifying a model and fitting any parameters, it would be useful to know that we cannot find evidence for dependence in the residuals of the model. Measuring and testing for dependence is therefore important for model specification and goodness-of-fit testing.

Furthermore, in certain areas of econometrics we maintain assumptions of independence or conditional independence which allow us either to appeal to laws of large numbers to justify an interpretation of the parameters in a model being estimated or to appeal to central limit theorems to justify an inference procedure. In some settings, these assumptions take the form of zero correlation between random variables, which can be tested using t- and F-tests. In other settings, however, these assumptions take the form of independence between random variables, which null hypothesis the preceding correlation-based tests may not have any power to reject. Measuring and testing for dependence, as opposed to correlation, is therefore important for testing such exogeneity assumptions.

3.1.1 Definition of independence

Random variables X and Y are independent if their joint distribution function F_{XY} satisfies $F_{XY}(x, y) = F_X(x)F_Y(y)$ for all x, y, where F_X and F_Y are the distribution functions of X and Y. This is the cross-sectional case, and for the time series case of a stochastic process $\{X_i : i \in \mathbb{N}\}$, we may set $X = X_i$ and $Y = X_j$ for some i, j. X and Y could be vectors. For example, in the time series case we may set $X = (X_1, X_2, X_3)$ and $Y = (X_4, X_5, X_6)$. Based on observations $\{(x_i, y_i) : i = 1, \ldots, n\}$ from X and Y, we would like to conduct a hypothesis test for dependence.

In the past, measuring dependence in real-valued random variables often amounted to measuring correlation between pairs of them. However, it is well known that non-zero correlation is a sufficient, but not a necessary, condition for dependence between two random variables. For example, if we consider a zero-mean symmetrically distributed random variable X, then X and X^2 have zero correlation with each other even though they are completely dependent.[1] Among sets of random variables, portmanteau statistics have been proposed to measure the aggregate pairwise correlation, but these cannot capture setwise dependencies that are not revealed pairwise. For example, if $X_1, X_2, Z \overset{\text{iid}}{\sim} N(0, 1)$ and $X_3 = \text{sign}(X_1 X_2) \cdot |Z|$, then (X_1, X_2, X_3) are pairwise but not jointly independent [Romano and Siegel, 1986].

In the cross-sectional context, we have two random variables or vectors X and Y and the null hypothesis is that they are independent, which we write as $\mathbb{H}_0$: $X \perp\!\!\!\perp Y$. In the time series context, we have a strictly stationary, ergodic stochastic process $\{X_t\}$ and the null hypothesis is that there is no serial dependence, or equivalently that the process is independent and identically distributed (iid), which we write $\mathbb{H}_0$: $\{X_t\}$ iid.

3.1.2 Correlation-based measures of and tests for dependence

Much of this literature review is based on the excellent review by Tjostheim [1996].

[1] X and X^2 have zero correlation because the covariance between them is $Cov(X, X^2) = \mathbb{E}X^3 = 0$. X and X^2 are not independent because $\sigma(X) \cap \sigma(X^2) = \sigma(X) \neq \emptyset$, where $\sigma(Y)$ denotes the sigma field generated by the random variable Y.

3.1.2.1 Pearson correlation

The population Pearson product-moment correlation coefficient between random variables X and Y is

$$\frac{\operatorname{cov}(X,Y)}{\sqrt{\mathbb{V}X \cdot \mathbb{V}Y}} \tag{3.1}$$

This population quantity can be consistently estimated by the sample Pearson correlation coefficient, which replaces the population covariance and variances in (3.1) by the analogous sample covariance and variances respectively.

3.1.2.2 Spearman's rho

Spearman's rho is defined as the Pearson product-moment correlation coefficient between the probability integral transforms (or *grades*) of X and Y as

$$\frac{\operatorname{cov}\left(F_X(X), F_Y(Y)\right)}{\sqrt{\mathbb{V}F_X(X) \cdot \mathbb{V}F_Y(Y)}} = 12\mathbb{E}[F_X(X)F_Y(Y)] - 3.$$

Spearman's rho is invariant under monotonic transformations of X and Y. Spearman's rho can be consistently estimated in a finite sample by the product-moment correlation coefficient of the ranks of the random variables, which we denote ρ_n. Fieller et al. [1957] showed for a bivariate normal distribution that

$$\sqrt{\frac{n-3}{1.06}}\arctan\rho_n = \sqrt{\frac{n-3}{1.06}}\frac{1}{2}\log\frac{1+\rho_n}{1-\rho_n}$$

has approximately a standard normal distribution, for n "not too large and not too small". Since ρ_n is invariant under monotonic transformations of X and Y, the bivariate normal parent distribution assumption is not too restrictive. Large sample means and variances of ρ_n for the bivariate normal distribution case have also been known since Moran [1948], Kendall [1949], David et al. [1951]:

$$\begin{aligned}
\mathbb{E}\rho_n &= \frac{6}{(n+1)\pi}\left(\arcsin\rho_n + (n-2)\arcsin(\rho_n/2)\right) \\
\mathbb{V}\rho_n &= \frac{1}{n}\big(1 - 1.563465\rho_n^2 + 0.304743\rho_n^4 + 0.155286\rho_n^6 \\
&\quad +0.061552\rho_n^8 + 0.022099\rho_n^{10} + \dots\big).
\end{aligned}$$

3.1.2.3 Kendall's tau

Kendall's tau is a measure of dependence between a pair of random variables based on a notion of *concordance.* Given random variables X and Y, Kendall's tau is defined as

$$\tau_{X,Y} = \mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) > 0\right) - \mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) < 0\right),$$

where (X_1, Y_1) and (X_2, Y_2) are independent random vectors, equal in distribution to (X, Y). For a random sample $\{(x_i, y_i)\}_{i=1}^n$ from the distribution of (X, Y), the particular realisations (x_i, y_i) and (x_j, y_j), $\forall i \neq j$, are said to be *concordant* if $(x_i - x_j)(y_i - y_j) > 0$ and *discordant* if $(x_i - x_j)(y_i - y_j) < 0$. This pair of realisations from (X, Y) is neither concordant nor discordant if $x_i = x_j$ or $y_i = y_j$. Denoting the number of concordant pairs in our sample by c and the number of discordant pairs by d, and noting that there are $\binom{n}{2}$ distinct pairs in our sample, Kendall's tau is defined for a random sample as

$$\tau_n = \frac{c - d}{c + d}.$$

If there are no ties in the ranks of $\{x_i\}$ or $\{y_i\}$, then $\tau_n \in [-1, 1]$. If there are ties in the ranks, then τ_n is usually modified. Kendall's tau also has the desirable property that it is invariant under monotonic transformations of X and Y.

Under the null hypothesis that X and Y are independent, we have $\mathbb{E}\tau_n = 0$ and $\mathbb{V}\tau_n = 2(2n + 5)/9n(n - 1)$. Typically for $4 \leq n \leq 10$, hypothesis testing is done by means of tables, while for $n > 10$ the distribution of τ_n is approximated by the normal distribution and independence is rejected if

$$|\tau_n| > \Phi(1 - \frac{\alpha}{2})\sqrt{\frac{2(2n + 5)}{9n(n - 1)}}.$$

3.1.3 Measures of and tests for dependence based directly on the distribution function

From the definition of dependence, it seems natural to measure the degree of dependence between two random variables X and Y using the distance between their joint distribution function $F_{XY}(x, y)$ and the product of their marginal distribution functions $F_X(x)F_Y(y)$. Alternatively, one could measure the degree of dependence between these two random variables using the distance between their joint density function $f_{XY}(x, y)$ and the product of their marginal density functions $f_X(x)f_Y(y)$, if these density functions exist. For example, Granger et al. [2004] propose to measure dependence in the population using

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sqrt{f_{XY}(x,y)}-\sqrt{f_X(x)f_Y(y)}\right)^2 dydx, \tag{3.2}$$

which is the Hellinger distance between $f_{XY}(x,y)$ and $f_X(x)f_Y(y)$. The authors propose to estimate f_{XY}, f_X and f_Y in a finite sample using nonparametric kernel density functions. For the stationary stochastic process $\{X_t\}$, the degree of dependence at lag s can be measured using (3.2) with $X = X_t$ and $Y = X_{t-s}$. They show that the nonparametric kernel density estimator test statistic converges in distribution to a normal distribution as the sample size increases, under the null hypothesis that X and Y are independent. The quantiles of this normal distribution therefore provide critical values that can be used in large samples to test the null hypothesis that X and Y are independent. However, the authors find that size and power performance of the test based on this asymptotic approximation are poor. They argue instead for simulating the distribution of the test statistic under the null hypothesis for any given sample size.

Blum et al. [1961] measure dependence between the random variables X and Y using the Cramér-von Mises population measure

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(F_{XY}(x,y)-F_X(x)F_Y(y)\right)^2 dF_{XY}(x,y), \tag{3.3}$$

where F_{XY} and F_X, F_Y are estimated in finite samples by the empirical joint and marginal distribution functions respectively. Genest and Rémillard [2004] apply the $p-$dimensional analog of the measure (3.3) to measure dependence in the set of p random variables $X_1, \ldots, X_p$.[2] The authors also aggregate this p-dimensional analog across all subsets of at least 2 random variables. They compute finite sample critical values under the null hypothesis of total independence in the set $\{X_1, \ldots, X_p\}$.

3.1.4 Measures of and tests for dependence based on entropy

An excellent introduction to entropy, data compression and related concepts is provided in Cover and Thomas [2006].

[2] The p-dimensional analog of (3.3) is

$$\mathbb{E}[F_{X_1\cdots X_p}(X_1,\ldots,X_p)-F_{X_1}(X_1)\cdots F_{X_p}(X_p)]^2$$

where the expectation is taken with respect to $F_{X_1\cdots X_p}$.

3.1.4.1 Definitions of entropy and the entropy rate

The *entropy* of a discrete random variable X with probability mass function $\{p_j\}$ is $-\sum_j p_j \log p_j$, which we write as $H(X)$. If the logarithm is taken to base 2, the entropy is measured in binary units called *bits*, and if the logarithm is taken to be the natural logarithm, the entropy is measured in *nats*. Note that X could be a vector or a scalar. In the class of discrete distributions, entropy is nonnegative and maximal for the uniform distribution ($p_j = p_{j'} \forall j, j'$).

The entropy, or *differential entropy*, of a continuous random variable X with probability density function f is defined as $-\int_{-\infty}^{\infty} f(x) \log f(x) dx$ and is written $H(X)$ or $H(f)$. It is readily seen that the entropy can be written as $-\mathbb{E} \log f(X)$, which by the weak law of large numbers is the limit of $-\frac{1}{n}\sum_{t=1}^{n} \log f(X_t)$ as $n \to \infty$ for the stationary and ergodic stochastic process $\{X_t\}$ where $X_t \sim f$ for all t.

The differential entropy in nats of the standard normal distribution is

$$\begin{aligned} H(N(0,1)) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \left(-\frac{1}{2}x^2 - \frac{1}{2}\log 2\pi \right) dx \\ &= \frac{1}{2}\mathbb{E}N(0,1)^2 + \frac{1}{2}\log 2\pi \\ &= \frac{1}{2}\log 2\pi e, \end{aligned}$$

and a similar argument can be used to show $H(N(\mu, \sigma^2)) = \frac{1}{2}\log 2\pi e \sigma^2$. The differential entropy in nats of the multivariate normal distribution with positive definite covariance matrix Σ is

$$\frac{1}{2}\log |2\pi e \Sigma|. \tag{3.4}$$

The *entropy rate* of a stochastic process $\{X_t\}$ is defined as the limit $h(\{X_t\}) = \lim_{t\to\infty} H(X_1, \ldots, X_t)/t$. It can be interpreted as the average amount of information contained in one symbol of the sequence. In particular, the entropy rate of an independent and identically distributed stochastic process $\{X_t\}$ is[3]

$$\lim_{t\to\infty} \frac{H(X_1, \ldots, X_t)}{t} = \lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} H(X_i) = H(X_1).$$

3.1.4.2 Measures for dependence using entropy

An important property of entropy for any two random variables X and Y is that

$$H(X, Y) \leq H(X) + H(Y) \tag{3.5}$$

[3]Note that $H(X_t) = H(X_1)$ for all t in this case, because the distribution of X_t is identical for each t.

with equality if and only if X and Y are independent. This property also holds for random vectors X and Y, and it shows why entropy arises naturally in the context of testing for dependence. If we can estimate $H(X,Y)$, $H(X)$ and $H(Y)$ based on a sample then we can estimate $H(X)+H(Y)-H(X,Y)$, which can be used as a test statistic. Large (positive) values of the test statistic would be evidence against the hypothesis that X and Y are independent.

Hong and White [2004] apply this measure of dependence to continuous random variables X and Y, which gives the Kullback–Leibler divergence function

$$\begin{aligned} D_{KL}(f_{XY}||f_X \otimes f_Y) &= \mathbb{E} \log \frac{f_{XY}(X,Y)}{f_X(X) f_Y(Y)} \\ &= \int_x \int_y f_{XY}(x,y) \log \frac{f_{XY}(x,y)}{f_X(x) f_Y(y)} dxdy. \end{aligned} \tag{3.6}$$

This statistic is the mutual information between X and Y, or equivalently the amount by which the entropy of the joint distribution (X,Y) falls short of the sum of the entropies of the marginal distributions of X and Y. It can be seen as the expected value of a log likelihood ratio between the joint and the product of the marginal densities. The authors propose to estimate (3.6) by estimating f_{XY}, f_X and f_Y using non-parametric kernel density functions. Their contribution is to show the asymptotic distribution of this test statistic under the null hypothesis that X and Y are independent random variables. They extend their results to the time series case by finding the asymptotic distribution of the portmanteau statistic

$$\sum_{s=1}^{T} D_s \tag{3.7}$$

where D_s is the distance (3.6) with $X = X_t$ and $Y = X_{t-s}$, and T is an integer lag length chosen by the researcher.

3.1.4.3 Measures for dependence using the entropy rate

Analogously to property (3.5), the entropy rate of a stationary, ergodic random process $\{X_t\}$ satisfies

$$h(\{X_t\}) \leq H(X_1) \tag{3.8}$$

with equality if and only if $\{X_t\}$ is independent and identically distributed. This property also leads itself naturally to a test for independence in any stationary, ergodic stochastic process $\{X_t\}$. If it is possible to estimate $H(X_1) - h(\{X_t\})$ from a sample, large (positive) values of this test statistic are evidence for serial dependence. It turns out that $h(\{X_t\})$ can be estimated consistently using the length of the binary sequence that arises when a data compression algorithm is applied to a sample from $\{X_t\}$.

A *compression algorithm* or *code* is a mapping from one discrete state space, the state space of discrete stochastic processes, to another discrete state space, the space of finite-length binary sequences. A code is therefore a rule for representing sequences of symbols as sequences of zeros and ones. A code φ is a map, which we can write $(x_1, \ldots, x_t) \mapsto \varphi(x_1, \ldots, x_t)$, where $(x_1, \ldots, x_t)$ is a realisation of length t from a discrete state space stochastic process. Most operating systems today come with standard compression algorithms, and these are examples of codes.[4] If each $x_t \in \{0, 1\}$, then a compression algorithm is useful if $|\varphi(x_1, \ldots, x_t)| < t$, where $|\cdot|$ denotes length, because then the compressed sequence is shorter than the original sequence.

A code φ is a *universal code* if

$$\frac{|\varphi(X_1, \ldots, X_t)|}{t} \xrightarrow{P} h(\{X_t\}) \text{ as } t \to \infty \tag{3.9}$$

for any discrete valued stochastic process $\{X_t\}$. Universal codes therefore provide a natural estimator for the entropy rate of any data generating process by counting the number of binary digits output by the compression algorithm when applied to a sample from $\{X_t\}$. The idea of using universal codes to test for dependence has been investigated by Ryabko and Astola [2005, 2006a,b]. They propose to estimate $h(\{X_t\})$ using (3.9), and they suggest estimating $H(X_1)$ using

$$-\sum_x \hat{p}(x) \log \hat{p}(x) \tag{3.10}$$

where $\hat{p}$ is the empirical probability mass function

$$\hat{p}(x) = \frac{1}{t} \sum_{s=1}^{t} \mathbb{I}\{X_s = x\} \tag{3.11}$$

for the random realisation $\{X_1, \ldots, X_t\}$ from $\{X_s\}$. Their argument can be adapted to place a lower bound on the distribution function of their estimator for $H(X_1) - h(\{X_s\})$ under the null hypothesis that $\{X_s\}$ is serially independent. This lower bound consequently limits the size of the rejection region of such a test in finite samples, for any choice of universal code φ. Let $x_1, \ldots, x_t$ be a realisation of length t from the stationary and ergodic stochastic process $\{X_s\}$ with discrete state space S. Estimate the measure of dependence $H(X_1) - h(\{X_s\})$ by

$$Y = -\sum_{x \in S} \hat{p}(x) \log \hat{p}(x) - \frac{|\varphi(X_1, \ldots, X_t)|}{t} \tag{3.12}$$

where φ is any universal code, $X_1, \ldots, X_t$ is a random sequence drawn from $\{X_s\}$ and $\hat{p}$ is

[4] An example of a well-known compression algorithm is Zip. In this paper we are only interested in lossless compression, which are codes that are uniquely decodable/invertible.

defined by (3.11). Then the event $Y > y$ for any non-negative number y implies

$$\sum_{x \in S} t\hat{p}(x) \log \hat{p}(x) < -yt - |\varphi(X_1, \ldots, X_t)|. \tag{3.13}$$

If natural logarithms are used, then the distribution function of Y can be bounded below using

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= 1 - \sum_{x_1,\ldots,x_t : Y > y} \mathbb{P}\left[(X_1, \ldots, X_t) = (x_1, \ldots, x_t)\right] \\
&= 1 - \sum_{x_1,\ldots,x_t : Y > y} \prod_{x \in S} p(x)^{t\hat{p}(x)} && (3.14) \\
&\geq 1 - \sum_{x_1,\ldots,x_t : Y > y} \prod_{x \in S} \hat{p}(x)^{t\hat{p}(x)} && (3.15) \\
&\geq 1 - \sum_{x_1,\ldots,x_t : Y > y} \exp(-yt - |\varphi(x_1, \ldots, x_t)|) && (3.16) \\
&= 1 - e^{-yt} \sum_{x_1,\ldots,x_t : Y > y} \exp(-|\varphi(x_1, \ldots, x_t)|) \\
&\geq 1 - e^{-yt} \sum_{(x_1,\ldots,x_t) \in S^t} \exp(-|\varphi(x_1, \ldots, x_t)|) && (3.17) \\
&\geq 1 - e^{-yt} && (3.18)
\end{aligned}$$

where (3.14) follows from the form of the likelihood for the realisation $x_1, \ldots, x_t$ under the null hypothesis that $\{X_s\}$ is iid, (3.15) follows from the fact that $\hat{p}$ maximises the likelihood, (3.16) follows from the fact that each $x_1, \ldots, x_t$ in the summation is an element of the set $\{(X_1, \ldots, X_t) : Y > y\}$ and (3.13) holds for any such element, (3.17) follows because more terms are added to the summation, and (3.18) follows from the Kraft inequality[5] for any universal code φ. This bound becomes $1 - 2^{-yt}$ if logarithms are taken to base 2. The inverse of the cumulative distribution function for Y is therefore bounded from above by the inverse of the expression (3.18) as a function of y. In particular, the $1 - \alpha$ quantile of Y is bounded above by $-\frac{1}{t} \log \alpha$.

3.1.5 Lempel–Ziv compression

Two important compression algorithms that are universal codes in the sense of (3.9) are sliding window LZ77 [Ziv and Lempel, 1977] and tree-structured LZ78 [Ziv and Lempel, 1978]. The LZ77 algorithm maintains a fixed number of data points in a window. It passes through the original sequence and encodes each recurring pattern within the window as a reference to the last time that pattern occurred within the window, while data points that do not match other data in the window are left unaltered. For example, consider the realisation $ACBACBBAC$

[5] Cover and Thomas [2006] define and discuss the Kraft inequality.

from a stochastic process with discrete state space $\{A, B, C\}$. LZ77 with a window of 4 data points divides up the sample into phrases $A|C|B|ACB|BAC$, which could be represented, for example, as

$$(0, A), (0, C), (0, B), (1, 3, 3), (1, 4, 3).$$

The last representation uses the initial digit to indicate whether the character in the original sample at that position was matched (0 or 1). Unmatched characters are encoded uncompressed and matched characters are encoded as a pair indicating the position of the match in the sliding window, as a number between 1 and 4 counting backwards along the last 4 characters seen, and the number of characters in the matched phrase.[6] In this example, the brackets and commas are not needed by the algorithm, so the original sequence $ACBACBBAC$ of length 9 characters is encoded by LZ77 as the sequence $0A0C0B133143$ of length 12 characters. The LZ77 algorithm does not succeed in shortening the length of the original sequence because the original sequence and windows are both relatively short.

The LZ78 algorithm maintains a tree of past phrases called a dictionary, without any window. At every point in the encoding procedure, LZ78 encodes the shortest subsequence (i.e. phrase) not yet seen as a reference to a phrase in the dictionary and an uncompressed character. The sample $ACBACBBAC$ is divided into phrases by defining each new phrase to be the shortest subsequence not yet seen. This gives the phrases $A|C|B|AC|BB|AC$. Because each phrase is the shortest subsequence not yet seen, each phrase consists of a previously seen phrase and one unseen character. These phrases could be represented as

$$(0, A), (0, C), (0, B), (1, C), (3, B), (4,),$$

where each pair in the encoding represents a phrase by a pointer to a previous phrase and an uncompressed character. The final pair points to the fourth phrase AC in the dictionary and no extra uncompressed character. The LZ78 algorithm therefore encodes the original sequence $ACBACBBAC$ of length 9 characters as the sequence $0A0C0B1C3B4$ of length 11, which again does not shorten the original sequence in this example, but could be expected to do so in applications to longer original sequences.

A common feature of the Lempel–Ziv algorithms is that they look for patterns in the original data. The more recurring subsequences that an algorithm is able to find, and the longer those subsequences, the fewer phrases the algorithm uses to divide up the original sequence. Encoding

[6] In the above example, the first three characters A, C, B are encoded in their original format. At this point the window consists of these three characters in their original order. The algorithm scans ahead and notes that the next three characters A, C, B in the sequence exist in the same order in the window. The algorithm records this fact with a 1 indicating a match, a 3 indicating that the first character A appears in the window three characters away from the right endpoint of the window and another 3 indicating that the match consists of three characters (A, C, B). At this point the window contains B, A, C, B and the algorithm notes that the subsequent phrase BAC appears in the window 4 characters from the right endpoint of the window and the matched phrase BAC is of length 3 characters.

a long subsequence as a reference substantially shortens its length, while unmatched characters like $(0, A)$, $(0, B)$ and $(0, C)$ require longer binary representations than their original characters A, B and C.

3.1.6 Structure of this paper

In Section 3.2.1, I first propose an estimator for the degree of dependence based on a modified version of a statistic due to Ryabko and Astola [2005] for univariate, discrete state space stochastic processes. I then discuss extensions of the statistic to measure dependence in cross-sections and in stochastic processes with continuous state spaces. In Section 3.2.2 I show that a test for serial dependence at lag one or more using the LZ78 compression algorithm can be constructed using an asymptotic normal distribution due to a result by Jacquet and Szpankowski [2011]. I also present new numerical evidence on the convergence properties of the test statistic under the null for the more modern LZMA and BZ2 compression algorithms, for which analytic results are not available.

In Section 3.3 I present Monte Carlo evidence on the size and power of these tests based on the modern LZMA compression algorithm. Section 3.4 concludes with directions for further research.

3.2 Method

3.2.1 Sample statistic for measuring dependence

3.2.1.1 Discrete state space stochastic process

Ryabko and Astola [2005, 2006a,b] use the population measure $H(X_1) - h(\{X_t\})$ and the sample statistic in (3.12) to measure dependence in a discrete state space stochastic process. If the support of a process $\{X_t\}$ is discrete with k mass points, we can scale the inequality in (3.8) by $\log k$ and express it as

$$\frac{H(X_1) - h(\{X_t\})}{\log k} = \left(1 - \frac{h(\{X_t\})}{\log k}\right) - \left(1 - \frac{H(X_1)}{\log k}\right), \tag{3.19}$$

which is the population measure of dependence that I consider. In the important special case where the stationary distribution of X_t is uniform, $H(X_1) = \log k$ and the second term in (3.19) is zero. The state space of $\{X_t\}$ could be numeric or categorical, which does not affect any of the quantities in (3.19). Define an ordering on the state space of X_1.[7] Then we have a population (marginal) cumulative distribution function $F(x) = \mathbb{P}\{X_1 \leq x\}$ and we can define the empirical

[7]When this state space is categorical it should be enumerated, for example in lexicographical order. The choice of ordering is arbitrary and does not affect the test statistic.

cumulative distribution function

$$\hat{F}(x) = \frac{1}{t}\sum_{s=1}^{t} \mathbb{I}\{X_s \leq x\} \tag{3.20}$$

for all $x \in \mathbb{R}$. When $\{X_t\}$ is stationary and ergodic, $\hat{F}(x) \xrightarrow{P} F(x)$ as $t \to \infty$ for all x. Any $\hat{F}$ with this pointwise consistency property would be suitable in place of (3.20). We know that $F(X_t)$ is uniformly distributed with k mass points and hence $H(F(X_1)) = \log k$.

Consider the population measure of dependence (3.19) applied to the discrete stochastic process $\{F(X_t)\}$, which is

$$\rho(\{X_t\}) \equiv 1 - \frac{h(\{F(X_t)\})}{\log k}. \tag{3.21}$$

Note that $\rho \in [0,1]$, and $\rho \geq 0$ with equality if and only if $\{X_t\}$ is independent and identically distributed. Quantity (3.21) is known as the redundancy rate or optimal compression ratio because it gives an upper bound on the extent to which a sample from any stationary, ergodic, discrete uniform stochastic process $\{F(X_t)\}$ can be compressed. Universal codes are compression algorithms that achieve this upper bound asymptotically, and therefore can be used to obtain a consistent estimator of ρ based on a realisation of $\{F(X_t)\}$. Because we only observe a realisation of $\{X_t\}$, we have to estimate F.

Consider the test statistic $\hat{\rho}$, which is the *compression ratio*, or *redundancy rate*. When $x_1, \ldots, x_t$ is a realisation from a discrete state space data generating process with k mass points, the test statistic can be computed directly as

$$\hat{\rho}(x_1, \ldots, x_t) = 1 - \frac{|\varphi(\hat{F}(x_1), \ldots, \hat{F}(x_t))|}{t \log k}, \tag{3.22}$$

where $|\cdot|$ denotes length. The transformed sample $(\hat{F}(x_1), \ldots, \hat{F}(x_t))$ is a realisation from the state space with k^t elements, which can be enumerated with $t \log_2 k$ binary digits.[8] When logarithms are taken to base 2, the quantity $1 - \hat{\rho}$ from (3.22) therefore has a natural interpretation as the ratio of the number of binary digits in the compressed sequence to the number of binary digits in the original sequence. By working with $\hat{F}(X_t)$ and (3.22) in place of X_t and (3.12), we avoid having to estimate $H(X_1)$. The scaling factor $1/\log k$ introduced in (3.19) ensures that $\hat{\rho}$ in (3.22) lies between 0, total independence, and 1, total dependence.

Note that the population quantity (3.21) is a property of the data generating process for $\{X_t\}$, while the sample quantity (3.22) is a function of a finite set of points. Using the fact that

$$\hat{\rho}(X_1, \ldots, X_s) \xrightarrow{P} \rho(\{X_t\}) \tag{3.23}$$

[8] Or $\lceil t \log_2 k \rceil$ binary digits if $t \log_2 k$ is not an integer.

as $s \to \infty$, and that $\rho \geq 0$ with equality if and only if $\{X_t\}$ is independent and identically distributed, we can interpret large (positive) values of the test statistic as evidence for serial dependence. Compression algorithms work by finding patterns in binary representations of data and then removing recurring patterns from the binary representation. On a given sample, if the compression algorithm φ shortens the sample's binary representation, then the compression algorithm has found patterns in the sample, which are evidence for dependence in the population $\{X_t\}$. If a stationary, ergodic, discrete stochastic process is independent through time, then it is not compressible, and samples from this data generating process will be compressible only by chance.

3.2.1.2 Continuous state space stochastic process

When the support of the data generating process is continuous, some discretisation has to be applied to the state space before the test statistic (3.22) is computed, in order for this statistic to have a well-defined probability limit. Multiple discretisations are possible for a given sample. We can restrict ourselves to discretisations of compact state spaces, because it is always possible to ensure that the state space of a random variable is compact.[9] A discretisation is defined by a function $[\cdot]_Q$ which maps all points in a compact state space to some discrete state space with k elements.

There are three ways of discretising the stationary, ergodic, continuous state space stochastic process $\{X_t\}$, with the stationary distribution function F, under the null hypothesis that $\{X_t\}$ is serially independent, choice of which depends on what is known in advance:

1. F known: In this case $F(X_t)$ is iid $U[0,1]$ and if $[\cdot]_Q$ partitions the state space into k equal-sized bins then $[F(X_t)]_Q$ is iid uniform on the discrete state space with k elements.

2. Estimate G of F given, which does not depend on the sample size: For example, G could be an estimate from another dataset, or some transformation like the sigmoid function. In this case $G(X_t)$ lies in $[0,1]$ and $[G(X_t)]_Q$ lies in the discrete state space with k elements. Both are iid although they are not necessarily uniformly distributed.

3. $\hat{F}$ estimated consistently, for example using the empirical cumulative distribution function on a sample from $\{X_t\}$: In this case $\hat{F}(X_t)$ is serially dependent. For the empirical cumulative distribution function, this dependence vanishes asymptotically because $\hat{F}(X_s) = F(X_s) + \hat{F}(X_s) - F(X_s) = F(X_s) + O_p(t^{-1/2})$. For each t, if $[\cdot]_Q$ partitions the state space into k equal-sized bins then $[\hat{F}(X_t)]_Q$ is uniformly distributed on the discrete state space with k elements.

[9] Entropy and hence our dependence measure ρ are invariant to monotone transformations, so any monotonic function that maps a continuous state space into $[0,1]$ can be used. For example, the sigmoid function $s(x) = (1+e^{-x})^{-1}$ is a monotonic transformation from $\mathbb{R}$ to $(0,1) \subset [0,1]$.

The entropy of the discretised random variable approaches the entropy of the original random variable as the discretisation becomes finer, although there is an additive correction term. If X is a random variable with support $\mathbb{R}$, we can partition the support of the random variable into a finite collection of intervals $\mathcal{P} = \{a_i\}$ such that $\cup_i a_i = \mathbb{R}, a_i \cap a_{i'} = \emptyset$ for all $i \neq i'$ and $|\mathcal{P}| < \infty$. If X has density function f, then by the mean value theorem, for every a_i with $||a_i|| < \infty$, there is an $x_i \in a_i$ such that $\int_{a_i} f(x)dx = f(x_i)||a_i||$. We can then define the discretised random variable

$$[X]_{\mathcal{P}} = \begin{cases} x_i & \text{if } X \in a_i \text{ and } ||a_i|| < \infty \\ \pm\infty & \text{if } X \in a_i \text{ and } ||a_i|| = \infty \end{cases}$$

and note that the entropy of the discretised random variable is

$$\begin{aligned} H([X]_{\mathcal{P}}) &= -\sum_{a_i \in \mathcal{P}} \left(\int_{a_i} f(x)dx \right) \log \left(\int_{a_i} f(x)dx \right) \\ &= -\sum_{i:||a_i||<\infty} f(x_i)||a_i|| \left(\log f(x_i) + \log ||a_i|| \right) \\ &\quad - \sum_{i:||a_i||=\infty} \left(\int_{a_i} f(x)dx \right) \log \left(\int_{a_i} f(x)dx \right). \end{aligned} \tag{3.24}$$

Now consider a sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \ldots$ such that $|\mathcal{P}_1| < |\mathcal{P}_2| < |\mathcal{P}_3| < \ldots$, every $a \in \mathcal{P}_j$ has $||a|| \to \infty$ as $j \to \infty$, and

$$\cup \{a \in \mathcal{P}_j : ||a|| < \infty\} \to \mathbb{R} \text{ as } j \to \infty$$

which corresponds to finer partitions of the support of X. By construction, we have

$$\int_{\cup\{a \in \mathcal{P}_j : ||a|| = \infty\}} f(x)dx \to 0 \text{ as } j \to \infty$$

and hence the second term in (3.24) approaches zero. Then as $i \to \infty$ we have

$$H([X]_{\mathcal{P}_i}) + \sum_{a_i \in \mathcal{P}_i : ||a_i|| < \infty} f(x_i)||a_i|| \log ||a_i|| \to H(X).$$

In the special case where $||a_i|| = \frac{1}{k_i}$ for all $||a_i|| < \infty$ and for $k_i \to \infty$, we have the simpler statement $H([X]_{\mathcal{P}_i}) - \log k_i \to H(X)$ as $i \to \infty$. For example, if the support of X is $[0, 1]$, then we can set $k_j = 2^j$ for integer j, and then the entropy (with logs taken to base 2) of the discretised random variable $[X]_Q$ is approximately equal to $H(X) + j$ when j is large.

3.2.1.3 Cross-section

Consider the p continuous state space random variables $X_1, \ldots, X_p$ with population cumulative distribution functions $F_{X_1}, \ldots, F_{X_p}$. For some discretisation function $[\cdot]_Q$ mapping the unit interval $[0, 1]$ to the discrete state space of k numbers, the random vector

$$X = ([F_{X_1}(X_1)]_Q, \ldots, [F_{X_p}(X_p)]_Q)$$

is a discrete state space random variable with k^p mass points. Property (3.5) ensures that

$$H(X) \leq \sum_{j=1}^{p} H([F_{X_j}(X_j)]_Q) \tag{3.25}$$

with equality if and only if the joint distribution function of $X_1, \ldots, X_p$ can be factorised into a product of the marginal distribution functions of $X_1, \ldots, X_p$, i.e. if the set $\{X_1, \ldots, X_p\}$ is totally independent. If the discretisation function $[\cdot]_Q$ divides the unit interval into k equally-sized bins, then each $[F_{X_j}(X_j)]_Q$ is uniformly distributed with k mass points and entropy $\log k$, so that the right hand side of (3.25) becomes $p \log k$.

The left hand side of (3.25) could be estimated consistently by estimating a $p-$dimensional joint probability mass function, but this estimator would suffer from poor finite sample properties due to the curse of dimensionality. Instead, define an independent and identically distributed stochastic process $\{X_t\}$ with $X_t \stackrel{D}{=} X$ for all t, which is a stationary ergodic stochastic process with k^p mass points. The entropy rate of $\{X_t\}$ is simply $H(X)$, which by (3.9) can be estimated by applying any universal code to samples from $\{X_t\}$.

The null and alternative hypotheses can therefore be written using (3.25) as

$$\mathbb{H}_0 : \quad 0 = 1 - \tfrac{H(X)}{p \log k} \tag{3.26}$$

$$\mathbb{H}_1 : \quad 0 < 1 - \tfrac{H(X)}{p \log k} \tag{3.27}$$

which suggests the test statistic

$$1 - \frac{|\varphi(x_1, \ldots, x_n)|}{p \log k} \tag{3.28}$$

based on the realisation $x_1, \ldots, x_n$ from the stochastic process $\{X_t\}$. Large (positive) values of the test statistic are evidence against the null hypothesis that the set of random variables $\{X_1, \ldots, X_p\}$ is totally independent. I suggest estimating each F_{X_j} using the empirical marginal cumulative distribution function for X_j where $j = 1, \ldots, p$.

3.2.2 Asymptotic distribution under the null

For a realisation $x_1, \ldots, x_n$ from a stationary, ergodic, discrete stochastic process $\{X_t\}$ with k mass points,

$$\limsup_{n\to\infty} \hat{\rho}(x_1, \ldots, x_n) \leq 1 - \frac{h(\{X_t\})}{\log k}, \tag{3.29}$$

almost surely where $\hat{\rho}$ is calculated using any code [Shannon, 1948]. Universal codes are codes that achieve this lower bound asymptotically. Two such codes are the Lempel–Ziv algorithms LZ77 and LZ78, for which Wyner and Ziv [1991, 1994] showed that

$$\liminf_{n\to\infty} \hat{\rho}(x_1, \ldots, x_n) \geq 1 - \frac{h(\{X_t\})}{\log k} \tag{3.30}$$

almost surely. Therefore Lempel–Ziv algorithms are universal codes and can be used in place of φ in $\hat{\rho}$ to get the consistency in (3.23).[10]

For the 1978 Lempel–Ziv algorithm (LZ78) applied to an independent and identically distributed[11] stochastic proces $\{U_t\}$ with uniform marginal distribution on k mass points, Kirschenhofer et al. [1994], Jacquet and Szpankowski [1995, 2011] show that the compression ratio

$$\hat{\rho}_t = 1 - \frac{|\varphi(U_1, \ldots, U_t)|}{t \log k} \tag{3.31}$$

converges weakly and in moments to a normal distribution

$$\frac{\hat{\rho}_t - \mathbb{E}\hat{\rho}_t}{\sqrt{\mathbb{V}\hat{\rho}_t}} \xrightarrow{D} N(0, 1), \tag{3.32}$$

and the mean and variance of $\hat{\rho}_t$ can be expressed asymptotically as

$$\mathbb{E}\hat{\rho}_t \sim \frac{1}{\log t} \tag{3.33}$$

$$\mathbb{V}\hat{\rho}_t \sim \frac{C + \delta(\log t)}{(\log k)^2} \frac{1}{t \log t}, \tag{3.34}$$

where $C = 0.2600\ldots$ is a constant and $\delta(x)$ is a periodic continuous function of period 1 and small amplitude ($< 10^{-6}$). I provide a sketch of their proof in Appendix 3.A. Note that $\mathbb{E}\hat{\rho}_t \to 0 = 1 - \frac{H(U_1)}{\log k}$ as $t \to \infty$.[12] In particular, the test statistic $\hat{\rho}_t$ has a finite-sample bias that vanishes at a logarithmic rate, and a standard deviation that approaches zero at a rate marginally faster than the square-root rate.

[10] Consistency of $\hat{\rho}$ therefore relies on a stationarity assumption, which requires in particular that each X_t is identically distributed. Stationary and ergodicity are also assumed in Ryabko and Astola [2005, 2006a,b].

[11] A sequence of independent random variables is a stationary stochastic process if the random variables are identically distributed. Therefore 'iid' is sufficient for a stochastic process to be stationary and hence to apply (3.29), but the converse is not true.

[12] $H(U_1) = \log k$ because U_1 is assumed to be uniformly distributed with k mass points.

For any sample of size t from the discrete state space independent and identically distributed stochastic process $\{X_s\}$ with marginal cumulative distribution function F and empirical cumulative distribution function $\hat{F}(x) \xrightarrow{P} F(x)$ for all x we have the asymptotic equality

$$\hat{F}(X_s) \sim U_s \tag{3.35}$$

for all s and for some independent and identically uniformly distributed discrete state space stochastic process $\{U_s\}$, to which we can apply (3.32). Similarly, consider any sample of size t from a continuous state space independent and identically distributed stochastic process $\{X_s\}$ with marginal cumulative distribution function F and empirical cumulative distribution function $\hat{F}(x) \xrightarrow{P} F(x)$ for all x. Let $[\cdot]_Q$ be a discretisation function dividing the unit interval $[0, 1]$ into k equally-sized intervals. We then have the asymptotic equality

$$[\hat{F}(X_s)]_Q \sim U_s \tag{3.36}$$

for all s and for some independent and identically uniformly distributed discrete state space stochastic process $\{U_s\}$, to which we can apply (3.32). Therefore, neither the serial dependence induced by applying $\hat{F}$ to X_t nor the discretisation $[\cdot]_Q$ of $\hat{F}(X_t)$ prevent an appeal to the asymptotic results in (3.32), (3.33) and (3.34).

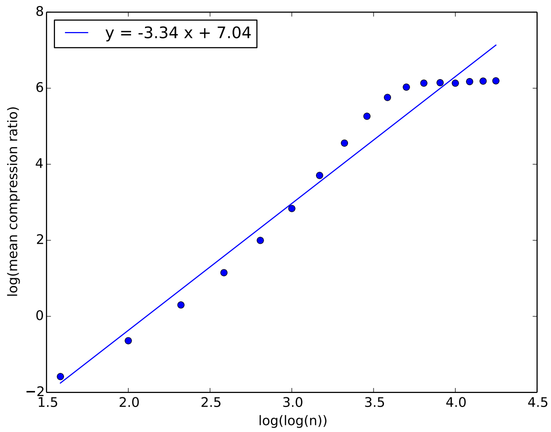
The current asymptotic theory is limited to the classic LZ77 and LZ78 algorithms. In practice, we would like to take advantage of more modern compression algorithms, which achieve faster asymptotic convergence rates than LZ78. Asymptotic convergence theorems for these modern algorithms are not yet available, so I investigate convergence numerically. To do so, I simulate $10,000$ independent realisations of time series of length n of independent and identically uniformly distributed random variables on 256 states, for different values of n.[13] For each time series sample, I calculate the summary statistic $\hat{\rho}$ in (3.22) with $t = n$, $\hat{F}$ known, $k = 256$, logarithms to base 2 and two modern compression algorithms for φ: the Lempel–Ziv Markov Algorithm (LZMA) and the Burrows–Wheeler BZip2 Algorithm (BZ2).[14] I summarise the distribution of $\hat{\rho}$ for each sample size n and compression algorithm.[15]

In Figure 3.1 I plot the mean and standard deviation across realisations of $\hat{\rho}$ against (transforms of) the sequence length n. The transform applied to each horizontal axis is my conjecture for the functional form of the relationship between the mean or standard deviation and the sample size n, and the slope of each (ordinary least squares) regression line gives an estimate

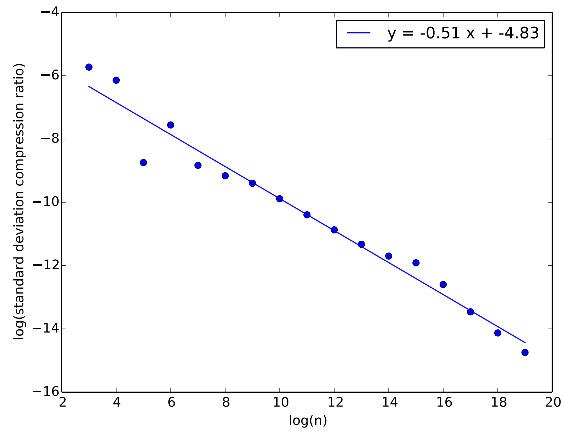
[13] Any integer between 1 and 256 can be represented by $\log_2(256) = 8$ binary digits, so that a realisation of length n of discrete 256-state uniformly distributed random variables is equivalent to a realisation of length $8n$ of independent and identically Bernoulli distributed random variables with common distribution parameter $1/2$. I use 256 states because the software implementations of the modern compression algorithms that I consider operate on sequences of blocks of length 8 of binary digits, called 'byte' sequences.

[14] $\hat{F}(x) = \min\{\max\{\lfloor x \rfloor/256, 0\}, 1\}$

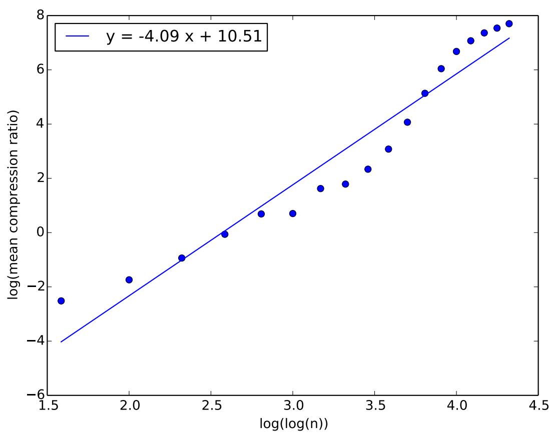
[15] For hypothesis testing, we are particularly interested in the quantiles of this distribution.



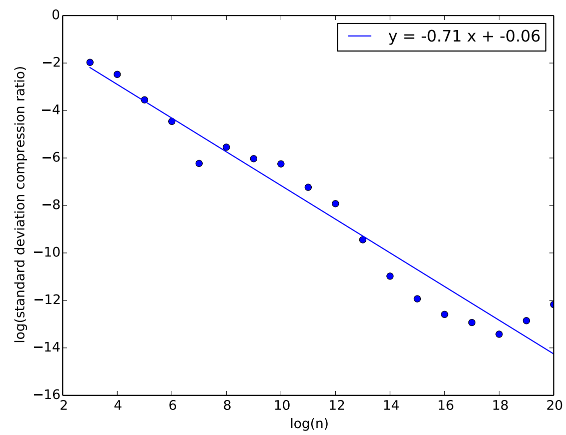
(a) LZMA mean



(b) LZMA standard deviation



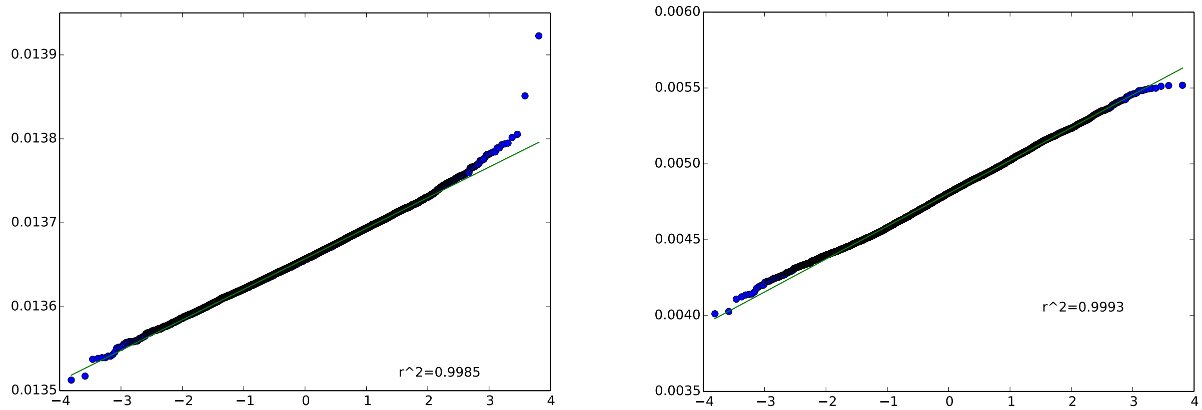
(c) BZ2 mean



(d) BZ2 standard deviation

Figure 3.1: Convergence in mean and standard deviation of the modern Lempel Ziv Markov and BZ2 algorithms when applied to sequences of discrete state space random variables, independent and identically distributed uniformly on 256 *states.*

of the free parameter describing the rate of convergence of these first and second moments. Figures 3.1a and 3.1c show that both algorithms achieve convergence in mean in polynomial-log time, with estimated rates $O((\log t)^{-3.34})$ and $O((\log t)^{-4.09})$ for LZMA and BZ2 respectively. Convergence in standard deviation is shown in Figures 3.1b and 3.1d. The standard deviation of the LZMA compression ratio decreases at approximately $\sqrt{t}$ rate, or $O(t^{-0.51})$, and the standard deviation of the BZ2 compression ratio decreases faster, approximately at rate $O(t^{-0.71})$. Modern compression algorithms, therefore, appear to achieve equal or faster rates of convergence in mean and standard deviation than the LZ78 compression algorithm. Figure 3.2 shows Q-Q plots against a normal distribution for the compression ratio test statistic under LZMA and BZ2 compression algorithms. The high correlation between the quantiles (with R^2 statistics in the Q-Q plots above 99%) suggests that an asymptotic normal distribution approximation is



(a) LZMA Q-Q plot, $n = 2^{19}$ *(b) BZ2 Q-Q plot,* $n = 2^{20}$

Figure 3.2: Q-Q plots for the distribution of the test statistic for samples of size n from independent and identically uniformly distributed random variables with 256 *states. The reference quantiles for the Q-Q plot are for the normal distribution.*

reasonable. The normal distribution approximation errors are largest in the tails of the test statistic distribution under the LZMA compression algorithm. Whether these errors are too large for a normal approximation to be useful depends on the particular application.

In Appendix 3.B I tabulate the upper quantiles of this distribution of $\hat{\rho}$ for each sample size n and compression alrgorithm. These quantiles can be used as exact critical values in a hypothesis test for dependence in a time series or cross section[16] where the stationary distribution is the symmetric Bernoulli,[17] or equivalently the uniform distribution on 2^m states for $m = 8, 12, \ldots, 21$.[18] The computed quantiles for the largest sequence lengths can be used as approximate asymptotic critical values in a hypothesis test for dependence in a time series or cross section regardless of the stationary (marginal) distribution, including discrete or continuous state space distributions.[19] The distributions of compression ratios are discrete rather than continuous. As the sample size (i.e. sequence length) grows, the distribution of compression ratios approaches a continuous distribution as in (3.32), but in small samples this discreteness is noticeable.[20]

[16] I discuss discrete time series and cross section applications in Section 3.2.1.

[17] The symmetric Bernoulli distribution is the Bernoulli distribution with parameter $p = 1/2$.

[18] I tabulate values of the upper quantiles of the distribution of $\hat{\rho}$ under the LZMA compression algorithm for random (iid) byte sequences of lengths $n = 2^k$ where $k = 3, 4, \ldots, 19$ from a marginal $U\{1, \ldots, 2^8\}$ distribution, which is the same as random (iid) binary sequences from the symmetric Bernoulli marginal distribution of length $8n = 2^{3+k}$, or random (iid) sequences of lengths 2^{3+k-m} for $3 + k - m = n$ from the $U\{1, \ldots, 2^m\}$ marginal distribution. For the BZ2 compression algorithm I obtain one extra set of quantiles for $n = 2^{20}$.

[19] I discuss the treatment of continuous state space distributions in Section 3.2.1.2.

[20] The discreteness of the distribution of compression ratios defined by $\hat{\rho}$ is common to all statistics based on lengths of sequences compressed by universal codes, in particular Ryabko and Astola [2005, 2006a,b].

3.2.3 Generalisation to detecting Markov lag order

Property (3.8) can be generalised to

$$H(X_m|X_{m-1}\dots X_1) \geq h(\{X_s\}) \tag{3.37}$$

for any positive integer m and any stationary, ergodic, discrete time Markov process $\{X_s\}$, with equality if and only if $\{X_s\}$ is Markovian of order at most $m-1$.[21] When $m = 1$, we get the special case (3.8), with equality if and only if $\{X_s\}$ is independent and identically distributed. The right hand side of (3.37) can be estimated consistently by (3.9) for any universal code φ. Ryabko and Astola [2006a] suggest estimating the left hand side of (3.37) when $\{X_s\}$ has discrete state space S using

$$-\sum_{(x_1,\dots,x_m)\in S^m} \hat{p}(x_m, x_{m-1}, \dots, x_1) \log \hat{p}(x_m|x_{m-1}, \dots, x_1) \tag{3.38}$$

where the $\hat{p}$s are the empirical joint and conditional probability mass functions

$$\hat{p}(x_m, x_{m-1}, \dots, x_1) = \frac{1}{t-m}\sum_{s=1}^{t-m} \mathbb{I}\{(X_{s+m}, \dots, X_{s+1}) = (x_m, \dots, x_1)\} \tag{3.39}$$

$$\hat{p}(x_m|x_{m-1}, \dots, x_1) = \frac{p(x_m, x_{m-1}, \dots, x_1)}{p(x_{m-1}, \dots, x_1)} \tag{3.40}$$

for all sequences $(x_m, \dots, x_1) \in S^m$ and all positive integers m. However, the size of the sample space S^m grows exponentially in m, making it computationally inefficient.

A computationally efficient alternative to computing (3.38) is to sample from (3.40) with replacement. This procedure defines a new stationary, ergodic stochastic process $\{\tilde{X}_s\}$ that is Markov of order $m-1$. This stochastic process satisfies property (3.37) with equality and hence any universal code estimator (3.9) of the entropy rate of $\{\tilde{X}_s\}$ is also an estimator of $H(X_m|X_{m-1}\dots X_1)$.

Re-writing (3.37) as

$$H(X_1) - h(\{X_s\}) \geq \dots \geq H(X_m|X_{m-1}\dots X_1) - h(\{X_s\}) \geq \dots \geq 0 \tag{3.41}$$

we get a decomposition of the serial dependence in $\{X_t\}$ into dependence at each lag m or beyond. Plotting the function $H(X_m|X_{m-1}\dots X_1) - h(\{X_s\})$ against m gives the *dependogram* analog of an autocorrelation function. Resampling methods give an estimate of the full

[21] (3.37) follows from the fact that $h(\{X_t\}) = \lim_{t\to\infty} H(X_t|X_{t-1}\dots X_1)$ for stationary $\{X_t\}$ and $H(X_t|X_{t-1}\dots X_1)$ is non-increasing in $t \in \mathbb{N}$. Additionally, a discrete time Markov process $\{X_t\}$ has $H(X_{t+m}|X_{t+m-1}\dots X_1) = H(X_{t+m}|X_{t+m-1}\dots X_{t+1}) = H(X_m|X_{m-1}\dots X_1)$ for $t, m > 0$ if and only if it is of order at most $m-1$, in which case $h(\{X_t\}) = \lim_{t\to\infty} H(X_{t+m}|X_{t+m-1}\dots X_1) = H(X_m|X_{m-1}\dots X_1)$.

distribution of $H(X_m|X_{m-1}\dots X_1) - h(\{X_s\})$ for each m, and the quantiles of each of these distributions can be used to construct confidence intervals around the dependogram.

3.3 Size and power experiments

3.3.1 Monte carlo experiments for autoregressive processes

My first Monte Carlo experiment is based on a mean zero, stationary, Gaussian AR(1) process $X_t = aX_{t-1} + \epsilon_t$, with $\epsilon_t \sim N(0, \sigma^2)$ for $t > 1$, $|a| < 1$ and initial conditions $\epsilon_1 \sim N(0, \frac{\sigma^2}{1-a^2})$ from the stationary distribution and $X_0 = 0$. I show that the Pearson correlation can be expected to be the best measure of dependence (in $\{X_t\}$) in the sense that it is larger than the measure of dependence (in $\{X_t\}$) based on entropy or entropy rate (in $\{X_t\}$) in the population. I then provide the Monte Carlo results that confirm this conjecture.

The entropy rate of this AR(1) process is

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} H(X_s|X_{s-1},\dots,X_1) &= \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} H(X_s|X_{s-1}) \\
&= \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} H(aX_{s-1}+\epsilon_s|X_{s-1}) \\
&= \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} H(\epsilon_s|X_{s-1}) \\
&= \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} H(\epsilon_s) \\
&= H(\epsilon_1) \\
&= \frac{1}{2}\log 2\pi e\sigma^2 \qquad (3.42) \\
&\le \frac{1}{2}\log 2\pi e\frac{\sigma^2}{1-a^2} \qquad (3.43) \\
&= H(X_1).
\end{aligned}$$

Using the inequality (3.8) we can express the population measure of dependence (3.21) for continuous state space stochastic processes as $H(X_1) - h(\{X_t\})$, which in the case of this AR(1) is

$$H(X_1) - h(\{X_t\}) = H(X_1) - H(\epsilon) = -\frac{1}{2}\log(1-a^2) \qquad (3.44)$$

An entropy rate-based test for dependence therefore compares (3.42) to (3.43). In particular, the measure of dependence based on the entropy rate does not depend on an arbitrary choice of lag.[22]

[22] The measure of dependence based on the entropy rate does not depend on the distribution of the pair

For the above AR(1) model the entropy-based measure of dependence in the pair (X_t, X_{t-s}) is

$$
\begin{aligned}
D_s &= H(X_s) + H(X_{t-s}) - H(X_s, X_{t-s}) && (3.45)\\
&= -\frac{1}{2}\log(1-a^{2s}) && (3.46)
\end{aligned}
$$

for all integer $s > 0$ using (3.4), which is the population dependence measure used in the Hong and White [2004] test and explained in Section 3.1.4.2. The authors also propose to use the portmanteau statistic (3.7) as a larger (population) measure of dependence.

The correlation at lag s is a^s, so the Pearson correlation-based test for dependence at lag s compares a^s to zero. This measure of dependence decays exponentially in integer s when $a > 0$. The portmanteau statistic $\sum_{s=1}^{T} a^s$ can also be estimated from a sample and used as a measure of dependence.[23]

I plot these measures of dependence against one another for various coefficients a in the above AR(1) model in Figure 3.3. The Pearson correlation at lag 1 is larger than the entropy rate for $0 < a < 0.91$, and the portmanteau statistic version of the Pearson correlation for lags 1 and 2 is larger than the entropy rate for $0 < a < 0.99$. As the number of lags grows, the unique value of $a \in (0, 1)$ at which the Pearson portmanteau statistic equals the entropy rate is increasing toward 1. Dependence in the AR(1) model is characterised by autocorrelation, and hence the Pearson correlation is an effective measure of dependence for this model. Comparing (3.44) and (3.46), we see that in the AR(1) model the entropy-based measure of dependence equals the entropy rate-based measure of dependence when $s = 1$, because serial dependence in this AR(1) is fully characterised by the joint distribution of X_t and X_{t-1}. Furthermore the portmanteau statistic version of the entropy-based measure (3.7) is larger for every $a \in (0, 1)$ when $T > 1$ lags are used.

The magnitudes of the population measures of dependence shown in Figure 3.3 are indicative of the power of these tests in distinguishing $a > 0$ from $a = 0$ in the above AR(1) model. Therefore, we might expect the correlation-based test to have highest power, followed by the entropy-based test and then the entropy rate-based test. However, the sampling uncertainty in the estimators of these quantities differ. Even if the population quantity being estimated is larger, the sampling uncertainty could limit the power of the test in finite samples.

I present the results of a Monte Carlo experiment of the size and power of the test under the LZMA compression algorithm in Table 3.1. I perform the test 1,000 times, with nominal size 5%, on independent draws from an AR(1) process with standard normal errors and various choices of the lag coefficient parameter a. I compare the size and power to that of a test based

(X_t, X_{t-s}) for some arbitrary s, unlike the measure of dependence based on the entropy of the joint distribution of (X_t, X_{t-s}) or the correlation between X_t and X_{t-s}.

[23] The portmanteau statistic based on the correlation coefficient or joint entropy then depends on the arbitrary choice of T rather than s.

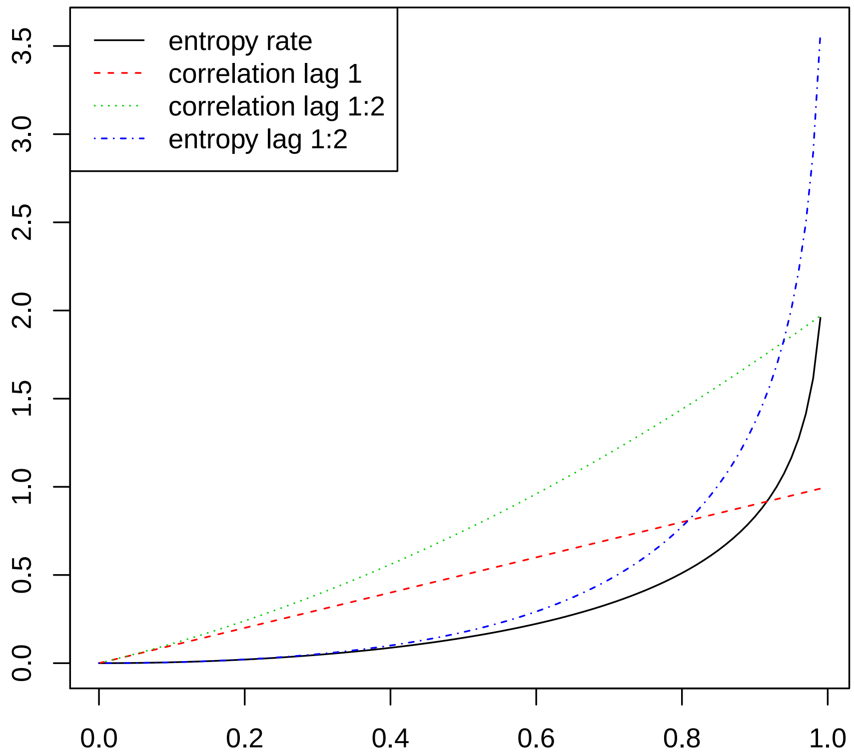


Figure 3.3: Population measures of dependence in the AR(1) plotted against the autocorrelation coefficient between 0 *and* 1.

on the lag-1 Pearson correlation coefficient. I perform the test using (3.22) with $\hat{F}$ set to the empirical cumulative distribution function and using a discretisation of 256 bins. The results are presented in Table 3.1 and show, as expected, that the Pitman-efficient test based on the lag-1 Pearson correlation coefficient achieves higher power.

3.3.2 Monte carlo experiments for discrete state markov chains

The entropy rate of a discrete time, finite state space, lag-1 Markov Chain $\{X_t\}$ with transition matrix P, state space S and stationary distribution $\pi = (\pi_j)$, which starts from its stationary

Table 3.1: Size and power experiments for tests of dependence in 1,000 AR(1) processes with unit error variance.

(a) The entropy rate-based test using finite-sample critical values.

n	a					
	0	0.0625	0.125	0.25	0.5	0.9
100	4.0	6.0	7.0	15.0	41.0	100.0
1000	7.0	3.0	8.0	40.0	100.0	100.0
10000	4.0	9.0	12.0	92.0	100.0	100.0
1e+05	4.0	7.0	35.0	100.0	100.0	100.0

(b) Lag-1 Pearson correlation test

n	a					
	0	0.0625	0.125	0.25	0.5	0.9
100	4.0	7.0	18.0	72.0	100.0	100.0
1000	6.0	48.0	99.0	100.0	100.0	100.0
10000	4.0	100.0	100.0	100.0	100.0	100.0
1e+05	2.0	100.0	100.0	100.0	100.0	100.0

distribution, is

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} H(X_1,\ldots,X_t) &= \lim_{t\to\infty} \frac{1}{t}\left[H(X_1) + H(X_2,\ldots,X_t|X_1)\right] \\
&= \lim_{t\to\infty} \frac{1}{t}\left[H(X_1) + H(X_2|X_1) + H(X_3,\ldots,X_t|X_2,X_1)\right] \\
&= \lim_{t\to\infty} \frac{1}{t}\left[H(X_1) + \sum_{s=2}^{t} H(X_s|X_{s-1})\right] \\
&= \lim_{t\to\infty} \frac{1}{t}\left[H(X_1) + (t-1)H(X_2|X_1)\right] \\
&= H(X_2|X_1). \qquad (3.47)
\end{aligned}$$

Consider the special lag-1 Markov Chain $\{X_t\}$ with state space $S = \{s_1,\ldots,s_K\}$ and symmetric transition matrix

$$P = \frac{1-p}{K}\begin{pmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 1 \end{pmatrix} + p\begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix},$$

where $p \in [0,1]$ is a parameter indexing the degree of serial dependence. Since P is symmetric, the stationary distribution of such a Markov Chain is uniform on S. Using (3.47), the entropy

Table 3.2: Size and power experiments for tests of dependence in 1,000 lag-1 Markov Chains with $|S| = 256$ *states.*

(a) The entropy rate-based test using finite-sample critical values.

p	0	0.0625	0.125	0.25	0.5	1
$h(\{X_t\})$	8.0	7.8	7.5	6.8	5.0	0.0
ρ	0.0	0.0	0.1	0.2	0.4	1.0
10	2.9	37.6	66.8	90.9	99.5	100.0
100	0.2	83.3	99.6	100.0	100.0	100.0
1000	2.8	75.4	100.0	100.0	100.0	100.0
10000	4.2	100.0	100.0	100.0	100.0	100.0
1e+05	4.4	100.0	100.0	100.0	100.0	100.0

(b) Lag-1 Pearson correlation test.

	0	0.0625	0.125	0.25	0.5	1
10	3.1	2.7	3.5	5.3		
100	6.2	9.9	21.0	66.4	99.5	
1000	5.5	49.0	96.7	100.0	100.0	
10000	4.9	100.0	100.0	100.0	100.0	
1e+05	6.2	100.0	100.0	100.0	100.0	

(c) Lag-1 Spearman rank correlation test.

	0	0.0625	0.125	0.25	0.5	1
10	4.1	4.1	3.6	5.3		
100	6.2	9.0	21.4	66.4	99.5	
1000	5.6	48.8	96.6	100.0	100.0	
10000	5.0	100.0	100.0	100.0	100.0	
1e+05	6.2	100.0	100.0	100.0	100.0	

rate of this particular Markov Chain is

$$
\begin{aligned}
H(X_2|X_1) &= \frac{1}{K}\sum_{x\in S} H(X_2|X_1 = x) \\
&= -\frac{1}{K}\sum_{x\in S}\left[(K-1)\frac{1-p}{K}\log\frac{1-p}{K} + \left(\frac{1-p}{K}+p\right)\log\left(\frac{1-p}{K}+p\right)\right] \\
&= -\left[(K-1)\frac{1-p}{K}\log\frac{1-p}{K} + \left(\frac{1-p}{K}+p\right)\log\left(\frac{1-p}{K}+p\right)\right] \qquad (3.48)\\
&\in [0, \log K],
\end{aligned}
$$

and we can define the *relative entropy rate* to be $H(X_2|X_1)/\log K \in [0, 1]$.

I simulate the hypothesis test using the LZMA compression algorithm over 1,000 draws of the above lag-1 Markov chain on $|S| = 256$ states. I use (nominal) critical values as close as possible to 5% upper quantiles of the discrete distribution of compression ratios tabulated in Appendix 3.B based on sample sizes as close as possible to the sequence lengths of the Markov chains being tested.[24] I vary the parameter p, which indexes the degree of serial dependence in the Markov chain being tested, from zero to one. From (3.48), these parameter values correspond to entropy rates from eight to zero, and hence asymptotic compression ratios of zero to one, respectively. I present the size and power performance in Table 3.2, where the columns index different values of the parameter p and the rows below the header index Markov chain samples of length $10, 100, \ldots, 10^5$. The first column of results in Table 3.2a show that the size of the test remains below its nominal 5% level and the other results show that the power approaches unity, although not monotonically, as the sample size grows and as the degree of dependence increases. The power scores in Table 3.2a for the test based on the entropy rate are generally better than the power scores in Tables 3.2b and 3.2c based on the Pearson and Spearman correlation coefficients respectively.

3.4 Conclusion

The statistic proposed in Section 3.2 gives the statistic due to Ryabko and Astola [2005] a natural interpretation as the degree of compressibility of a stochastic process, measured between 0 and 1. The asymptotic properties of the statistic for independent and identically distributed discrete state space stochastic processes under the LZ78 compression algorithm follow from a result in Jacquet and Szpankowski [2011].

For more modern algorithms like LZMA and BZ2, analytic asymptotic results are not available. However, numerical evidence presented in Section 3.2 suggests that the statistic using these compression algorithms also converges in the L_2-norm to a well defined limit distribution. For some purposes, normality might be a sufficiently good approximation. Nevertheless, I provide tables of the distribution under the serially independent null that could be used, for example, to pick critical values for a hypothesis test of serial dependence (at lag 1 or more).

The nominal size of a test for serial dependence at lag 1 or more based on the LZMA compression algorithm is controlled in Section 3.3. For shorter samples, the null distribution takes on discretely many values, so choosing an exact size is not possible in general. Nevertheless, the state space of the distribution of the test statistic under the null could be made more continuous by a procedure of padding shorter sequences with zeros or ones. In other words, given that the compression algorithms operate on sequences of blocks of 8 binary digits, the user has some freedom in encoding the original sample $x_1, \ldots, x_n$ from the time series $\{X_t\}$ into such blocks of binary digits. If the state space of X_t is binary, a natural choice would be to pack $x_{t+1}, \ldots, x_{t+8}$ into the t^{th} block of 8 binary digits, which would encode the original

[24] The distribution of compression ratios is discrete rather than continuous. As the sample size (i.e. sequence length) grows, the distribution of compression ratios approaches a continuous distribution as in (3.32), but in small samples this discreteness is noticeable. I choose the exact finite sample critical value to achieve a nominal size as close as possible to 5%, but the nominal size cannot exactly equal 5% due to the discreteness of the distribution of compression ratios.

sample $x_1, \ldots, x_n$ with $8(\lfloor n/8 \rfloor + 1)$ binary digits, but this is not the only choice. One could pack $x_t, 0, \ldots, 0$ into the t^{th} block of 8 binary digits, which would encode the original sample $x_1, \ldots, x_n$ with $8n$ binary digits.[25] The distribution of $\hat{\rho}$ applied to encoded sequences of length $8n$ can be expected to have a finer discrete state space than the distribution of $\hat{\rho}$ applied to encoded sequences of length $8(\lfloor n/8 \rfloor + 1) \leq 8n$, so that critical values for dependence tests can be chosen closer to a desired significance level. Test size could therefore be controlled more finely by studying this procedure further.

The statistic proposed here has power one asymptotically under all forms of dependence. However, the experiments in Section 3.3 suggest that correlation-based measures of dependence provide more power when the true dependence relationship is linear. The reverse is true in the lag-1 Markov Chain example, where the entropy rate-based measure performs better. Population measures of dependence for other examples of interest could be analysed to see where the entropy rate-based measure is most promising. An important case to consider is the class of location–scale models, for which correlation-based measures might be expected to have weak power when the mean component is small relative to the (persistence of the) variance component. A second important case to consider is a higher-order Markov Chain, where it might be possible to make the entropy rate large relative to the sum of lagged pairwise entropies used by Hong and White [2004].

3.A Sketch of the central limit theorem proof for Lempel-Ziv compression of iid Bernoulli stochastic processes

LZ78 passes through a sequences and parses it into 'phrases' where the next phrase is the shortest subsequence not yet seen as a previous phrase. Hence

$$11001010001000100 \mapsto (1)(10)(0)(101)(00)(01)(000)(100) \tag{3.49}$$

where parentheses denote phrases. Thus each new phrase is just an old phrase with one extra character appended. Jacquet and Szpankowski [1995] consider stationary Bernoulli data generating processes where $\mathbb{P}(X_t = 0) = p$ and $\mathbb{P}(X_t = 1) = 1 - p = q$ for all t. The special case $p = \frac{1}{2}$ is the *symmetric* Bernoulli model. The sequence in (3.49) could be seen to be a realisation of this symmetric Bernoulli stochastic process. It is easy to generalise the results for processes with a binary state space to processes with discrete state spaces bigger than $\{0, 1\}$. Jacquet

[25] Another alternative scheme would be to encode $x_{t+1}, \ldots, x_{t+8}$ into the t^{th} block of 8 binary digits and then if the encoded sequence length $8(\lfloor n/8 \rfloor + 1)$ is small enough for the discrete state space of $\hat{\rho}$ to be noticeable, zeros could be appended to the encoded sequence to lengthen it to the point where the discreteness of the state space of $\hat{\rho}$ is negligible.

and Szpankowski [2011] also generalise the results to the asymmetric Bernoulli model, where $p \neq \frac{1}{2}$.

A Digital Search Tree is a way of organising the phrases parsed by the LZ78 algorithm, starting from an empty phrase $\square$ called the root of the tree. For example, we can arrange the phrases in (3.49) to get the Digital Search Tree in Figure 3.4. The *internal path length* of this tree is the sum of the number of edges from the root to every node. The internal path length of the tree in Figure 3.4 is $2 \cdot 1 + 3 \cdot 2 + 3 \cdot 3 = 17$, which by construction is also equal to the length of the realisation (3.49) from the stochastic data generating process.

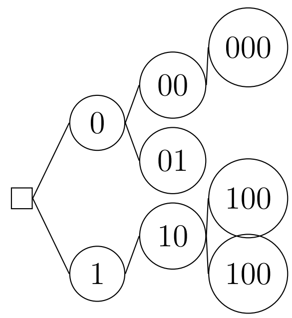


Figure 3.4: Digital Search Tree constructed from equation (3.49)

3.A.1 Link between L_m and M_n

Suppose that we apply LZ78 to compress a realisation of length n from the data generating process. Define M_n to be the number of phrases needed to parse the realisation and L_m to be the internal path length in the Digital Search Tree made up of the first m phrases. (We just saw that L_m is the same as the length of the sequence formed by concatenating the first m phrases together.) M_n and L_m are random variables, which depend on the particular realisation of the data generating process to which we are applying the LZ78 parsing algorithm. This section describes the relationship between L_m and M_n.

Consider a realisation of length $n = 10$ from the data generating process that we parse using LZ78. The maximum compression achievable is if the realisation consists entirely of zeros or entirely of ones, for example

$$0000000000 \mapsto (0)(00)(000)(0000) \tag{3.50}$$

which corresponds to $M_{10} = 4$. Maximal compression is achieved in this case because LZ78 needs only to represent each long subsequence of zeros as a reference to the previous subsequence and the extra 0 character. The minimum possible compression will occur when we observe the maximum number of phrases in the realisation of length 10, for example we could observe

$$0100011011 \mapsto (0)(1)(00)(01)(10)(11) \tag{3.51}$$

in which $M_{10} = 6$ is maximal. So $4 \leq M_{10} \leq 6$. The Digital Search Trees formed by the best and worst cases are given in Figures 3.5 and 3.6 respectively.

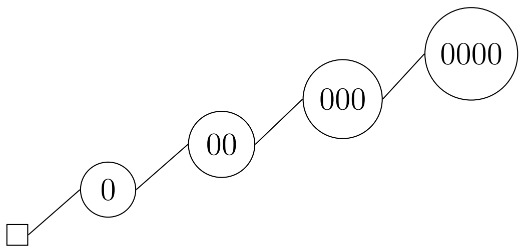


Figure 3.5: best case

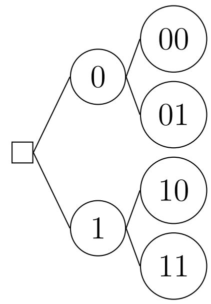


Figure 3.6: worst case

In Figure 3.5, $L_1 = 1, L_2 = 3, L_3 = 6$ and $L_4 = 10$. Therefore L_m looks like the integral of the sequence $1, 2, 3, 4, \ldots$ as a function of m, which is the integral of the $45°$ line, which ignoring remainders is $\int_0^m x dx = m^2$.

In Figure 3.6, $L_1 = 1, L_2 = 2, L_3 = 4, L_4 = 6, L_5 = 8$ and $L_6 = 10$. Therefore L_m looks like the integral of the sequence $1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, \ldots$ which ignoring remainder terms is $\int_0^m \log_2 x dx = m(\log_2 m - 1)$.

If 4 phrases are enough to parse the original sequence, then $L_4 = 10 = n$ and similarly if m phrases are sufficient to parse the original sequence then $L_m = 10 = n$. By definition of M_n therefore, $L_{M_n} = n$ and[26]

$$\{M_n = m\} \iff \{L_m = n\}. \tag{3.53}$$

Similarly, M_n and L_m are related by a *renewal equation*

$$\{M_n > m\} \iff \{L_m < n\}. \tag{3.54}$$

In the best case, the number of phrases m satisfies $m^2 = n$, ignoring remainder terms, so that $M_{10} = 4 \approx \sqrt{10}$. In the worst case, the number of phrases m satisfies $m(\log_2 m - 1) = n$ so that $M_{10} = 6 \approx 6.2$ where $6.2(\log_2 6.2 - 1) = 10$.

The renewal equation that characterises the relationship between L_m and M_n allows one to appeal to the following result from Billingsley [2008].

Theorem 1 (Billingsley [2008] Theorem 17.3)**.** *If L_n satisfies $\frac{L_n - \mathbb{E}L_n}{\mathbb{V}L_n} \xrightarrow{D} N(0,1)$ as $n \to \infty$ then M_n defined via the renewal equation $M_n = \max\{m : L_m \leq n\}$ satisfies*

$$\frac{M_n - 1/\mathbb{L}_n}{\sqrt{\mathbb{V}L_n}\,(\mathbb{L}_n/n)^{-3/2}} \xrightarrow{D} N(0,1)$$

as $n \to \infty$.

[26] More precisely,

$$M_n = \max\{m : L_m \leq n\} \tag{3.52}$$

Therefore, Jacquet and Szpankowski [1995] show that the normalised L_m has a limiting normal distribution as $m \to \infty$ in the knowledge that they can apply this theorem to find the limiting distribution for M_n as $n \to \infty$.

3.A.2 Finding the asymptotic distribution of L_m

To work out the asymptotic distribution of L_m, start by defining $L_m(u) = \mathbb{E}u^{L_m}$, the probability generating function of L_m.[27] Then we can show that

$$L_{m+1}(u) = u^m \sum_{k=0}^{m} \binom{m}{k} p^k q^{m-k} L_k(u) L_{m-k}(u) \tag{3.55}$$

for $m = 0, 1, \ldots$ by the following conditioning argument.

Let R_m be the number of phrases within the first m phrases that start with a zero. (Note $R_m \in [0, m]$.) Then $\mathbb{P}(R_m = k) = \binom{m}{k} p^k q^{m-k}$. Conditional on $R_m = k$, the internal path length of the Digital Search Tree where all phrases start with a zero is L_k and the internal path length of the Digital Search Tree where all phrases start with a one is L_{m-k}. So conditioning on $R_m = k$ we get $L_{m+1} = m + L_k + L_{m-k}$. Hence

$$\begin{aligned} \mathbb{E}u^{L_{m+1}}|R_m &= u^m \mathbb{E}(u^{L_k} u^{L_{m-k}}|R_m) &(3.56)\\ &= u^m \mathbb{E}(u^{L_k}|R_m)\mathbb{E}(u^{L_{m-k}}|R_m) &(3.57) \end{aligned}$$

where the last equality follows from the (conditional) independence of L_k and L_{m-k}.[28]

Now introduce the exponential generating function $L(z, u)$ of the sequence of probability generating functions $L_0(u), L_1(u), \ldots$ as $L(z, u) = \sum_{m=0}^{\infty} L_m(u) \frac{z^m}{m!}$ because we expect a simpler form for $L(z, u)$ than for $L_m(u)$, which turns out to be true [Jacquet and Szpankowski, 1995].

[27] Define $L_0(u) = 1$.

[28] L_k and L_{m-k} are independent because all the digits in the sequence are independent, so all the digits in all the phrases are independent.

Then

$$\begin{aligned}
\frac{\partial}{\partial z}L(z,u) &= \frac{\partial}{\partial z}\sum_{m=1}^{\infty} L_m(u)\frac{z^{m-1}}{(m-1)!} && (3.58)\\
&= \sum_{m=0}^{\infty} L_{m+1}(u)\frac{z^m}{m!} && (3.59)\\
&= \sum_{m=0}^{\infty}\left[u^m\sum_{k=0}^{m}\begin{pmatrix} m \\ k \end{pmatrix} p^k q^{m-k} L_k(u)L_{m-k}(u)\right]\frac{z^m}{m!} && (3.60)\\
&= \sum_{m=0}^{\infty}\sum_{k=0}^{m} L_k(u)\frac{(pzu)^k}{k!}L_{m-k}(u)\frac{(qzu)^{m-k}}{(m-k)!} && (3.61)\\
&= \left[\sum_{m=0}^{\infty} L_m(u)\frac{(pzu)^m}{m!}\right]\left[\sum_{m=0}^{\infty} L_m(u)\frac{(qzu)^m}{m!}\right] && (3.62)\\
&= L(pzu,u)L(qzu,u) && (3.63)
\end{aligned}$$

where (3.62) follows by definition of the Cauchy Product. So (3.63) gives a partial functional differential equation that Jacquet and Szpankowski [1995] solve for $L(z,u)$. The authors say that the solution to (3.63) is challenging because (3.63) appears in multiplicative form. Once we solve for $L(z,u)$, we can recover $L_m(u)$ by a Taylor expansion, and $L_m(u)$ gives the probability law of L_m and hence of M_n as discussed above.[29]

Begin the derivation by writing the Taylor expansion of $\log L(z,e^t)$ around $t=0$

$$\log L(z,e^t) = \log L(z,1) + \frac{\partial}{\partial t}\log L(z,e^t)\bigg|_{t=0} t + \frac{\partial^2}{\partial t^2}\log L(z,e^t)\bigg|_{t=0}\frac{t^2}{2} + O(t^3) \qquad (3.64)$$

from which the coefficient of t is the exponential generating function of the sequence of means $\mathbb{E}L_0, \mathbb{E}L_1, \mathbb{E}L_2, \ldots$ and the coefficient of $\frac{t^2}{2}$ is the exponential generating function of the sequence of variances $\mathbb{V}L_0, \mathbb{V}L_1, \mathbb{V}L_2, \ldots$. Denote these generating functions by $\tilde{X}(z)$ and $\tilde{V}(z)$ respectively. To prove that $(L_m - \mathbb{E}L_m)/\sqrt{\mathbb{V}L_m}$ converges in probability and moments to a standard normal distribution as $m\to\infty$, it suffices to show that the $O(t^3)$ term disappears for large m.By solving for $\tilde{X}(z)$ and $\tilde{V}(z)$ we obtain $\mathbb{E}L_m$ and $\mathbb{V}L_m$ for all m. The first and second derivatives of (3.63) can be used to find $\tilde{X}(z)$ and $\tilde{V}(z)$ respectively.[30]

[29]Note that the moment generating function of L_m evaluated at argument t is $L_m(e^t)$ and a Taylor expansion of the logarithm of this moment generating function around $t=0$ gives

$$\log L_m(e^t) = t\mathbb{E}L_m + \frac{t^2}{2}\mathbb{V}L_m + O(t^3).$$

[30]Note also that the leading term in the Taylor expansion is $\log L(z,1) = z$.

3.B Exact finite sample critical values

For LZMA compression ratios applied to length-n random sequences $\{X_i : i = 1, \ldots, n\}$ where $X_i \sim U\{1, \ldots, 2^8\}$ independently for each i we get the following upper percentiles:

$\log_2 n$	0.01	0.03	0.05	0.10
3	-2.875000	-3.000000	-3.000000	-3.000000
4	-1.500000	-1.500000	-1.500000	-1.562500
5	-0.812500	-0.812500	-0.812500	-0.812500
6	-0.437500	-0.437500	-0.437500	-0.437500
7	-0.250000	-0.250000	-0.250000	-0.250000
8	-0.136719	-0.136719	-0.136719	-0.136719
9	-0.074219	-0.074219	-0.074219	-0.074219
10	-0.040039	-0.040039	-0.041016	-0.041016
11	-0.024414	-0.024414	-0.024902	-0.024902
12	-0.017334	-0.017334	-0.017578	-0.017822
13	-0.014404	-0.014526	-0.014648	-0.014771
14	-0.013550	-0.013672	-0.013733	-0.013855
15	-0.013580	-0.013672	-0.013733	-0.013794
16	-0.013840	-0.013916	-0.013962	-0.014038
17	-0.013664	-0.013695	-0.013719	-0.013748
18	-0.013603	-0.013622	-0.013638	-0.013655
19	-0.013576	-0.013588	-0.013598	-0.013612

For BZ2 compression ratios applied to length-n random sequences $\{X_i : i = 1, \ldots, n\}$ where $X_i \sim U\{1, \ldots, 2^8\}$ independently for each i we get the following upper percentiles:

$\log_2 n$	0.01	0.03	0.05	0.10
3	-5.125000	-5.250000	-5.250000	-5.375000
4	-2.937500	-3.000000	-3.062500	-3.125000
5	-1.718750	-1.750000	-1.781250	-1.812500
6	-0.968750	-0.968750	-0.984375	-1.000000
7	-0.593750	-0.593750	-0.601562	-0.601562
8	-0.558594	-0.570312	-0.578125	-0.585938
9	-0.308594	-0.310547	-0.310547	-0.312500
10	-0.258789	-0.263672	-0.267578	-0.272461
11	-0.181152	-0.184082	-0.186523	-0.188965
12	-0.109863	-0.110840	-0.111816	-0.113281
13	-0.056396	-0.056885	-0.057251	-0.057739
14	-0.027344	-0.027527	-0.027649	-0.027832
15	-0.014587	-0.014679	-0.014771	-0.014862
16	-0.009399	-0.009460	-0.009506	-0.009567
17	-0.007141	-0.007187	-0.007225	-0.007271
18	-0.005882	-0.005909	-0.005936	-0.005966
19	-0.004990	-0.005081	-0.005152	-0.005211
20	-0.004340	-0.004404	-0.004455	-0.004520

The critical values in these tables do not converge on critical values of the standard normal distribution because of the bias and variance shown in Figure 3.1 that depend on the sequence length n and the normal distribution approximation errors for the largest sequence lengths shown in Figure 3.2.

Bibliography

Jean-Louis Arcand, Enrico Berkes, and Ugo Panizza. Too much finance? Technical Report WP/12/161, International Monetary Fund, 2012.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58 (2):277–297, 1991.

Manuel Arellano and Olympia Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.

Bank for International Settlements. Interest rate risk in the banking book. Consultative document, Bank for International Settlements, June 2015.

Bank of England. Technical report, Bank of England, 2011.

Robert J. Barro. Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2):407–443, 1991. doi: 10.2307/2937943. URL `http://qje.oxfordjournals.org/content/106/2/407.abstract`.

Robert J Barro and Jong-Wha Lee. Losers and winners in economic growth. In *Proceedings of the World Bank Annual Conference on Development Economics*, pages 267–97, 1994.

Robert J Barro and Jong Wha Lee. A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198, 2013.

Basel Committee on Banking Supervision. Principles for the management and supervision of interest rate risk. Technical Report July, Basel Committee on Banking Supervision, Basel, Switzerland, 2004.

Thorsten Beck, Ross Levine, and Norman Loayza. Finance and the sources of growth. *Journal of Financial Economics*, 58(1):261–300, 2000.

Juliane Begenau, Monika Piazzesi, and Martin Schneider. Banks' risk exposures. Working Paper 21334, National Bureau of Economic Research, July 2015. URL `http://www.nber.org/papers/w21334`.

Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

JR Blum, J Kiefer, and M Rosenblatt. Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, pages 485–498, 1961.

Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, 1998.

Stephen R Bond, Anke Hoeffler, and Jonathan RW Temple. GMM estimation of empirical growth models. Discussion Paper 3048, CEPR, 2001.

Clive Bowsher. On testing overidentifying restrictions in dynamic panel data models. *Economics Letters*, 77(2):211–220, 2002.

Francesco Caselli, Gerardo Esquivel, and Fernando Lefort. Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth*, 1(3):363–389, 1996.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. Wiley-interscience, 2006.

ST David, Maurice George Kendall, and Alan Stuart. Some questions of distribution in the theory of rank correlation. *Biometrika*, pages 131–140, 1951.

Angus Deaton and Alan Heston. Understanding PPPs and PPP-based national accounts. *American Economic Journal: Macroeconomics*, pages 1–35, 2010.

Matthias Doepke and Martin Schneider. Inflation and the redistribution of nominal wealth. *Journal of Political Economy*, 114(6):1069–1097, 2006.

Mathias Drehmann, Steffen Sorensen, and Marco Stringa. The integrated impact of credit and interest rate risk on banks: A dynamic framework and stress testing application. *Journal of Banking & Finance*, 34(4):713–729, 2010.

Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4): 379–406, 1996.

Darrell Duffie and Kenneth J. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12(4):687–720, 1999. doi: 10.1093/rfs/12.4.687. URL `http://rfs.oxfordjournals.org/content/12/4/687.abstract`.

Steven N Durlauf, Paul A Johnson, and Jonathan RW Temple. Growth econometrics. *Handbook of Economic Growth*, 1:555–677, 2005.

Federal Deposit Insurance Corporation. The banking crises of the 1980s and early 1990s: Summary and implications. Technical report, Federal Deposit Insurance Corporation, 1994.

Federal Deposit Insurance Corporation. Instructions for preparation of consolidated reports of condition and income: FFIEC 031 and 041. Technical report, Federal Deposit Insurance Corporation, 2014.

Edgar C Fieller, Herman O Hartley, and Egon S Pearson. Tests for rank correlation coefficients. I. *Biometrika*, 44(3-4):470–481, 1957.

Financial Accounting Standards Board. Statement of Financial Accounting Concepts No. 6: Elements of Financial Statements. Technical report, Financial Accounting Standards Board, 1985.

Financial Accounting Standards Board. Statement of Accounting Standards No. 91: Accounting for Nonrefundable Fees and Costs Associated with Originating or Acquiring Loans and Initial Direct Costs of Leases. Technical report, Financial Accounting Standards Board, 1986.

Financial Accounting Standards Board. Statement of Accounting Standards No. 115: Accounting for Certain Investments in Debt and Equity Securities. Technical report, Financial Accounting Standards Board, 1993.

Financial Accounting Standards Board. Statement of Accounting Standards No. 144: Accounting for the Impairment and Disposal of Long-Lived Assets. Technical report, Financial Accounting Standards Board, 2001.

Financial Accounting Standards Board. Proposed Statement of Financial Accounting Standards. Technical Report 1201-100, Financial Accounting Standards Board, Norwalk, CT, June 2004.

Financial Accounting Standards Board. Statement of Accounting Standards No. 159: The Fair Value Option for Financial Assets and Financial Liabilities. Technical report, Financial Accounting Standards Board, 2007.

Financial Accounting Standards Board. Statement of Accounting Standards No. 65: Accounting for Certain Mortgage Banking Activities. Technical report, Financial Accounting Standards Board, 2010.

Mark J Flannery and Christopher M James. The effect of interest rate changes on the common stock returns of financial institutions. *Journal of Finance*, pages 1141–1153, 1984.

Blaise Gadanecz. The syndicated loan market: Structure, development and implications. *BIS Quarterly Review, December*, 2004.

Christian Genest and Bruno Rémillard. Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369, 2004.

Mark Gertler, Nobuhiro Kiyotaki, et al. Financial intermediation and credit policy in business cycle analysis. *Handbook of Monetary Economics*, 3(3):547–599, 2010.

CW Granger, Esfandiar Maasoumi, and Jeffrey Racine. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis*, 25(5):649–669, 2004.

Refet S Gürkaynak, Brian Sack, and Eric Swanson. The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American Economic Review*, pages 425–436, 2005.

Refet S Gürkaynak, Brian Sack, and Jonathan H Wright. The US treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304, 2007.

Andrew G Haldane. Speech: Accounting for bank uncertainty, 2011.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054, 1982.

David Heath, F Delbaen, JM Eber, and P Artzner. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

Douglas Holtz-Eakin, Whitney Newey, and Harvey S Rosen. Estimating vector autoregressions with panel data. *Econometrica*, pages 1371–1395, 1988.

Yongmiao Hong and Halbert White. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. 2004.

James V Houpt and James A Embersit. A method for evaluating interest rate risk in US commercial banks. *Federal Reserve Bulletin*, (August):625–637, 1991.

Nazrul Islam. Growth empirics: A panel data approach. *The Quarterly Journal of Economics*, pages 1127–1170, 1995.

Philippe Jacquet and Wojciech Szpankowski. Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees. *Theoretical Computer Science*, 144(1):161–197, 1995.

Philippe Jacquet and Wojciech Szpankowski. Limiting distribution of Lempel-Ziv 1978 redundancy. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 1509–1513. IEEE, 2011.

Gabriel Jiménez, Steven Ongena, José-Luis Peydró, and Jesús Saurina Salas. Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking? 2009.

Gabriel Jiménez, Steven Ongena, José-Luis Peydró, and Jesús Saurina. Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications. *The American Economic Review*, 102(5):pp. 2301–2326, 2012. ISSN 00028282. URL http://www.jstor.org/stable/41724622.

Anil K Kashyap and Jeremy C Stein. What do a million observations on banks say about the transmission of monetary policy? *American Economic Review*, pages 407–428, 2000.

George G Kaufman. Measuring and managing interest rate risk: A primer. *Economic Perspectives*, 8(1):16–29, 1984.

David G Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–282, 1949.

Robert G King and Ross Levine. Finance and growth: Schumpeter might be right. *The Quarterly Journal of Economics*, pages 717–737, 1993.

Peter Kirschenhofer, Helmut Prodinger, and Wojciech Szpankowski. Digital search trees again revisited: The internal path length perspective. *SIAM Journal on Computing*, 23(3):598–616, 1994.

Ross Levine. Finance and growth: Theory and evidence. *Handbook of Economic Growth*, 1: 865–934, 2005.

Ross Levine and David Renelt. A sensitivity analysis of cross-country growth regressions. *The American Economic Review*, pages 942–963, 1992.

Ross Levine, Norman Loayza, and Thorsten Beck. Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, 46(1):31–77, 2000.

Norman V Loayza and Romain Ranciere. Financial development, financial fragility, and growth. *Journal of Money, Credit and Banking*, pages 1051–1076, 2006.

N Gregory Mankiw, David Romer, and David N Weil. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–437, 1992.

Patrick AP Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251, 1948.

Charles R Nelson and Andrew F Siegel. Parsimonious modeling of yield curves. *Journal of Business*, pages 473–489, 1987.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica*, pages 1417–1426, 1981.

Daniel A Nuxoll. Differences in relative prices and international differences in growth rates. *The American Economic Review*, pages 1423–1436, 1994.

James O'Brien. Estimating the value and interest rate risk of interest-bearing transactions deposits. *FEDS Working Papers*, (2000-53), 2000.

Monika Piazzesi and Martin Schneider. Interest rate risk in credit markets. *The American Economic Review*, 100(2):pp. 579–584, 2010. ISSN 00028282. URL `http://www.jstor.org/stable/27805062`.

Joseph P Romano and Andrew F Siegel. *Counterexamples in probability and statistics*. CRC Press, 1986.

Boris Ryabko and Jaakko Astola. Application of data compression methods to hypothesis testing for ergodic and stationary processes. In *International Conference on Analysis of Algorithms DMTCS proc. AD*, volume 399, page 408, 2005.

Boris Ryabko and Jaakko Astola. Universal codes as a basis for time series testing. *Statistical Methodology*, 3(4):375–397, 2006a.

Boris Ryabko and Jaakko Astola. Universal codes as a basis for nonparametric testing of serial independence for time series. *Journal of Statistical Planning and Inference*, 136(12):4119–4128, 2006b.

Ratna Sahay, Martin Cihak, Papa N'Diaye, Adolfo Barajas, Diana Ayala Pena, Ran Bi, Yuan Gao, Annette Kyobe, Lam Nguyen, Christian Saborowski, et al. *Rethinking financial deepening: Stability and growth in emerging markets*. Number 15-18. International Monetary Fund, 2015.

Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 (July & October):379–423 & 623–656, 1948.

Galen Sher and Giuseppe Loiacono. Maturity transformation and interest rate risk in large European bank loan portfolios. EcoMod2013 5442, EcoMod, 2013. URL `http://EconPapers.repec.org/RePEc:ekd:004912:5442`.

Galen Sher and Pedro Vitoria. An information-theoretic test for dependence with an application to the temporal structure of stock returns. *arXiv preprint arXiv:1304.0353*, 2013.

Gregory E Sierra. Can an accounting-based duration model effectively measure interest rate sensitivity? *Available at SSRN 1488884*, 2009.

Gregory E Sierra and Timothy J Yeager. What does the Federal Reserve's Economic Value Model tell us about interest rate risk at US community banks? *Federal Reserve Bank of St. Louis Review*, 86(6):45–60, 2004.

Lars E. O. Svensson. Estimating and interpreting forward interest rates: Sweden 1992-1994. IMF Working Papers 94/114, International Monetary Fund, September 1994. URL `http://ideas.repec.org/p/imf/imfwpa/94-114.html`.

Dag Tjostheim. Measures of dependence and tests of independence. *Statistics*, 28(3):249–284, 1996.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, pages 1–25, 1982.

Frank Windmeijer. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1):25–51, 2005.

David M Wright and James V Houpt. An analysis of commercial bank exposure to interest rate risk. *Federal Reserve Bulletin*, 82(2):115, 1996.

A. Wyner and J. Ziv. On entropy and data compression. *IEEE Trans. Inf.*, 1991.

A.D. Wyner and J. Ziv. The sliding-window Lempel-Ziv algorithm is asymptotically optimal. *Proceedings of the IEEE*, 82(6):872–877, 1994. ISSN 0018-9219. doi: 10.1109/5.286191.

Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Information Theory*, IT-23(3):337–343, 1977. ISSN 0018-9448.

Jacob Ziv and Abraham Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory*, 24(5):530–536, 1978. ISSN 0018-9448. doi: 10.1109/TIT.1978.1055934. URL `http://dx.doi.org/10.1109/TIT.1978.1055934`.