DISCUSSION PAPER SERIES

DP14115

THE LEVERAGE FACTOR: CREDIT CYCLES AND ASSET RETURNS

Alan M. Taylor and Josh Davis

ECONOMIC HISTORY

FINANCIAL ECONOMICS

MACROECONOMICS AND GROWTH

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

THE LEVERAGE FACTOR: CREDIT CYCLES AND ASSET RETURNS

Alan M. Taylor and Josh Davis

Discussion Paper DP14115
Published 12 November 2019
Submitted 05 November 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE LEVERAGE FACTOR: CREDIT CYCLES AND ASSET RETURNS

Abstract

Research finds strong links between credit booms and macroeconomic outcomes like financial crises and output growth. Are impacts also seen in financial asset prices? We document this robust and significant connection for the first time using a large sample of historical data for many countries. Credit boom periods tend to be followed by unusually low returns to equities, in absolute terms and relative to bonds. Return predictability due to this leverage factor is distinct from that of established factors like momentum and value and generates trading strategies with meaningful excess profits out-of-sample. These findings pose a challenge to conventional macro-finance theories.

JEL Classification: E17, E20, E21, E32, E44, G01, G11, G12, G17, G21, N10

Keywords: debt, leverage, Cycles, Asset Pricing, return predictability, asset allocation

Alan M. Taylor - amtaylor@ucdavis.edu
University of California, Davis and CEPR

Josh Davis - josh.davis@pimco.com
PIMCO

Acknowledgements
We thank our colleagues for helpful discussions. All errors are ours.

The Leverage Factor: Credit Cycles and Asset Returns*

Josh Davis† Alan M. Taylor‡

October 2019

Abstract

Research finds strong links between credit booms and macroeconomic outcomes like financial crises and output growth. Are impacts also seen in financial asset prices? We document this robust and significant connection for the first time using a large sample of historical data for many countries. Credit boom periods tend to be followed by unusually low returns to equities, in absolute terms and relative to bonds. Return predictability due to this leverage factor is distinct from that of established factors like momentum and value and generates trading strategies with meaningful excess profits out-of-sample. These findings pose a challenge to conventional macro-finance theories.

Keywords: debt, leverage, cycles, asset pricing, return predictability, asset allocation.

JEL classification codes: E17, E20, E21, E32, E44, G01, G11, G12, G17, G21, N10.

*We thank our colleagues for helpful discussions. All errors are ours.

†PIMCO (josh.davis@pimco.com).

‡Department of Economics and Graduate School of Management, University of California, Davis; PIMCO; NBER; and CEPR (amtaylor@ucdavis.edu; alan.taylor@pimco.com)).

1. Introduction

The financial crisis of 2008 and its aftermath served as a reminder that rapid credit growth conditions can profoundly influence macroeconomic and financial outcomes. Since then, a wave of academic research, often guided by macroeconomic history, has shown that periods of more rapid growth in economywide leverage are more likely to be followed by slower growth, deeper recessions, and a greater chance of financial crises (e.g., Mian and Sufi 2009; Gourinchas and Obstfeld 2012; Schularick and Taylor 2012; Jordà, Schularick, and Taylor 2013; Mian, Rao, and Sufi 2013).[1]

But if economywide leverage has strong predictive power for these macroeconomic outcomes, shouldn't it also predict dramatic shifts in major asset markets? Until now this corollary has not been formally tested, but we provide evidence that an economy's recent history of aggregate credit growth may be a robust predictor of future asset returns, with implications for the development of future models in macroeconomics and finance. Using long-run annual panel data for the advanced economies we show that credit booms are a negative predictor for real GDP growth looking forward, as has been noted (Mian, Sufi, and Verner 2017); but they are also a negative predictor for future returns to the major asset classes, equities and bonds. Higher credit growth predicts lower returns to equities in absolute terms, and for equities relative to bonds. In short, when a credit boom is maturing, the investor exposed to equities would face a more likely drawdown, while the investor exposed to bonds would not.

Our findings on returns constitute an additional layer in the puzzle for macroeconomic theory: Mian, Sufi, and Verner (2017) argue that the standard neoclassical model would normally associate more borrowing today with better growth prospects in the future. We now find that after credit booms the holders of equity claims also tend to do worse in the future, not better. If neither aggregate additional flows of output, nor aggregate additional returns to equity, compensate the agent-investor, the ex post outcome disappoints the ex ante desire to borrow. Our paper is thus complementary with an emerging view that financial crisis events—which are themselves strongly correlated with past credit growth—are also events that take the form of a re-pricing. See, for example, Krishnamurthy and Muir (2017) on the losses triggered by shifts in credit spreads, and Baron and Xiong (2017) on crashes in bank equity values. However, we emphasize the broader losses sustained in the entire equity market, and the predictable losses that occur even in normal times, not only financial crises. Even so, our results are consistent with proposed explanations based on mis-pricings and corrections to distorted, over-optimistic, or extrapolated beliefs (e.g., Greenwood and Hanson 2015; Bordalo, Gennaioli, and Shleifer 2018).

Our findings also pose a puzzle for finance theory and asset pricing models. Since past credit growth is in the information set of today's investors, one might think that, in an efficient market, it should not predict future asset returns. Yet our findings indicate that it does. Is this predictive power be too small to matter? We argue the contrary. In the second part of the paper we show

[1] Some works prior to the financial crisis also noted the warning signs in credit booms, a notable example being Borio and Lowe (2002).

that credit growth may be able to improve portfolio performance via asset allocation, and that the marginal value to an investor of our *leverage factor* (L) is as important as two widely-accepted predictive factors, momentum (M) and value (V), when judged by a variety of fit, profitability, and risk-reward metrics. In the context of the related literature, note that the leverage story we explore here has two important features. First, it is a global aggregate phenomenon, in other eras (going back to the 19th century) and other countries (not just recent U.S. experience), encompassing a wide range of policy, instititional, and financial regimes. Second, our L factor is predictive, known ex ante by the investor, and can be used to build portfolios, just like the M and V factors. Our findings are thus distinct, but related to, empirical work on impact of U.S. broker dealer leverage (e.g., Adrian, Moench, and Shin 2010) and they link to theoretical work on intermediaries, credit and asset pricing (e.g., Longstaff and Wang 2012; He and Krishnamurthy 2013).

The structure of the paper is summarized as follows, and the results are based on up to 150 years of asset return and credit data for a set of advanced economies provided by the annual panel data in the newly-expanded Jordà-Schularick-Taylor Macrohistory Database.

Section 2 presents a simple nonparametric test, binning annual observations using a crude binary high-low private credit growth indicator, based on a leverage factor (L) defined as the three-year lagged change in the ratio of total bank loans to GDP, denoted D3CREDGDP. Differences in future mean equity returns across the two bins are shown to be highly significant, statistically and quantitatively. For bonds there is virtually no effect. This finding uses no formal parametric model, so it avoids data-mining, but the concern arises that a lack of control might be biasing the result.

Section 3 therefore develops a linear regression model with controls, using in-sample local projections to predict future asset returns out to 3 years based on our leverage factor. The initial findings are robust and higher levels of the leverage factor (L), D3CREDGDP, today are associated with lower equity returns in the next 3 years, again with almost no effects for bonds. The result holds with and without controls for other well-established predictors of asset returns, the momentum factor (M) and the value factor (V), and also with macro controls. Thus the L factor is not simply capturing what is already contained in the M or V factors.

Section 4 explains that these forecasts may be able to generate a material improvement in terms of both fit and profitability, based on out-of-sample (OOS) performance. Whether compared to a null model, or models using M or V alone, or M and V together, the addition of the L factor improves OOS fit measured by mean square error. The addition of the L factor also enhances OOS profitability-weighted directional accuracy measured using weighted-AUC statistics, and the differences are statistically significant. This naturally raises the question of whether the L factor enhances investment performance in a meaningful way. We document that the OOS local projection return forecasts, fed into the machinery of a standard textbook Markowitz portfolio problem, can deliver meaningful gains in investment performance, as measured by increases in Jensen's α and Sharpe ratio. Based on this historical evidence, when a credit boom is maturing an investor should likely expect stocks to underperform bonds.

Section 5 concludes.

2. Simple Summary Statistics

We begin with a very basic in-sample test to keep assumptions to a minimum. In any given historical sample of data, we divide every country-year observation into high and low levels of recent credit growth at each point in time. We then ask whether that binary classification provides any statistically robust guide to observed differences in countries' asset market outcomes going forward.

Working at an annual frequency, we use the Jordà-Schularick-Taylor (JST) Jordà-Schularick-Taylor Macrohistory Database (Jordà, Schularick, and Taylor 2017; Jordà, Knoll, Kuvshinov, Schularick, and Taylor 2019).[2] This data source runs over the period 1870–2015 for as many as 17 advanced economies. It provides a private credit measure based on bank loans-to-GDP, and it now includes total real returns to four major asset classes (equities, housing, government bonds, and government bills). This study focuses mainly on post-1950 data, where data are believed to be of better quality, and on a subset of 14 countries which includes long series on important variables from which we later construct momentum and value factors as auxiliary controls.[3] Even so, we shall see here that the key findings also hold in a larger pooled sample that is expanded to include the late 19th century era of liberalized financial markets from 1870 to 1913.

For our empirical work, we need a suitable candidate credit boom signal. We use the change in the credit (bank loans)-to-GDP ratio over three years, denoted as D3CREDGDP_{it} in country i at time t. Researchers have used a variety of lag structures, but a window of about three to five years captures medium-term credit cycles and has already been shown to have a good predictive performance for macroeconomic outcomes (see, e.g., Mian, Sufi, and Verner 2017; Jordà, Schularick, and Taylor 2013).[4] For comparability, and to avoid data mining and optimization based on asset returns, we take this variable directly from these empirical macro settings and apply it here naïvely.

We use a simple approach. We classify every country-year observation into a high or low credit growth bin based on lagged credit growth, and then we see how asset performance after that year differs across the two bins. For a simple cut, we pool all countries and years since 1950 and we classify them based on whether D3CREDGDP_{it} in the current year is above or below the full sample median for all countries and years. We call these bins *hi* and *lo*.

The outcome measures are the real total return for equities (the broad benchmark index) and government bonds (approximately 10-year maturity) in the next year, where the return is in local-currency terms deflated by local-currency CPI inflation. (Later in the paper, when we construct portfolios, we shall switch to real returns in USD terms, but the findings here are robust to such a change in the units of the outcome measure.)

[2] The database can be found at `http://www.macrohistory.net/data/`.

[3] Canada, Portugal, and Switzerland were dropped due to missing dividend yield data in some years.

[4] The paper by Mian-Sufi-Verner uses the three-year change in total private-credit-to-GDP to study debt and the business cycle, with a focus on predicting future real GDP outcomes. We can replicate here their findings for GDP outcomes, so we adopt their 3-year change measure, except that our focus is on implications for asset returns. The papers from Jordà-Schularick-Taylor generally use five years of individual or averaged lag changes in bank-loans-to- GDP, but the results are not highly sensitive to this choice.

We ask: Are the credit bin indicators correlated with subsequent return performance? To assess that, in Table 1 we display the average annual rate of real total returns for equities and bonds in the next year. Look first at columns (1) and (2) for the postwar period 1950–2015. The first row shows the unconditional mean for all years, from 1950 to 2015, and all 14 countries. The mean real total return for equities (in local currency) is about 9%, and for government bonds is about 3%. The next two rows report mean returns for the *hi* and *lo* bins for each asset class. In column (1), the differences are dramatic for equities. For observations where D3CREDGDP$_{it}$ is below its historic median, equities return 11.5%. For observations where D3CREDGDP$_{it}$ is above its historic median, equities return just 6.8%. The final row shows that the differences in mean returns of 4.7% for equities is statistically significant. In column (2) the differences for bonds are much smaller, only 1.5%, and not statistically significant. Lastly, look at columns (3) and (4) which repeat the exercise pooling the postwar period 1950–2015 with the late nineteenth century period 1870–1913. Note that we exclude turbulent interwar sample, where inference is clouded by very volatile markets and a small sample size. Even so, the historically enlarged sample supports broadly the same conclusions. High versus low recent credit growth is associated with worse versus better equity return outcomes going forward, but smaller differences in bond return outcomes.

To sum up, our findings indicate that credit booms are followed by below-par returns to the two asset classes, but the differences are more pronounced for equities than for bonds. Our next step is to more rigorously test formal models of asset returns, and control for other factors to ensure the result is robust, before moving on to investment implications.

3. Regression Analysis of Predicted Returns

To investigate return forecasting ability, we apply the method of local projections (Jordà 2005). We use panel OLS regressions to forecast real total returns to the two assets using lagged three-year average credit growth D3CREDGDP$_{it}$, which we term the leverage factor, L, and other controls. As above, for now, we still measure real total return in local-currency terms deflated by local-currency CPI inflation and we use only the post-WW2 sample from 1950 to 2015 from now on.

Our exact model specification for the h-period cumulative total real return forecast is

$$\frac{\text{total real return index}_{i,t+h}}{\text{total real return index}_{i,t}} - 1 = \alpha_i^h + \beta^h \, \text{D3CREDGDP}_{it} + \mathbf{\Gamma^h} \, \mathbf{X}_{it} + e_{it} \,, \tag{1}$$

where the outcome variable refers to the return on equities or bonds in country i from year t, in percent. Country fixed effects are denoted α_i^h. The controls $\mathbf{X}_{it}$ include two established asset pricing factors, momentum (defined as one-year lagged total return) and value (defined as equity dividend yield or nominal bond yield); we refer to these factors as M and V, following convention.[5]

[5] On using M and V factors for return forecasts, see, e.g., Asness, Moskowitz, and Pedersen (2013). The literature is huge. Seminal works on momentum include, e.g., Poterba and Summers (1988) and Jegadeesh and Titman (1993); on value they include, e.g., Fama and French (1988), and Goyal and Welch (2003).

Table 1: *Binary credit boom indicator and future real total returns for equities and bonds*

	Real total returns (one year ahead, %)			
	(1)	(2)	(3)	(4)
Asset class	Stocks	Bonds	Stocks	Bonds
Sample	Post-WW2 only	Post-WW2 only	Pre-WW1 & Post-WW2	Pre-WW1 & Post-WW2
Mean	8.91^{***}	2.96^{***}	7.48^{***}	3.19^{***}
	(0.80)	(0.32)	(0.52)	(0.22)
D3CREDGDP Below Median	11.60^{***}	3.90^{***}	10.22^{***}	3.77^{***}
	(1.11)	(0.41)	(0.62)	(0.29)
D3CREDGDP Above Median	6.82^{***}	2.29^{***}	6.05^{***}	2.56^{***}
	(0.89)	(0.35)	(0.58)	(0.35)
Difference	-4.74^{*}	-1.53	-3.07^{**}	-1.37^{*}
	(2.01)	(0.80)	(0.95)	(0.58)
N	910	910	1526	1526

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Sample countries: AUS, BEL, DEU, DNK, ESP, FIN, FRA, GBR, ITA, JPN, NLD, NOR, SWE, USA.

The results are also robust to the inclusion of macro variables in the form of observed CPI inflation and real GDP per capita growth in country i at year t, but these are not reported here.

Given our focus on using credit to forecast future asset returns, the key coefficients of interest are the impulse responses β^h on the lagged credit growth variable D3CREDGDP$_{it}$. In the analysis, we ensure all of the controls are centered and standardized, so coefficients can be interpreted as the change in the forecast due to a +1 standard deviation (s.d.) shock in the corresponding regressor. For reference, in the sample used here D3CREDGDP$_{it}$ has a mean of 3.65% and an s.d. of 8.46% .

Figure 1 displays the impulse responses in graphical form, using the forecast model on the post-1950 advanced economy panel. In these charts, the solid line shows the response out to a 3-year horizon for a +1 s.d. shock to D3CREDGDP$_{it}$, with confidence intervals of $\pm$ 1 and $\pm$ 2 s.e. shown by dark and light shading, respectively.

In panel (a) the local projection for cumulative real returns to equities and bonds omits all controls, and in panel (c) it includes momentum and value controls as defined above. The choice of outcome units is not crucial; similar results, not shown here, are obtained for equity and bond returns expressed in U.S. dollars, and also as an excess over local 3-month T-bills or over 3-month U.S. Treasury bills. As a consistency check, the chart in panel (b) based on real GDP per capita growth, and no controls, shows that we can replicate on our data the main finding of Mian, Sufi, and Verner (2017): higher D3CREDGDP$_{it}$ today is a strong predictor of slower real GDP per capita growth in the future. But, as noted, our interest is whether similar patterns hold for asset returns and so henceforth we set to one side discussion of GDP outcomes.

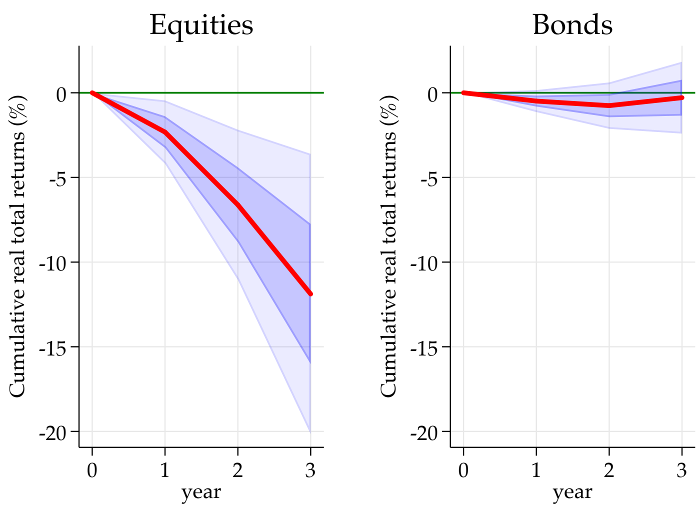
The key test here is whether the local projection coefficients β^h on D3CREDGDP$_{it}$ are nonzero. We are also interested in whether they have the expected sign. The null hypothesis is clearly rejected, and the coefficient on D3CREDGDP$_{it}$ is statistically significant at all years for equities but not for bonds. More importantly, the responses differ quantitatively by a large factor, and by an order of magnitude at three years. Again, in a more formal test, and with or without controls, we find that larger credit booms measured by D3CREDGDP$_{it}$ go hand in hand with depressed returns in equities relative to bonds. Over three years, given a +1 s.d. change in D3CREDGDP$_{it}$, forecast real total equity returns drop by an average of 300 to 400 bps per year, but forecast real total bond returns are virtually flat.

These results provide further support for an approach to return prediction based on a leverage factor like D3CREDGDP. In addition, since the results are robust to controls, we can be reassured that this leverage factor (L) is not simply capturing, in a different form, established sources of predictive ability already captured in well-known factors like momentum (M), or value (V).

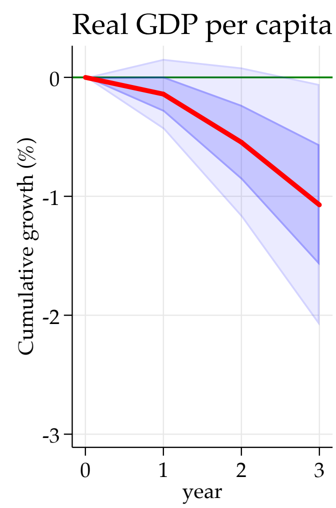
To sum up, we see that, from a forecasting perspective, the leverage factor contains distinct predictive information on the margin, and in what follows we seek to find out how valuable that information might be.

Figure 1: *Predicted response of cumulative real total return for equities and bonds, and real GDP per capita, to a +1 s.d. change in D3CREDGDP$_{it}$, in-sample local projection*

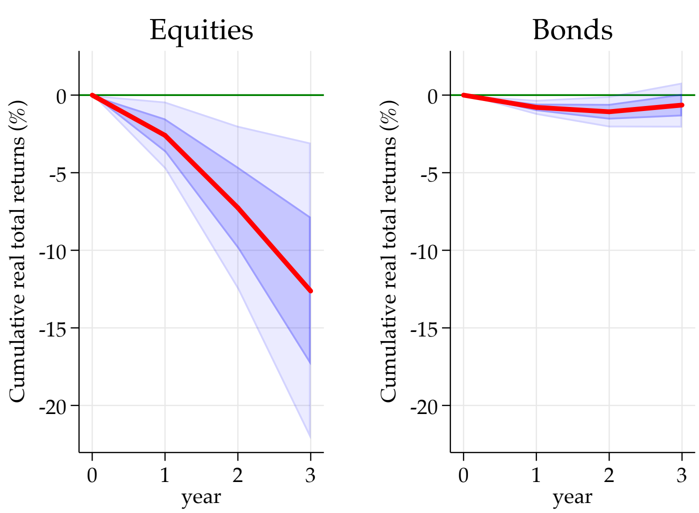
(a) *Returns, no controls*



(b) *GDP per capita, no controls*



(c) *Returns, with momentum and value controls*



Notes: See text. Confidence intervals of $\pm$ 1 and $\pm$ 2 s.e. shown by dark and light shading, respectively. Note that Panel (b) uses GDP per capita outcomes to replicate the Mian, Sufi, and Verner (2017) on our sample as a consistency check.

4. Trading on the Leverage Factor

The results so far suggest that returns to one major asset class, equities, can be forecast in absolute terms and relative to bonds using lagged credit growth information. We view this as indicative of the existence of a new ex ante asset pricing factor, the leverage factor, which we denote L. However, this does not yet fully document how important or significant this new factor is as a driver of asset returns, either on its own or as compared to other well established factors. Nor does it address the natural question as to whether the information in the L factor would generate a significant excess returns to an investor who used the information to suitably tailor their portfolio allocations.

The purpose of this final section of the paper is to address these open questions. We proceed using various empirical tests, and basing our work on a single, standard, global portfolio allocation problem to equities and bonds. In the back of our mind we have a global equal-weight 60-40 as a simple benchmark portfolio. Our forecasting results above suggest that investment performance gains can be made relative to the benchmark by an asset allocation strategy that tilts away from countries in high credit boom states and toward countries low credit boom states. The idea is to test this empirically.

We apply a model of asset returns like those above to set optimal portfolio allocation in each country based the state of the credit cycle and other factors. We allow the allocation to equities and bonds to vary over time and over countries, with an offsetting long or short position in each country's three-month bills as necessary. All returns to all positions are now measured in USD on an annual basis, and the portfolio is rebalanced each year based on rolling returns forecast in that year using an expanding window of lagged information. That is, we apply a standard out-of-sample recursive backtest using the model's annual return forecasts. We then solve a Markowitz stock/bond allocation problem for each country.

We use regression models as above to make rolling forecasts of a vector of excess returns for each country and each year for equities and bonds relative to three-month bills. We focus on forward-looking total returns at a holding period horizon of one year $h = 1$. In the model, we can include as controls, as before, the three candidate factors: a momentum factor (M) measured by the lagged USD return, a value factor (V) exactly as before, and the new leverage factor (L) which is still defined as D3CREDGDP_{it}.

In detail, using the annual data from 1950 up to 1969 as a training sample, from 1970 on we make a rolling one-step ($h = 1$) forecast of annual excess returns for the vector of the two assets, equities and bonds, in 14 countries, using the fitted values from panel country fixed effect regressions for each asset class at the one-year horizon, where each asset class forecast is given by the predicted value

$$\underbrace{\hat{r}_{i,t+1}}_{\text{forecast excess return}} = \underbrace{\hat{\alpha}_i^1}_{\text{null}} + \underbrace{\hat{\beta}^1 \,\text{D3CREDGDP}_{it}}_{\text{L}} + \underbrace{\hat{\gamma}^1 \,\text{Momentum}_{it}}_{\text{M}} + \underbrace{\hat{\delta}^1 \,\text{Value}_{it}}_{\text{V}} . \tag{2}$$

Combining the resulting equity and bond return predicted values as a vector, we denote the stacked actual returns by $\mathbf{r_{it}}$, and the forecasts by $\mathbf{\hat{r}_{it}}$.

We can implement this forecasting model in various ways; we can include all controls, but we can also look at nested alternatives which use only some controls, or even a null model with no controls. As the bracketed labels denote, when using no controls (apart from country fixed-effects) we will refer to the forecast as the null model. When using leverage factor, we refer to the L model. When using momentum factor, we refer to the M model. When using value factor, we refer to the Vmodel. When using momentum and value factors, we refer to the MV model. When using all three factors, we refer to the MVL model.

We performed the rolling regressions with both unconstrained and constrained coefficients, with similar results. In the more conservative constrained results shown below, the coefficients on momentum and value in the rolling regressions were restricted to conventional positive values and coefficient on credit was restricted to negative values. We did this to ensure that the signal coefficients would not invert arbitrarily in some windows and create spurious support for the assumed model or artificially enhance performance.

For reference, Table 2 reports unconstrained in-sample panel OLS regression forecast results for the maximal 1950 to 2015 window. The constraints would not be rejected here: V (+) and L (−) have statistically significant coefficients of the correct sign, and the M coefficients are not statistically different from zero. That said, M is usually a better signal at higher frequency than annual, and it still makes a statistically significant contribution in some but not all sample windows in the recursive estimation.

Once we have the rolling forecasts using the out-of-sample constrained regressions, we take the covariance matrix of pooled excess returns $\mathbf{\hat{V}}$, again from the lagged expanding window of the raw data, and we then compute the optimal tangent portfolio weights of the Markowitz problem, given by $\mathbf{w}_{it} = \mathbf{\hat{V}_{it}^{-1}\hat{r}_{it}}$. The weights $\mathbf{w}_{it}$ are then used in each country i, in each year t, this portfolio is then used as the asset allocation rule.

We then apply simple global weights so that in the overall allocation is always $1/N$ to each country, but the model-based optimal weights for that country are used to set the stock-bond tilt which attains a realized USD return $\psi_{it} = \mathbf{w}_{it}'\mathbf{\hat{r}}$ for that country. Under equal country weights, the global portfolio's realized USD return is then given by $\bar{\psi}_t = \frac{1}{N}\sum_{i=1}^{N}\psi_{it}$. The performance of realized returns can then be analyzed and compared across different strategies—i.e., for different subsets of signals used as control variables. Note that these portfolios are unconstrained, so both long and short positions are permitted, in contrast to say a simple long-only strategy like 60-40 in equities and bonds.

We present four sets of diagnostics statistics to show that the L factor matters for asset pricing, on its own terms, as well as in comparison to the more established M and V factors. Whenever the L factor is added to a model it increases both the OOS fit of the model return forecasts and the OOS performance of the Markowitz portfolios constructed with those forecasts. These contributions are statistically and quantitatively significant, as we now show.

Table 2: *Excess return forecast models for equities and bonds, full post-WW2 sample*

(a) *Excess return to equities*

	Excess returns in USD terms (one year ahead, %)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	7.66***	7.05***	7.67***	7.06***	7.66***	7.23***	7.73***	7.23***
	(0.83)	(0.84)	(0.84)	(0.84)	(0.82)	(0.82)	(0.82)	(0.82)
D3CREDGDP		-2.74**		-2.70**		-2.46**		-2.47**
		(0.92)		(0.91)		(0.91)		(0.91)
Momentum			1.22	0.89			-0.50	-0.62
			(1.05)	(1.06)			(1.12)	(1.13)
Value					5.22***	4.58***	5.10***	4.82***
					(0.92)	(0.94)	(0.99)	(1.01)
Observations	896	854	882	854	896	854	882	854
R^2	0.005	0.020	0.008	0.021	0.044	0.048	0.039	0.049

(b) *Excess return to bonds*

	Excess returns in USD terms (one year ahead, %)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	1.87***	1.85***	1.86***	1.85***	1.87***	1.83***	1.85***	1.83***
	(0.31)	(0.32)	(0.32)	(0.32)	(0.31)	(0.32)	(0.31)	(0.32)
D3CREDGDP		-0.77*		-0.85*		-0.70*		-0.78*
		(0.34)		(0.35)		(0.34)		(0.35)
Momentum			-0.69	-0.76			-0.60	-0.67
			(0.40)	(0.40)			(0.40)	(0.41)
Value					0.94*	0.90*	0.87*	0.82
					(0.42)	(0.43)	(0.43)	(0.43)
Observations	896	854	882	854	896	854	882	854
R^2	0.006	0.014	0.012	0.020	0.016	0.022	0.020	0.027

Notes: Standard errors in parentheses clustered by country and year. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Unconstrained OLS panel regression with country fixed effects on full 1950–2014 sample.
Sample countries: AUS, BEL, DEU, DNK, ESP, FIN, FRA, GBR, ITA, JPN, NLD, NOR, SWE, USA.

4.1. *MSE* ratio statistics

Figure 2(a) begins by comparing the mean-squared error (*MSE*) of each model's OOS forecasts to the *MSE* of the null model, over all observations in the panel regression, for equities and bonds.

We use this as a simple yardstick to gauge whether the leverage factor L adds any explanatory power on its own, compared to that added by the well-established momentum and value factors, M and V. We also want to know if, when L is added to the MV model, it also makes any further distinct contribution to model performance. Thus, as in the case of other performance statistics which follow, we take the null model and we compare it against the M, V, L, MV, and MVL models.

Starting with equities, the *MSE* ratio statistics in the left panel of Figure 2(a) are persuasive. We see that, as expected, the M factor on its own has little explanatory power at the annual frequency, and reduces *MSE* by only 1.2% relative to the null (0.988). But the V factor on its own contributes more to fit, and reduces *MSE* by 4.3% relative to the null (0.957). These are both established factors is the asset pricing literature. Can the L factor achieve anything comparable? Yes it can. The L factor on its own contributes to fit, and reduces *MSE* by 3.4% relative to the null (0.966). This fit improvement is much more than that of the accepted M factor, and close to that of the equally well-established V factor. But is the L factor merely picking up what the M and V factors already detect? This possibility could be entertained. Perhaps the credit boom indicated by L factors is also a time when M and V factors also shift and deliver essentially the same information? This is not the case. Figure 1 also shows that although the MV factors jointly reduce *MSE* by 4.6% relative to the null (0.954), adding the leverage factor allows the MVL model to achieve the maximal reduction of *MSE* by 7.4% relative to the null (0.926), that is, by a further 2.8%

Turning to bonds, the *MSE* ratio statistics in the right panel of Figure 2(a) are very similar in terms of ranking and magnitudes. The M factor is here quite weak, but the V factor reduces *MSE* by more here, by 4.9% relative to the null (0.951). The L factor on its own again contributes to fit, and reduces *MSE* by 3.3% relative to the null (0.967). The MV factors jointly reduces *MSE* by 4.6% relative to the null (0.954), adding the leverage factor allows the MVL model to achieve the maximal reduction of v by 7.1% relative to the null (0.929), that is, by a further 2.5%

Taken together these results give strong support to the idea that L factor matters for asset pricing. When used to forecast asset returns, its contribution in terms of predictive ability is nontrivial, either on its own or when in combination with M and V factors. It reduced the model's annual forecast error by more than the M factor, and almost as much as the V factor. If these factors are broadly accepted based on this kind yardstick, then the L factor might be taken seriously too.

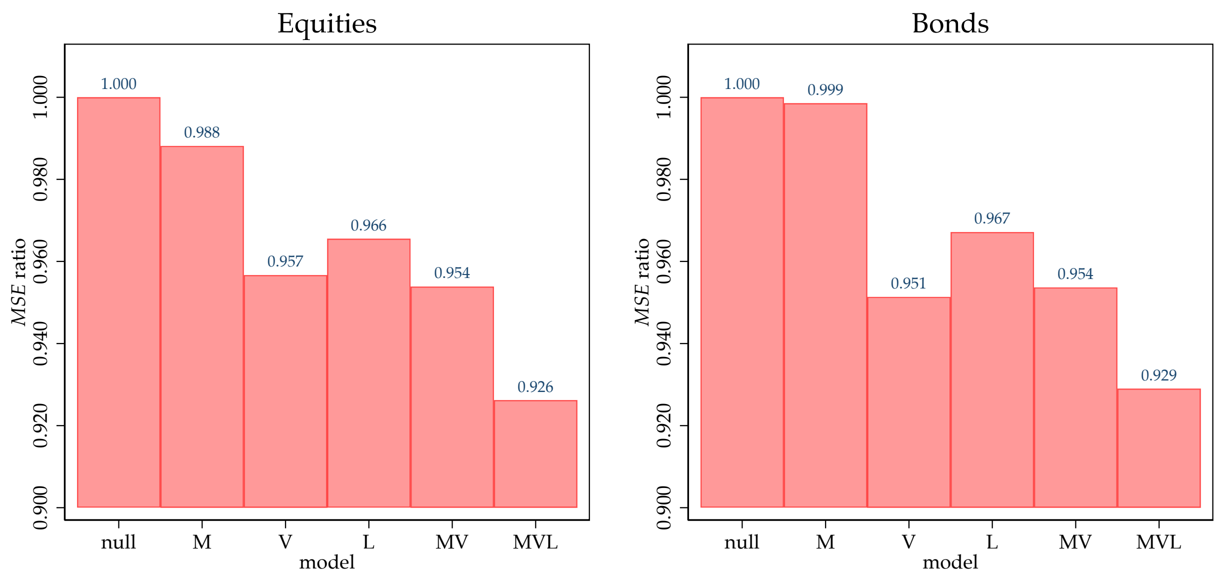
4.2. Weighted AUC^* statistics

We now ask whether performance can be assessed by more than the standard measure of statistical fit *MSE*, and judged by a more economically based criterion.

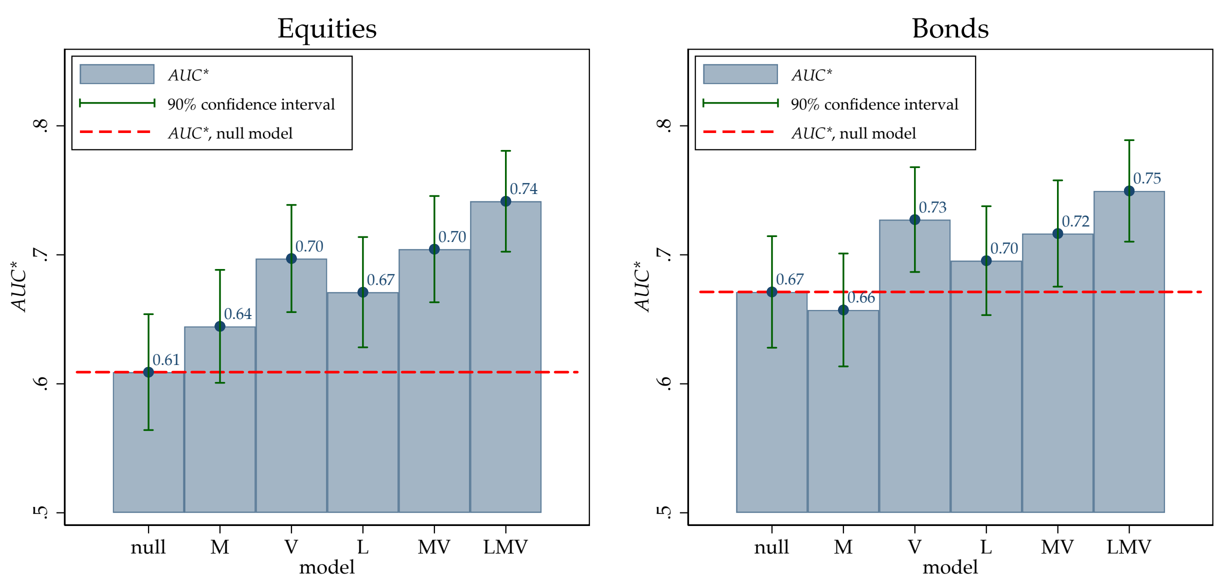
We borrow from Jordà and Taylor (2011), who rightly note that from a profit standpoint, the

Figure 2: *Excess return forecast evaluation, out-of-sample MSE ratio and AUC* statistics*

(a) *MSE ratio of model relative to the null*



(b) *AUC* of models versus benchmark 60-40 and null*



Notes: See text.

investor does not care in the last about MSE, they care about returns. It is no good if the model has great fit most of the time when small profits and losses are at stake, but then makes rare but large directional mistakes when big profits and losses are at stake. Jordà and Taylor (2011) show how to import into finance tools from binary classification theory and machine learning to better gauge the investment performance of any strategy, specifically a statistic known as the area under the curve or AUC, which comes from the ROC curve, or receiver operating characteristic curve and its variants.

Jordà and Taylor (2011) advocate that in finance applications, the classification of each instance receive a weight w_t proportional a gain or loss associated with that observation. This type of weighting was proposed earlier by Fawcett (2006) and applied in Berge, Jordà, and Taylor (2011) to the analysis of FX returns. Here we consider the mean of the return as the "hurdle rate" and reward the classifier with weight based on return relative to mean. The area under this weight-adjusted curve or $AUC^\star$, inherits all the same properties and methods of inference as the simple AUC statistic. The weighted ROC* curve traces out weighted true positives against weighted true negatives, It is straightforward to show how these map into to the gain-loss statistics proposed by Bernardo and Ledoit (2000), as explained by Jordà and Taylor (2011), so this is a way to judge a strategy on it upside-downside profit capture. See Appendix A for further details.

We again implement the test using all observations in the panel regression, for equities and bonds separately. On this metric, the statistics reported in Figure 2(b) are also persuasive. Looking at equities, in the left panel, the null model achieves an $AUC^\star$ of 0.61. Adding any single factor to the model improves $AUC^\star$, and in a statistically significant way with V and L. M achieves 0.64, V achieves 0.70, and L achieves 0.67. This is the same performance ordering seen with the MSE ratios. However, judged by $AUC^\star$, the MV model adds little. Its $AUC^\star$ is 0.70, the same as V alone. In contrast, adding the leverage factor here, MVL has $AUC^\star$ equal to 0.74, and the change versus MV is statistically significant.

For bonds, in the right panel, similar findings appear. The null model achieves an $AUC^\star$ of 0.67. Here, among single factor models, M is actually marginally worse than the null and achieves 0.66, but V and L perform better than the null at 0.73 and 0.70, with the difference for V being statistically significant. The MV model achieves only 0.72, but adding L the MVL model achieves 0.75, the best of all the models shown, and the increment versus MV is close to statistically significant.

Summing up, in the case of both stocks and bonds, adding the L factor to the null or to the MV model again results in meaningful performance gain, now measured by the $AUC^\star$ statistic. The leverage factor makes a difference.

4.3. Jensen's α statistics

We now turn to investment performance of the models judged by the annual returns to the Markowitz portfolios. A well-established gauge of whether any given method of portfolio construction leads to improvement in performance is the measure known as Jensen's α. This is given by the intercept in a regression of the portfolio excess return on some benchmark excess return.

In our application, we perform the following estimation on annual portfolio returns,

$$r_{it}^{m} = \alpha^{m} + \beta^{m} r_{it}^{B6040} , \tag{3}$$

where m is one of the set of models considered here, {B6040, null, M, V, L, MV, MVL}, and where we add in here as the benchmark (B6040) the static 60-40 global portfolio, which is a fixed-weight allocation of 60% equity, 40% bonds, with each country equally-weighted.

In our setting, the two key hypothesis tests are: do models deliver meaningful gains relative to the benchmark, $\alpha^{m} > 0 \equiv \alpha^{B6040}$; and do they deliver meaningful gains relative to the null, $\alpha^{m} > \alpha^{null}$.

On this metric, the statistics reported in Figure 3(a) are again supportive. First, even the null model outperforms the Benchmark 60-40. It uses no signals to make a forecast, but just computes a Markowitz portfolio based on historical first and second moments of returns, but this on its own delivers α^{null} that is positive and statistically significant, and equal to 188 bps.

Adding the M, V, and L factors, increases performance even more with α^{m} rising as we move through these signals, from 302 bps, to 368 bps, to 455 bps. Clearly, the established M and V signals live up to their reputation as useful return predictors, but on this metric L on its own does even better than either one of these traditional asset pricing factors. Finally, for multiple signals, the MV model achieves 391 bps, but adding the L factor makes a big improvement, and delivers 585 bps.

All told, looking at the confidence intervals for these estimates of α^{m} for all models, it is striking that all beat the Benchmark 60-40 by a statistically significant margin, but only models including the L factor beat the null by a statistically significant margin. The leverage factor again looks to be at least as useful as its better known rivals.

4.4. Sharpe ratio statistics

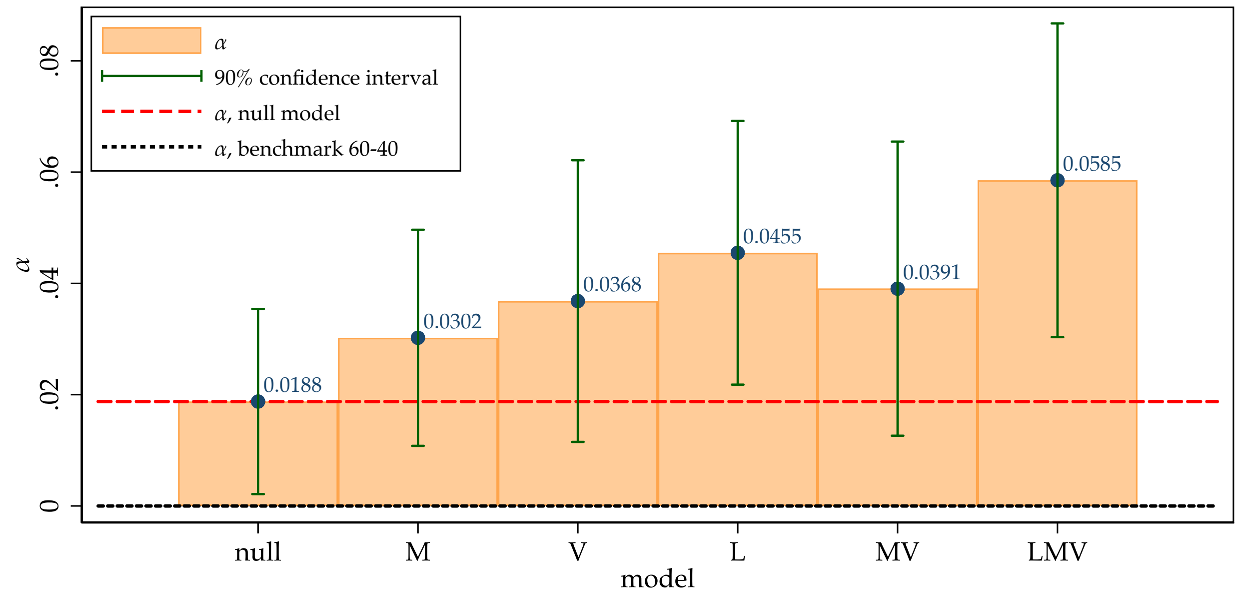
Our final test is based on the best known performance statistic of all, the Sharpe ratio. For any portfolio m, this is defined as the mean of the excess return $\mu(r_{it}^{m})$ divided by its standard deviation $\sigma(r_{it}^{m})$, for all of the model portfolios m considered here, {B6040, null, M, V, L, MV, MVL}.

On this metric, the statistics reported in Figure 3(b) are also favorable. The Benchmark 60-40 achieves a Sharpe ratio of 0.45. This is substantially bettered even by the null model, which attains 0.62. Moving through the models with only one signal, M achieves 0.70, V achieves 0.82, and L achieves 0.78. So among individual signals L performs quite well. The MV model achieves a Sharpe ratio of 0.83, but adding the leverage factor to this the MVL model achieves a Sharpe ratio of 0.94, a nontrivial gain of +0.11 Sharpe units on the margin relative to the established MV signals.

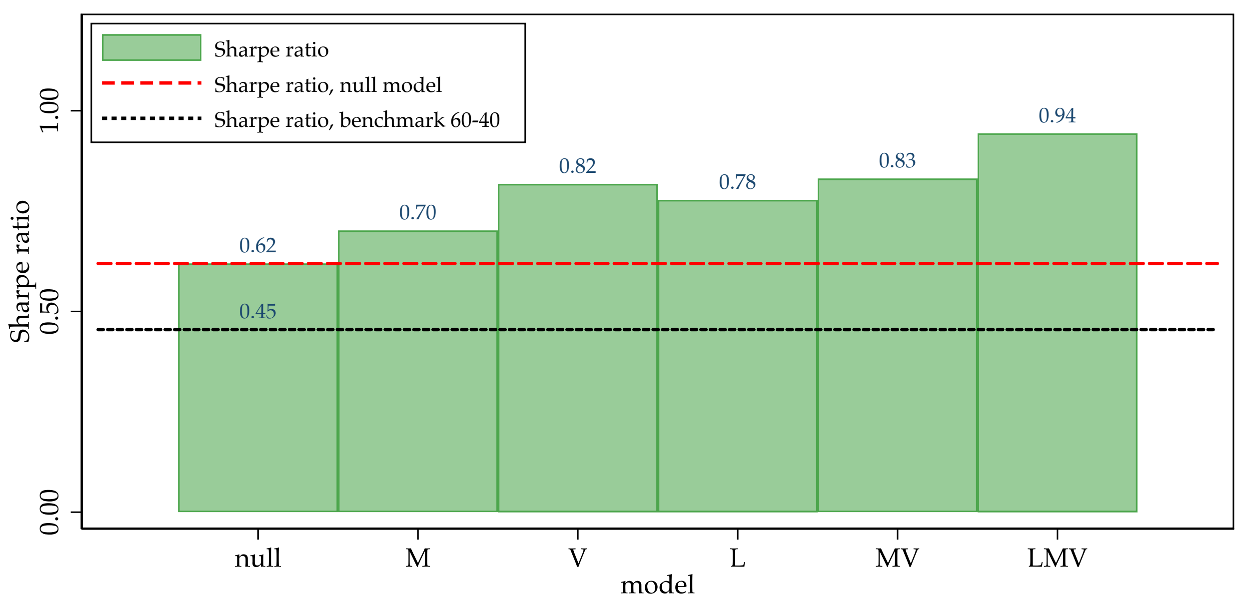
Again the leverage factor does well in this contest. It delivers meaningful performance gains as measured by Sharpe ratio, when the three signals are compared individually, or when L is added to an MV model.

Figure 3: *Portfolio performance evaluation, out-of-sample Jensen's α and Sharpe ratio statistics*

(a) *Jensen's α of models versus benchmark 60-40 and null*



(b) *Sharpe ratio of models versus benchmark 60-40 and null*



Notes: See text.

5. Conclusion

We set out to explore whether leverage cycles leave a signature on asset returns and whether these patterns have predictive value for investors. Already, a wave of research has shown that credit boom and bust episodes deeply influence future macroeconomic outcomes, so it would be surprising if the same were not true of financial markets.

The evidence supports the hypothesis. Credit boom periods tend to be followed by equity return underperformance in the near future. We find that credit growth signals can be a useful input for a tactical asset allocation strategy, alongside such tried and tested signals as momentum and value.

Further research is needed, but our findings suggest that accounting for the role of credit booms and busts could be as important for asset pricing studies as it has become for mainstream macroeconomics. The evidence also adds to the growing set of macro-finance puzzles which standard models struggle to explain.

References

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. Value and Momentum Everywhere. *Journal of Finance* 68(3): 929–85.

Bernardo, Antonio E., and Olivier Ledoit. 2000. Gain, Loss, and Asset Pricing. *Journal of Political Economy* 108(1): 144–72.

Berge, Travis, Òscar Jordà, and Alan M. Taylor, 2011. Currency Carry Trades. In *NBER International Seminar on Macroeconomics 2010* edited by R. Clarida and F. Giavazzi. Chicago: University of Chicago Press, pp. 357–388.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2018. Diagnostic Expectations and Credit Cycles. *Journal of Finance* 73(1): 199–227.

Borio, Claudio, and Philip Lowe. 2002. Assessing the Risk of Banking Crises. *BIS Quarterly Review* (December): 43–54.

Fama, Eugene F., and Kenneth R. French. 1988. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22(1): 3–25.

Fawcett, Tom. 2006. ROC Graphs with Instance-Varying Costs. *Pattern Recognition Letters* 27(8): 882–91.

Goyal, Amit, and Ivo Welch. 2003. Predicting the Equity Premium with Dividend Ratios. *Management Science* 49(5): 639–54.

Gourinchas, Pierre-Olivier, and Maurice Obstfeld. 2012. Stories of the Twentieth Century for the Twenty-First. *American Economic Journal: Macroeconomics* 4 (1): 226–65.

Greenwood, Robin, and Samuel G. Hanson. 2015. Waves in Ship Prices and Investment. *Quarterly Journal of Economics* 130(1): 55–109.

He, Zhiguo, and Arvind Krishnamurthy. 2013. Intermediary Asset Pricing. *American Economic Review* 103(2): 732–70.

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48(1): 65–91.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. The Rate of Return on Everything, 1870–2015. *Quarterly Journal of Economics* 134(3): 1225–98.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. When Credit Bites Back. *Journal of Money, Credit and Banking* 45(s2): 3–28.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. Macrofinancial History and the New Business Cycle Facts. In *NBER Macroeconomics Annual 2016*, vol. 31, edited by M. Eichenbaum and J. Parker. Chicago: University of Chicago Press.

Jordà, Òscar, and Alan M. Taylor. 2011. Performance Evaluation of Zero Net-Investment Strategies. NBER Working Paper 17150.

Longstaff, Francis A., and Jiang Wang. 2012. Asset Pricing and the Credit Market. *Review of Financial Studies* 25(11): 3169–3215.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. Household Balance Sheets, Consumption, and the Economic Slump. *Quarterly Journal of Economics* 128(4): 1687–1726.

Mian, Atif, and Amir Sufi. 2009. The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *Quarterly Journal of Economics* 124(4): 1449–96.

Poterba, James M., and Lawrence H. Summers. 1988. Mean Reversion in Stock Prices: Evidence and Implications. *Journal of Financial Economics* 22(1): 27–60.

Schularick, Moritz, and Alan M. Taylor. 2012. Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008. *American Economic Review* 102(2): 1029–61.

A. Binary classification methods with and without weights

This appendix note draws on Jordà and Taylor (2011) and Taylor (2015).

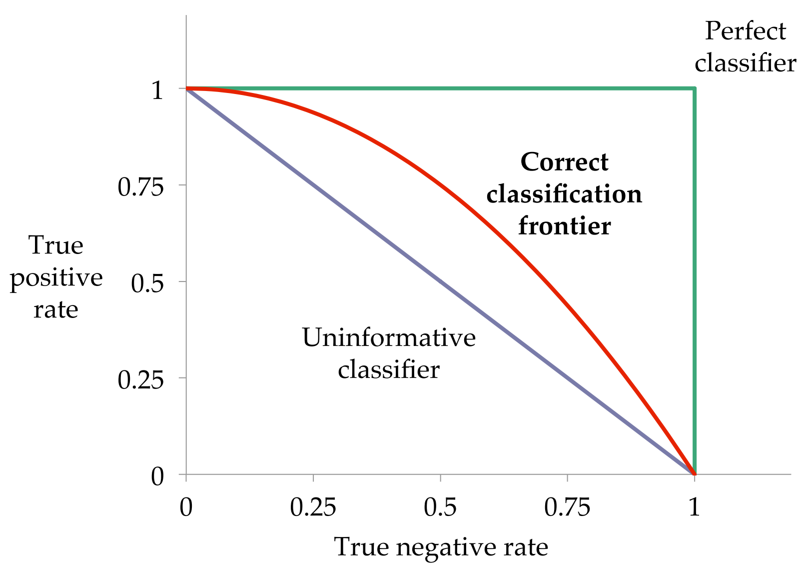
The unweighted *AUC* statistic To frame this for our application, suppose we have an asset return time series r_t and want to ask if a model forecast $\widehat{r}_t$ has profit potential. A naïve binary classification approach might be to ask if the forecast is informative as to whether the asset return beats its average, in any given period. To assess that make an indicator that is 1 for yes, and 0 for no, whenever the return minus its average $\widehat{r}_t - \bar{r}$ exceeds some threshold c, and find the true positives (a subset of all positives T_P) and true negatives (a subset of all positives T_N) of the associated classifier for each value of c:

$$\widehat{TN}(c) = \sum_{j=1}^{T_N} I(\widehat{r}_t - \bar{r} < c), \quad \widehat{TP}(c) = \sum_{i=1}^{T_P} I(\widehat{r}_t - \bar{r} > c), \tag{4}$$

We define the *Correct Classification Frontier* or CCF as the plot of the true positive rate $TP(c)$ against the true negative rate $TN(c)$, for all real thresholds c.

If the threshold c gets large and negative, the classifier is very aggressive in making positive calls, almost all signals are above the threshold, and (TN, TP) converges to $(0, 1)$ as $c \to -\infty$; conversely, when c gets large and positive, the classifier is very conservative in making positive calls, almost all signals are below the threshold, and (TN, TP) converges to $(1, 0)$ as $c \to +\infty$. In between these extremes, an informative classifier should deliver a CCF curve above the simplex or the 45-degree line of the null uninformative (or "coin toss") classifier, as shown in the accompanying Figure A1. The CCF is a variant of the better known ROC or Receiver Operating Characteristic curve.

Figure 4: *A Correct Classification Frontier or CCF*



Notes: The area under the curve (AUC) is 0.5 for the null uninformative classifier and 1 for a perfect classifier, and between these values for a useful classifier. See text.

The area under the CCF, known as the area under the curve (AUC) is 0.5 for the null uninformative classifier and 1 for a perfect classifier. Concerning inference, testing whether a classifier is informative, or better than an alternative classifier, is simple with the AUC statistic since it is asymptotically normally distributed with a variance that can be easily estimated. The test is also independent of the operator's preferences. Such tests are available in most statistical packages.

The weighted AUC statistic The key idea in Jordà and Taylor (2011), also in Fawcett (2006), is to note that the basic AUC statistics value each correct or incorrect classification with an equal weight. Instead, they propose that the classification of each instance receive a weight w_t proportional a gain or loss associated with that observation.

Suppose we consider the mean of the return as the "hurdle rate" and reward the classifier with weight based on return differential relative to mean: a positive "gain" if the differential is positive, a negative "loss" if the differential is negative. For a market timer, a positive is better than the mean, and is when one would want to go longer; a negative is worse than the mean, and is when one would want to go shorter.

The analysis is then shown to be a straightforward transform of the ordinary AUC setup, when the weights for positives (negatives) are scaled by the sum of the maximum possible weights for positives (negatives) that would arise from a perfect classifier.

With those weights applied, a transformed CCF, called the CCF* can be built from the transformed $TN^\star$ and $TP^\star$ statistics given by

$$\widehat{TN^\star}(c) = \sum_{j=1}^{T_N} w_t I(\widehat{r}_t - \bar{r} < c), \quad \widehat{TP^\star}(c) = \sum_{i=1}^{T_P} w_t I(\widehat{r}_t - \bar{r} > c), \tag{5}$$

The area under this curve or $AUC^\star$, inherits all the same properties and methods of inference as the AUC statistic.

Crucially, by dint of the weight design, $AUC^\star$ measures upside and downside capture: $TP^\star(c)$ is the extent to which the strategy succeeds in making a (relative) profit from long positions (i.e., upside captured when returns exceed c). $TN^\star(c)$ is the extent to which the strategy avoids in making a (relative) loss from short positions (i.e., downside avoided when returns are below c).