The Equity Premium and the One Percent

Toda, Alexis Akira and Walsh, Kieran James

University of California, San Diego, Darden School of Business, University of Virginia

31 March 2014

The Equity Premium and the One Percent*

Alexis Akira Toda† Kieran James Walsh‡

First draft: March 2014
This version: March 1, 2017

Abstract

We show that in a general equilibrium model with heterogeneity in risk aversion or belief, shifting wealth from an agent who holds comparatively fewer stocks to one who holds more reduces the equity premium. Since empirically the rich hold more stocks than do the poor, the top income share should predict subsequent excess stock market returns. Consistent with our theory, we find that when the income share of top earners in the U.S. rises, subsequent one year excess market returns significantly decline. This negative relation is robust to (i) controlling for classic return predictors such as the price-dividend and consumption-wealth ratios, (ii) predicting out-of-sample, and (iii) instrumenting with changes in estate tax rates. Cross-country panel regressions suggest that the inverse relation between inequality and returns also holds outside of the U.S., with stronger results in relatively closed economies (emerging markets) than in small open economies (Europe).

Keywords: equity premium; heterogeneous risk aversion; return prediction; wealth distribution; international equity markets.

JEL codes: D31, D52, D53, F30, G12, G17.

1 Introduction

Does the wealth distribution matter for asset pricing? Intuition tells us that it does: as the rich get richer, they buy risky assets and drive up prices. Indeed, over a century ago prior to the advent of modern mathematical finance, Fisher (1910) argued that there is an intimate relationship between prices, the heterogeneity of agents in the economy, and booms and busts. He contrasted

*We benefited from comments by Daniel Andrei, Brendan Beare, Dan Cao, Vasco Carvalho, Peter Debaere, Graham Elliott, Nicolae Gârleanu, John Geanakoplos, Emilien Gouin-bonenfant, Jim Hamilton, Gordon Hanson, Fumio Hayashi, Toshiki Honda, Jiasun Li, Semyon Malamud, Larry Schmidt, Allan Timmermann, Frank Warnock, and seminar participants at Boston College, Cambridge-INET, Carleton, Darden, Federal Reserve Board of Governors, HEC Lausanne, Hitotsubashi ICS, Kyoto, Simon Fraser, Tokyo, UBC, UCSD, Vassar, Yale, Yokohama National University, 2014 Australasian Finance and Banking Conference, 2014 Northern Finance Association Conference, 2015 Econometric Society World Congress, 2015 ICMAIF, 2015 Midwest Macro, 2015 SED, 2015 UVa-Richmond Fed Jamboree, and 2017 AFA. We especially thank Snehal Banerjee, Daniel Greenwald, Stavros Panageas, and Jessica Wachter for detailed comments. Earlier drafts of this paper were circulated with the title "Asset Pricing and the One Percent."

†Department of Economics, University of California San Diego. Email: atoda@ucsd.edu.

‡Darden School of Business, University of Virginia. Email: walshk@darden.virginia.edu.

(p. 175) the "enterpriser-borrower" with the "creditor, the salaried man, or the laborer," emphasizing that the former class of society accelerates fluctuations in prices and production. Central to his theory of fluctuations were differences in preferences and wealth across people.

To see the intuition as to why the wealth distribution affects asset pricing, consider an economy consisting of investors with different attitudes towards risk or beliefs about future dividends. In this economy, equilibrium risk premiums and prices balance the agents' preferences and beliefs. If wealth shifts into the hands of the optimistic or less risk averse, for markets to clear, prices of risky assets must rise and risk premiums must fall to counterbalance the new demand of these agents. In this paper, we establish both the theoretical and empirical links between inequality and asset prices.

This paper has two main contributions. First, we theoretically explore the asset pricing implications of general equilibrium models with heterogeneous agents. In a two period economy populated by CRRA agents with arbitrary risk aversion, belief, and wealth heterogeneity, we prove that there exists a unique equilibrium, and that in this equilibrium, increasing wealth concentration in the hands of stockholders leads to a decline in the equity premium. Although the inverse relationship between wealth concentration and risk premiums under heterogeneous risk aversion has been recognized at least since Dumas (1989) and recently emphasized by Gârleanu and Panageas (2015), in order to test the existing theory one needs to identify the preference types, which is challenging. In contrast, we show that it is sufficient to identify the portfolio types, or agents that have larger portfolio shares of stocks. It does not matter why some agents hold more stocks: while we prove that high risk tolerance or optimism are sufficient conditions for investing more in stocks, it may also be due to other reasons such as lower participation costs.

Second, we empirically explore our theoretical predictions. Given the empirical evidence that the rich invest relatively more in stocks,[1] rising inequality should negatively predict subsequent excess stock market returns. Consistent with our theory, we find that when the income share of the top 1% income earners in the U.S. rises, the subsequent one year U.S. stock market equity premium falls on average. That is, current inequality appears to forecast the subsequent risk premium of the U.S. stock market.

Because top income shares appear nonstationary, we use a stationary component of inequality, "`cgdiff` (capital gains difference)," which we define to be the difference between the top 1% income share with and without realized capital gains income. Regressions of the year t to year $t+1$ excess return on the year t top 1% income share indicate a strong and significant negative correlation: when `cgdiff` rises by one percentage point, subsequent one year excess market returns decline on average by about 3–5%, depending on the controls included. Overall, our evidence suggests that the top 1% income share is not simply a proxy for the price level, which previous research shows correlates with subsequent returns, or for aggregate consumption factors: the top 1% income share predicts excess returns even after we control for some classic return predictors such as the price-dividend ratio (Fama and French, 1988) and the consumption-wealth ratio (Lettau and Ludvigson, 2001). Our findings are also

[1] See, for example, Haliassos and Bertaut (1995), Carroll (2002), Campbell (2006), Wachter and Yogo (2010), Bucciol and Miniaci (2011), and Calvet and Sodini (2014).

robust to the inclusion of macro control variables, such as GDP growth. Since we get very similar results when we detrend the top 1% income share using a variety of trend extraction methods, the construction of `cgdiff` is not driving our findings. Across nearly all of our specifications, the inverse relationship between top income shares and excess returns is large and statistically significant. Using five year excess returns or the top 0.1% or 10% income share also yields similar results.

The empirical literature on return prediction is not without controversy. While many papers find evidence for return predictability,[2] others find mixed evidence (Ang and Bekaert, 2007), and some point out econometric issues such as small sample bias when regressors are persistent (Nelson and Kim, 1993; Stambaugh, 1999) and problems with overlapping data (Valkanov, 2003; Boudoukh et al., 2008). In an influential study, Welch and Goyal (2008) show that excess return predictors suggested in the literature by and large perform poorly out-of-sample. How does the top 1% share fare out-of-sample? Using the methodologies of McCracken (2007) and Hansen and Timmermann (2015), we show that including the top 1% as a predictor significantly decreases out-of-sample forecast errors relative to using the historical mean excess return. That is, top income shares predict returns out-of-sample as well.

Given that in our regressions we lag top shares and given that our results are robust to the inclusion of many macro/financial control variables, we do not suspect our findings stem from reverse causation or omitted variable bias. However, because top tax rates have an inverse relationship with top income shares (Roine et al., 2009), as an additional robustness check, we explore two approaches to using tax changes as an instrument for inequality in predicting returns. First, we identify seven periods in U.S. history (over 1915-2004) where top marginal income and estate tax rates were either trending upwards or downwards. We find that tax hike periods are on average associated with a declining 1% share, flat price-earnings ratios, and positive subsequent excess returns. Tax cut periods, however, are accompanied by a rising 1% income share, increasing price-earnings ratios, and negative subsequent excess returns. Second, since contemporaneous and lagged changes in top estate tax rates explain a substantial portion of the variation in `cgdiff` (Table 5), we also estimate the effect of inequality on returns using GMM with instrumental variables (Table 6). Including `cgdiff`, industrial production growth, and the log price-earnings ratio as endogenous explanatory variables and using lags of top estate tax rate changes and the log price-earnings ratio as instruments, top income shares are still significant in predicting excess returns. This finding addresses another concern, which is that part of the variation in `cgdiff` is not due to inequality but rather from the timing of realizing capital gains. Including one-year-ahead changes in capital gains tax rates as an additional instrument, we separately identify how the timing and inequality components predict returns (Table 8). The coefficient on the inequality component is negative and significant, while the timing

[2]Classic examples are the price-dividend ratio (Campbell and Shiller, 1988; Fama and French, 1988; Hodrick, 1992; Cochrane, 2008) and the consumption-wealth ratio (Lettau and Ludvigson, 2001). Campbell and Thomson (2008) suggest that many economic variables predict returns by imposing weak restrictions such as a nonnegative equity premium. Rapach et al. (2010) show that instead of using a single predictive regression model, combining forecasts significantly decreases the out-of-sample forecast errors. See Lettau and Ludvigson (2010) and Rapach and Zhou (2013) for reviews on forecasting stock returns.

coefficient is insignificant.

We uncover a similar pattern in international data on inequality and financial markets: post-1969 cross-country fixed-effects panel regressions suggest that when the top 1% income share rises above trend by one percentage point, subsequent one year market returns significantly decline on average by 2%. This relationship is particularly strong for relatively "closed" economies such as emerging markets. In countries with low levels of investing home bias ("small open economies"), we find a large and significant inverse relationship between the *U.S. 1% share* (a potential proxy of the global 1% share) and subsequent domestic excess returns. These results are consistent with our theory because our model suggests that what predicts returns is the wealth distribution amongst the set of potential stock and bond holders. For small open economies, local agents comprise a small fraction of this set of investors. In large or relatively closed economies, domestic agents are a substantial proportion of the universe of investors.

1.1 Related literature

For many years after Fisher, in analyzing the link between individual utility maximization and asset prices, financial theorists either employed a rational representative agent or considered cases of heterogeneous agent models that admit aggregation, that is, cases in which the model is equivalent to one with a representative agent. Extending the portfolio choice work of Markowitz (1952) and Tobin (1958), Sharpe (1964) and Lintner (1965a,b) established the Capital Asset Pricing Model (CAPM).[3] These original CAPM papers, which concluded that an asset's covariance with the aggregate market determines its return, actually allowed for substantial heterogeneity in endowments and risk preferences across investors. However, their form of quadratic or mean-variance preferences admitted aggregation and obviated the role of the wealth distribution. Largely inspired by the limited empirical fit of the CAPM and asset pricing puzzles that arise in representative-agent models, since the 1980s theorists have extended macro/finance general equilibrium models to consider meaningful investor heterogeneity. Such heterogeneous-agent models fall into two groups.

In the first group, agents have identical standard (constant relative risk aversion) preferences but are subject to uninsured idiosyncratic risks.[4] Although the models of this literature have had some success in explaining returns in calibrations, the empirical results (based on consumption panel data) are mixed and may even be spuriously caused by the heavy tails in the cross-sectional consumption distribution (Toda and Walsh, 2015, 2016a). In the second group, markets are complete and agents have either heterogeneous CRRA preferences (see Section 2.3) or identical but non-homothetic preferences (Gollier, 2001; Hatchondo, 2008). In this class of models the marginal rates of substitution are equalized across agents and a "representative agent" in the sense of Constantinides (1982) exists, but aggregation in the sense of Gorman (1953) fails. Therefore there is room for agent heterogeneity to matter for asset pricing. However, this type of agent heterogeneity is generally considered to be irrele-

[3]See Geanakoplos and Shubik (1990) for a general and rigorous treatment of CAPM theory.

[4]Examples are Mankiw (1986), Constantinides and Duffie (1996), Heaton and Lucas (1996), Brav et al. (2002), Cogley (2002), Balduzzi and Yao (2007), Storesletten et al. (2007), Kocherlakota and Pistaferri (2009), among many others. See Ludvigson (2013) for a review.

vant for asset pricing because in dynamic models the economy is dominated by the richest agent (the agent with the largest expected wealth growth rate) in the long run (Sandroni, 2000; Blume and Easley, 2006). One notable exception is Gârleanu and Panageas (2015), who study a continuous-time overlapping generations endowment economy with two agent types with Epstein-Zin constant elasticity of intertemporal substitution/constant relative risk aversion preferences. Even if the aggregate consumption growth is i.i.d. (geometric Brownian motion), the risk-free rate and the equity premium are time-varying, even in the long run. The intuition is that when the risk tolerant agents have a higher wealth share, they drive up asset prices and the interest rate. The effect of preference heterogeneity persists since new agents are constantly born. Consistent with our empirical findings and model, the calibration of Gârleanu and Panageas (2015) suggests that increasing the consumption share of more risk tolerant agents pushes down the equity premium. All of the above works are theoretical, and our paper seems to be the first in the literature to empirically test the asset pricing implications of models with preference heterogeneity. In Section 2, we both present our theoretical results and further highlight how we contribute to these literatures.

Although the wealth distribution theoretically affects asset prices, there are few empirical papers that directly document this connection. To the best of our knowledge, Johnson (2012) and Campbell et al. (2016) are the only ones that explore this issue. Using incomplete markets models, they show that top income shares or top income growth innovations are cross-sectional asset pricing factors. However, they do not explore the ability of top income shares to predict excess market returns (our main empirical result).[5]

Lastly, our study is related to the findings of Greenwald et al. (2016), whose contribution is twofold. First, they use a cointegrating vector autoregression to identity orthogonal innovations to consumption, labor income, and wealth ($e_{c,t}$, $e_{y,t}$, and $e_{a,t}$), which explain most of the post-WWII variation in the stock market. Moreover, $e_{a,t}$ alone explains over 75% *and* strongly and significantly predicts low subsequent excess returns. Second, the authors put structure on these innovations using an equilibrium model with a representative laborer who does not hold stock, a representative stockholder, and shocks to productivity, the labor share of output, and stockholder risk aversion. The key result is that the model-generated $e_{a,t}$ series is almost perfectly correlated with shocks that decrease the risk aversion parameter of the stockholder. This suggests that $e_{a,t}$ captures the risk tolerance of the representative stockholder. Our analysis adds further microfoundation and corroboration of their story: interestingly, there is substantial correlation between $e_{a,t}$ and our inequality predictor variable `cgdiff`.[6] As we show in Section 2, in heterogeneous risk aversion models without aggregation, rising wealth concentration can effectively decrease the risk aversion of the corresponding representative stockholder/planner. Therefore, an alternative interpretation/microfoundation of $e_{a,t}$ is that it reflects the wealth share of relatively risk tolerant stockholders vs. more risk averse ones.

[5]Campbell et al. (2016) do explore market return prediction in their online appendix, but they uncover no relationship between the income of the rich and subsequent stock returns. Our findings are different likely because they use income instead of the income share and since they detrend top income linearly.

[6]Also, the component of `cgdiff` orthogonal to $e_{a,t}$ does not significantly predict excess returns. We thank Daniel Greenwald for discovering this.

2 Wealth distribution and equity premium

In this section we present a theoretical model in which the wealth distribution across heterogeneous agents affects the equity premium. In Section 2.1, we consider a static model with agents that have heterogeneous risk aversion and beliefs, and prove the uniqueness of equilibrium. In Section 2.2, we prove that shifting wealth from an agent that holds comparatively fewer stocks to one that holds more pushes down the equity premium. Section 2.3 compares our results to the existing literature. All proofs are in Appendix A.

2.1 Uniqueness of equilibrium

Consider a standard general equilibrium model with incomplete markets consisting of I agents and J assets (Geanakoplos, 1990). Time is denoted by $t = 0, 1$: agents trade assets at $t = 0$ and consume only at $t = 1$. At $t = 1$, there are S states denoted by $s = 1, \dots, S$. Let $e_i \in \mathbb{R}^S_{++}$ be agent i's endowment vector of consumption goods in each state and $A = (A_{sj}) \in \mathbb{R}^{SJ}$ be the $S \times J$ payoff matrix of assets. By redefining the initial endowments of goods if necessary, without loss of generality we may assume that the initial endowments of assets are zero. By removing redundant assets, we may also assume that the matrix A has full column rank.

Given the asset price $q = (q_1, \dots, q_J)' \in \mathbb{R}^J$, agent i's utility maximization problem is

$$\begin{aligned} &\text{maximize} && U_i(x) \\ &\text{subject to} && q'y \le 0, x \le e_i + Ay, \end{aligned}$$

where $U_i(x)$ is the utility function and $y = (y_1, \dots, y_J)' \in \mathbb{R}^J$ denotes the number of asset shares. $q'y \le 0$ is the $t = 0$ budget constraint. $x \le e_i + Ay$ is the $t = 1$ budget constraint. A general equilibrium with incomplete markets (GEI) consists of asset prices $q \in \mathbb{R}^J$, consumption $(x_i) \in \mathbb{R}^{SI}_+$, and portfolios $(y_i) \in \mathbb{R}^{JI}$ such that (i) agents optimize, and (ii) asset markets clear, so $\sum_{i=1}^I y_i = 0$.

We make the following assumptions.

Assumption 1 (Heterogeneous CRRA preferences)**.** *Agents have constant relative risk aversion (CRRA) preferences:*

$$U_i(x) = \begin{cases} \left(\sum_{s=1}^S \pi_{is} x_s^{1-\gamma_i}\right)^{\frac{1}{1-\gamma_i}}, & (\gamma_i \ne 1) \\ \exp\left(\sum_{s=1}^S \pi_{is} \log x_s\right), & (\gamma_i = 1) \end{cases} \tag{2.1}$$

where $\gamma_i > 0$ is agent i's relative risk aversion coefficient and $\pi_{is} > 0$ is agent i's subjective probability of state s.

Note that if $\gamma_i \ne 1$, through the monotonic transformation $x \mapsto \frac{1}{1-\gamma_i} x^{1-\gamma_i}$, U_i is equivalent to

$$\frac{1}{1-\gamma_i} U_i(x)^{1-\gamma_i} = \frac{1}{1-\gamma_i} \sum_{s=1}^S \pi_{is} x_s^{1-\gamma_i},$$

the standard additive CRRA utility function. The same holds when $\gamma_i = 1$ by considering $\log U_i(x)$. The expression (2.1) turns out to be more convenient since the utility function is homogeneous of degree 1.

Assumption 2 (Collinear endowments)**.** *Agents have collinear endowments: letting* $e = \sum_{i=1}^I e_i \gg 0$ *be the aggregate endowment, we have* $e_i = w_i e$*, where* $w_i > 0$ *is the wealth share of agent* i*, so* $\sum_{i=1}^I w_i = 1$*.*

While the collinearity assumption is strong, it is indispensable in order to guarantee the uniqueness of equilibrium: Mantel (1976) shows that if we drop collinear endowments, then even with homothetic preferences "anything goes" for the aggregate excess demand function, and hence there may be multiple equilibria.[7] With multiple equilibria, comparative statics may go in opposite directions, depending on the choice of equilibrium.

Assumption 3 (Tradability of aggregate endowment)**.** *The aggregate endowment is tradable:* e *is spanned by the column vectors of* A*.*

Under these assumptions, we can prove the uniqueness of GEI and obtain a complete characterization.

Theorem 2.1. *Under Assumptions 1–3, there exists a unique GEI. The equilibrium portfolio* (y_i) *is the solution to the planner's problem*

$$\begin{aligned} \underset{(y_i)\in\mathbb{R}^{JI}}{\text{maximize}} \quad & \sum_{i=1}^I w_i \log U_i(e_i + Ay_i) \\ \text{subject to} \quad & \sum_{i=1}^I y_i = 0. \end{aligned} \tag{2.2}$$

Letting

$$\sum_{i=1}^I w_i \log U_i(e_i + Ay_i) - q' \sum_{i=1}^I y_i$$

be the Lagrangian with Lagrange multiplier q*, the equilibrium asset price is* q*.*

Chipman (1974) shows that under complete markets, heterogeneous homothetic preferences, and collinear endowments, aggregation is possible and hence the equilibrium is unique. Our Theorem 2.1 is a stronger result since we prove the same for incomplete markets and we also obtain a complete characterization of the equilibrium portfolio. Uniqueness is important for our purposes because it rules out unstable equilibria and thus allows for the below unambiguous comparative statics regarding the wealth distribution.[8]

2.2 Comparative statics

Assuming that only a stock and a bond are traded, we can show that a redistribution of wealth from an investor that holds comparatively fewer stocks to one that holds more reduces the equity premium. To make the statement precise, we introduce the following assumption and notations.

Assumption 4. *The only assets traded are the aggregate stock and a risk-free asset:* $J = 2$ *and* $A = [e, 1]$*, where* $1 = (1, \dots, 1)' \in \mathbb{R}^S_{++}$*.*

[7] See Toda and Walsh (2016b) for concrete examples of multiple equilibria with canonical two-agent, two-state economies.

[8] See Kehoe (1998) and Geanakoplos and Walsh (2016) for further discussion of uniqueness in the presence of heterogeneous preferences.

Let $q = (q_1, q_2)'$ be the vector of asset prices. By the proof of Theorem 2.1, we have $q \gg 0$. Since there is no consumption at $t = 0$, we can normalize asset prices, so without loss of generality we may take the gross risk-free rate $R_f = 1/q_2$ as given. The vector of gross stock returns is $R := e/q_1$. Since by Assumption 2 we have $e_i = w_i e$, the initial wealth of agent i is $q_1 w_i$ and the budget constraint is

$$q_1 y_1 + q_2 y_2 \leq 0 \iff q_1(y_1 + w_i) + \frac{1}{R_f} y_2 \leq q_1 w_i.$$

Letting $\theta = \frac{y_1 + w_i}{w_i}$ be the fraction of wealth invested in the stock, by the budget constraint with equality we have $1 - \theta = \frac{y_2}{q_1 R_f w_i}$. Therefore the consumption vector satisfies

$$x \leq e_i + Ay = (y_1 + w_i)e + y_2 1 = q_1 w_i (R\theta + R_f(1 - \theta)1).$$

Letting

$$u_i(x) = \begin{cases} \frac{1}{1-\gamma_i} x^{1-\gamma_i}, & (\gamma_i \neq 1) \\ \log x, & (\gamma_i = 1) \end{cases}$$

by homotheticity the utility maximization problem is equivalent to

$$\max_\theta \mathrm{E}_i[u_i(R(\theta))],$$

where $R(\theta) := R\theta + R_f(1 - \theta)$ and E_i denotes the expectation under agent i's belief.

Now we can state our main theoretical result.

Theorem 2.2. *Suppose Assumptions 1–4 hold and let $\{q, (x_i), (y_i)\}$ be the unique GEI with corresponding portfolio (θ_i). Suppose that in the initial equilibrium agent 1 holds comparatively fewer stocks than agent 2, so $\theta_1 < \theta_2$. If we transfer wealth from agent 1 to 2, then the new equilibrium has a higher stock price q_1. The equity premium $\mathrm{E}[R] - R_f$ (computed using any fixed probability distribution) becomes lower.*

The intuition for Theorem 2.2 is straightforward. In an economy with financial assets, the equilibrium risk premiums and prices balance the agents' preferences and beliefs. If wealth shifts into the hands of the natural buyer (either the risk tolerant or optimistic agent), for markets to clear, prices of risky assets must rise and risk premiums must fall to counterbalance the new demand of these agents. While the conclusion of Theorem 2.2 is quite natural and intuitive, proving it is another story. Since the direction of comparative statics depends on the choice of equilibrium if there are multiple equilibria, we need to rule out this possibility. Only with the uniqueness result in Theorem 2.1 are we able to make the intuition rigorous.

The following propositions show that when agents have heterogeneous risk aversion or beliefs, the portfolio share of the risky asset is ordered as risk tolerance or optimism. To define optimism, we take the following approach. First, by relabeling states if necessary, without loss of generality we may assume that states are ordered from bad to good ones: $e_1 < \cdots < e_S$. Consider two agents $i = 1, 2$ with subjective probability $\pi_{is} > 0$. We say that agent 1 is more pessimistic than agent 2 if the likelihood ratio $\lambda_s := \pi_{1s}/\pi_{2s} > 0$ is monotonically decreasing: $\lambda_1 \geq \cdots \geq \lambda_S$, with at least one strict inequality.

Proposition 2.3. *Suppose Assumptions 1–4 hold and agents have common beliefs. If* $\gamma_1 > \cdots > \gamma_I$*, then* $0 < \theta_1 < \cdots < \theta_I$.

Proposition 2.4. *Suppose Assumptions 1–4 hold and agents 1, 2 have common risk aversion. Assume that agent 1 is more pessimistic than agent 2 in the above sense. Then* $\theta_1 < \theta_2$.

Combining Theorem 2.2 together with either Proposition 2.3 or 2.4, shifting wealth from a more risk averse or pessimistic agent to a more risk tolerant or optimistic agent reduces the equity premium. In particular, if the rich are relatively more risk tolerant, optimistic, or simply more likely to buy risky assets (for example due to fixed information or transaction costs), rising inequality should forecast declining excess returns.

2.3 Discussion

Dumas (1989) solves a dynamic general equilibrium model with constant-returns-to-scale production and two agents (one with log utility and the other CRRA). He shows (Proposition 17) that when the wealth ratio of the less risk averse agent increases, then the risk-free rate goes up and the equity premium goes down. Although this prediction is similar to ours, he imposes an assumption on endogenous variables (see his equation (8)).

Following Dumas (1989), a large theoretical literature has studied the asset pricing implication of preference heterogeneity under complete markets.[9] All of these papers characterize the equilibrium and asset prices by solving a planner's problem. However, this approach is not suitable for conducting comparative statics exercises of changing the wealth distribution, for two reasons. First, although by the second welfare theorem, for each equilibrium we can find Pareto weights such that the consumption allocation is the solution to the planner's problem, since in general the Pareto weights depend on the initial wealth distribution, changing the wealth distribution will change the Pareto weights, and consequently the asset prices. But in general it is hard to predict how the Pareto weights change. Second, even if we can predict how the Pareto weights change, there is the possibility of multiple equilibria. In such cases the comparative statics often go in the opposite direction depending on the choice of the equilibrium. Thus our results are quite different since we prove the uniqueness of equilibrium and derive comparative statics with respect to the initial wealth distribution.

Gollier (2001) studies the asset pricing implication of wealth inequality among agents with identical preferences. He shows that more inequality increases (decreases) the equity premium if and only if agents' absolute risk tolerance is concave (convex). In particular, wealth inequality has no effect on asset pricing when agents have hyperbolic absolute risk aversion (HARA) preferences, for which the absolute risk tolerance is linear. He also calibrates the model and finds that the effect of wealth inequality on the equity premium is small. Our results are different and complementary since our model features heterogeneous CRRA agents.

Gârleanu and Panageas (2015) study a continuous-time overlapping generations endowment economy with two agent types with Epstein-Zin preferences.

[9]Examples are Wang (1996), Chan and Kogan (2002), Hara et al. (2007), Cvitanić et al. (2012), Longstaff and Wang (2012), Bhamra and Uppal (2014), and the references therein. Basak and Cuoco (1998) consider a limited market participation model with log utility.

Unlike other papers on asset pricing models with heterogeneous preferences, all agent types survive in the long run due to birth/death, and they also solve the model without appealing to a planning problem. As a result, all endogenous variables are expressed as functions of the state variable, the consumption share of one agent type. They find that the concentration of wealth to the more risk-tolerant type ("the rich") tends to lower the equity premium. When the preferences are restricted to additive CRRA, then the relation between the consumption share and equity premium (more precisely, market price of risk) is monotonic (see their discussion on p. 10). Thus our results are closely related to theirs, but again are different and complementary since our model features many agents, discrete time (hence our shocks are arbitrary), and incomplete markets (without spanning).

3 Predictability of returns with inequality

In Theorem 2.2, we have theoretically shown that shifting wealth from an agent who holds comparatively fewer stocks to one who holds more reduces the subsequent equity premium. Many empirical papers show that the rich hold relatively more stocks than the poor and argue that the rich are relatively more risk tolerant.[10] Therefore, rising inequality should negatively predict subsequent excess stock market returns. In this section we construct a stationary measure of inequality and show that it predicts subsequent returns.[11]

3.1 Connecting theory to empirics

The ideal way to test our theory is to run regressions of the form

$$\text{ExcessReturn}_{t+1} = \alpha + \beta \times \text{WealthInequality}_t + \gamma \times \text{Controls}_t + \varepsilon_{t+1}$$

and test whether $\beta = 0$. The first obstacle is that it is difficult to measure wealth, and hence of wealth inequality. The 1916-2000 top wealth share series (based on estate tax data) from Kopczuk and Saez (2004) are missing many years in the 50s, 60s, and 70s. The wealth share data of Saez and Zucman (2016) cover 1913-2012 but are *estimates* created by capitalizing income. Due to these limitations, we perform our analysis with top income share data, which are correlated with top wealth shares.[12] We can justify this point by the fact that the rich are more likely to be entrepreneurs, and the income from capital investment is proportional to invested capital.

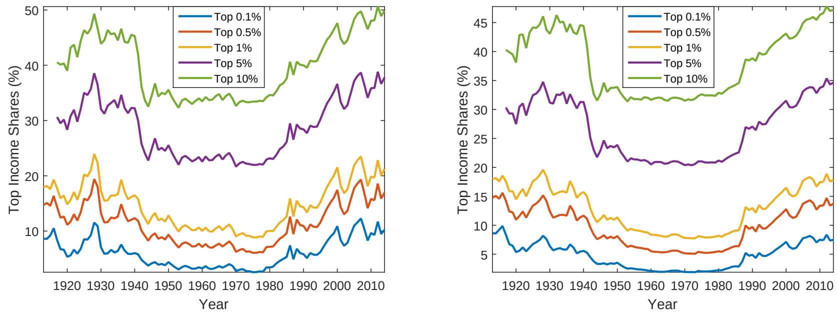
We employ the Piketty and Saez (2003) inequality measures for the U.S., which are available on the website of Alvaredo et al. (2015). In particular, we consider top income share measures based on tax return data, which are at the annual frequency and cover the period 1913-2014. These series reflect in a given year the percent of income earned by the top 1% of earners pretax. We also

[10] See Haliassos and Bertaut (1995), Carroll (2002), Vissing-Jørgensen (2002), Campbell (2006), Wachter and Yogo (2010), Bucciol and Miniaci (2011), and Calvet and Sodini (2014).

[11] In this section we are only concerned with predictability, or correlation. We address causality in Section 4.

[12] Using first differences, the correlation between the 1% income share and the Saez and Zucman (2016) 1% wealth share is about 0.5. In levels, the correlation is 0.7. In previous versions of this paper, we showed that these top wealth series also predict excess returns. The results are omitted to save space but are available upon request.

employ the top 0.1% share, the top 10% share, and the corresponding series that exclude realized capital gains income. Figure 1 show these series, both including capital gains (Figure 1a) and excluding capital gains (Figure 1b). We can immediately see that all series seem to share a common U-shaped trend over the century, and the series including capital gains are more volatile than those without capital gains.



(a) Including capital gains. (b) Excluding capital gains.

Figure 1: U.S. top income shares (1913-2014).

In order to use the top income share as a proxy for top wealth share, a large fraction of the income of top earners must come from capital income. To see if this assumption is satisfied, Figure 2 plots the "capital gains ratio", defined by

$$1 - \frac{\text{Top income share excluding capital gains}}{\text{Top income share including capital gains}}.$$

Although this number is different from the fraction of realized capital gains income relative to total income of any particular individual, it gives some idea for a typical number in each income group. The time series average of "capital gains ratio" is 12.6% for the top 1% earners and 22.4% for the top 0.1% earners. Since capital gains ratios reflect only *realized* capital gains in that particular year, the ratios between capital income (including both realized and unrealized capital gains as well as other capital income not reported as capital gains) and total income are likely even larger. Therefore it seems reasonable to assume that the top income share is a good proxy for the top wealth share.

Finally, there is an econometric issue to overcome. While excess returns are clearly stationary, the top income shares in Figure 1 are nonstationary. For example, the Phillips-Perron test (Phillips and Perron, 1988) fails to reject a unit root in the top 0.1%, 1%, and 10% shares (the p-values are 0.46, 0.59, and 0.84). Therefore, the regression specification

$$\text{ExcessReturn}_{t+1} = \alpha + \beta \times \text{TopShare}_t + \varepsilon_{t+1}$$

implies the error term ε_{t+1} cointegrates with the regressor TopShare_t, which violates OLS assumptions. If we were to regress a stationary variable Y_t (returns) on a nonstationary variable X_t (top 1% income share), the OLS coefficient (ignoring the constant term) $\widehat{\beta}_T = \sum X_t Y_t / \sum X_t^2$ would converge to 0 almost surely because the denominator diverges to ∞ faster than does the numerator. Granger (1981) explains this problem, stating on p. 127 that "one would not

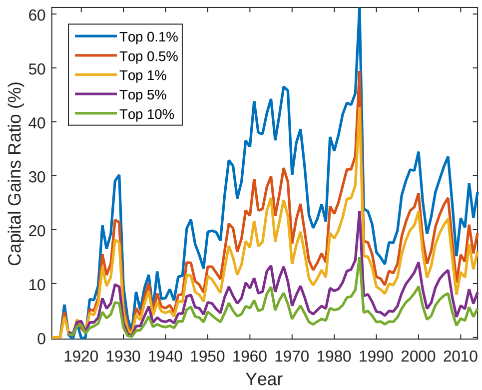


Figure 2: Capital gains ratio.

expect to build a model of [this] form. . . [because] one is attempting to explain a finite variance series by an infinite variance one." See Phillips and Lee (2013) for a recent treatment of this issue.

Given the nonstationarity in the raw top income share series, for our baseline results we use "`cgdiff` (capital gains difference)," which we define to be the top income share (0.1%, 1%, or 10%) including capital gains minus the top share without capital gains. Since `cgdiff` is stationary,[13] it ensures the validity of standard error calculations and inference and prevents spurious regressions (Granger and Newbold, 1974). `cgdiff` can be interpreted as a measure of the component of the top 1% income share due to variation in capital income.[14] Since the top income share is equal to `cgdiff` plus the top income share excluding capital gains, `cgdiff` represents an additive and stationary component of inequality. As we show in Appendix B, we get similar results when we detrend inequality using the Kalman filter with an AR(1) cyclical component (see Appendix C for details). This is unsurprising since, as Figure 3 shows, the two series behave similarly (correlation 0.73).

3.2 In-sample predictions

We calculate real annual one year U.S. stock excess market returns using the annual data updated from Welch and Goyal (2008).[15] The spreadsheet contains historical one year interest rates and price, dividend, and earnings series for the S&P 500 index, which are all put into real terms using consumer price index (CPI) inflation. These data are used to calculate the series P/D and P/E, which are the price-dividend and price-earnings ratios (in real terms) for the S&P 500. The spreadsheet also contains the Lettau and Ludvigson (2001) consumption-wealth ratio, commonly referred to as CAY, which spans the period 1945-2014.

[13]The Phillips-Perron test rejects a unit root in `cgdiff` at the 1% level.

[14]In order for this interpretation to be valid, we need to assume that agents realize capital gains randomly (*i.e.*, do not time when to realize capital gains). This assumption is not strictly satisfied. For example, the largest `cgdiff` observation in Figure 3 occurred in 1986, but the capital gains tax rate increased from 20% in 1986 to 28% in 1987 (but announced in 1986), which gave investors an incentive to realize capital gains in 1986. In Section 4 we disentangle the inequality and timing components of `cgdiff`.

[15]`http://www.hec.unil.ch/agoyal/`

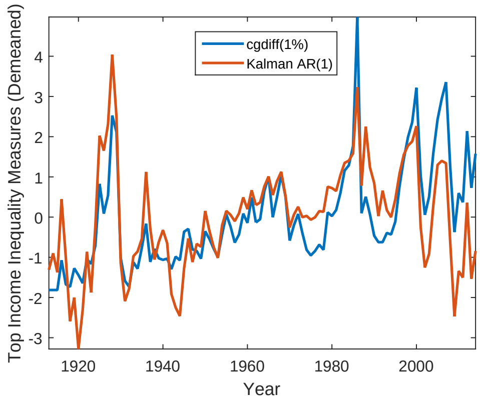


Figure 3: Time series plot of the stationary component of the top 1% income share using `cgdiff` and the AR(1) Kalman filter (both demeaned).

For presentation, we multiply CAY by 100.

Our other controls are GDP growth and, inspired by Lettau et al. (2008) and Bansal et al. (2014), consumption growth variance. Annual data for real GDP and real consumption are from the website of the Federal Reserve Bank of St. Louis (FRED)[16] and span 1930-2014. We estimate consumption growth variance using an AR(1)-GARCH(1,1) model for consumption growth.

Table 1 shows the results of regressions of one year (t to $t+1$) excess stock market returns on our top share measure (time t `cgdiff`), some classic return predictors (time t), and macro factors (time t). In column (1) we find that when the top 1% income share with capital gains (January to December of year t) rises above the series excluding capital gains by one percentage point, subsequent one year excess market returns (January to December of year $t+1$) decline on average by 3.5%. The coefficient is significant at the 1% level (using a Newey-West standard error), and the R-squared statistic is 0.05. It is clear, at least in sample, that the stationary top 1% share component (`cgdiff`) forecasts the subsequent excess return on the stock market.

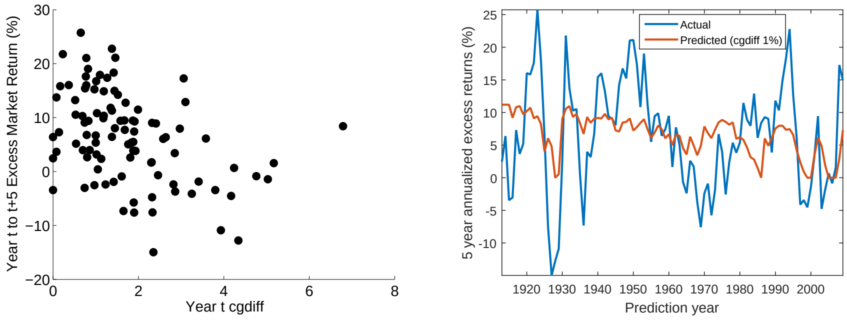
In Tables 13 and 14 (in Appendix B), we see that the inverse relationship between the top 1% share and subsequent excess returns is larger in magnitude and significant at the 1% and 5% levels, respectively, when we predict with `cgdiff`(10%) or `cgdiff`(0.1%). Table 2 shows that all three versions of `cgdiff` also significantly predict five year excess returns. Figures 4a and 4b show the corresponding scatter and time series plots for five year returns. The top 1% income share appears to forecast subsequent five year excess returns well except around 1970 and the 1980s. Overall, a one percentage point increase in the `cgdiff` component of inequality is associated with, roughly, a 3–5% decline in subsequent excess returns.

[16]`http://research.stlouisfed.org/fred2/`

Table 1: Regressions of one year excess stock market returns on `cgdiff` (top 1% − top 1% (without capital gains)) and other predictors

	Dependent Variable: t to $t+1$ Excess Market Return						
Regressors (t)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	14.42	14.32	15.18	16.42	16.76	15.31	23.62
	(2.91)	(4.05)	(3.37)	(15.21)	(11.10)	(3.72)	(14.82)
`cgdiff`(1%)	-3.51***	-3.72**	-3.76**	-3.33*	-3.37**	-3.48**	-2.90*
	(1.27)	(1.57)	(1.55)	(1.77)	(1.38)	(1.57)	(1.60)
Real GDP Growth		0.27					-0.11
		(0.40)					(0.80)
ΔCons. Growth Variance			-10.14				-23.72
			(15.10)				(16.56)
log(P/D)				-0.71			-2.88
				(5.13)			(4.48)
log(P/E)					-0.97		
					(4.26)		
CAY						1.46*	1.47*
						(0.74)	(0.77)
Sample	1913-	1930-	1935-	1913-	1913-	1945-	1945-
	-2014	-2014	-2014	-2014	-2014	-2014	-2014
R^2	0.05	0.06	0.07	0.05	0.05	0.13	0.15

Note: Newey-West standard errors in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). "`cgdiff`" is top 1% minus top 1% (no cg), neither detrended, where top 1% is the pre-tax share of income going to the top 1% of earners (including capital gains). Consumption growth volatility is from an $AR(1) - GARCH(1,1)$ model. P/D and P/E are the S&P500 price-dividend and price-earnings ratios. CAY is the consumption/wealth ratio.



(a) Scatter plot.

(b) Time series plot.

Figure 4: Year t to year $t+5$ excess stock market return (annualized) vs. year t `cgdiff` (top 1% income share including capital gains minus top 1% share without capital gains), 1913-2014.

Table 2: Regressions of five year excess stock market returns on `cgdiff` (top income share − top income share (without capital gains))

	Dependent Variable: t to $t+5$ Excess Market Return		
Regressors (t)	(1)	(2)	(3)
Constant	10.79	11.19	13.12
	(2.02)	(2.18)	(2.30)
`cgdiff`(0.1%)	-3.59***		
	(1.28)		
`cgdiff`(1%)		-2.71***	
		(1.01)	
`cgdiff`(10%)			-3.72***
			(1.22)
Sample	1913- -2014	1913- -2014	1917- -2014
R^2	0.17	0.17	0.21

Note: Newey-West standard errors in parentheses ($k=8$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). Five year excess returns are annualized. `cgdiff`(0.1%) is top 0.1% minus top 0.1% (no cg), neither detrended, where top 0.1% is the pre-tax share of income going to the top 0.1% of earners (including capital gains). `cgdiff`(1%) and `cgdiff`(10%) are defined analogously.

Given the strength of the relationship, a question immediately arises. Is there some mechanical, non-equilibrium explanation for the relationship between inequality and subsequent excess returns? For example, might stock returns somehow be *determining* the top share measures? For a few reasons, the answer is likely no. First, the relationship is between *initial* inequality and *subsequent* returns. Returns could affect contemporaneous top shares but not lagged top shares. One might still worry that our results are driven by our transformation of the top share series into `cgdiff`. However, as we see in Appendix B, we get similar results with other methods of creating a stationary series.[17] Furthermore, instrumental variables estimates in Section 4 suggest both that `cgdiff` represents inequality (vs. capital gains timing) and that the link between inequality and subsequent returns in causal.

But, one might say, we have known at least since Fama and French (1988) that when prices are high relative to either earnings or dividends, subsequent excess market returns are low. The current price could indeed affect current inequality. Are the top shares series simply proxying for the price-dividend or price-earnings ratios, which are known to predict returns? Again, the answer seems to be no for two reasons. First, excluding capital gains from income does not mitigate the relationship (see footnote 17), and capital gains are the main avenue through which prices would determine inequality. Second, as we see in regressions (4) and (5) from Table 1, top shares predict excess returns *even when controlling for the log price-dividend or price-earnings ratio.* Including these controls does decrease the top share coefficient slightly, but it remains large and significant. The P/D and P/E ratios, however, are not significant after controlling for top income shares. Controlling for P/E or P/D barely impacts the `cgdiff`(10%) results (columns (4) and (5) of Table 13). With respect to `cgdiff`(0.1%), we lose significance when controlling for P/D, but the top share coefficient remains large and significant when including P/E.

In regressions (2), (3), (6), and (7) from Table 1, we also control for real GDP growth, consumption growth variance (Lettau et al., 2008; Bansal et al., 2014),[18] and CAY, which Lettau and Ludvigson (2001) show forecasts excess market returns. Including these controls (which also shortens the sample), we still see a strong relationship between the top income share and subsequent returns. When controlling for CAY, consumption growth variance, GDP growth, or all three and log(P/D) (column 7), the 1% coefficient is around -3 and significant at the 10%, 5% or 1% level. Our results are almost uniformly stronger with `cgdiff`(10%), the Kalman filter (Table 15), or the HP filter (Table 16).

Our empirical analysis thus far has relied on creating stationary inequality series. Do the raw data indicate a relationship between asset prices and the one percent? Figure 5 suggests that the answer is yes. Over 1913-2014, there is a clear positive correlation between the top 1% income share (not detrended) and the contemporaneous price-dividend ratio. Of course, this scatter plot does not establish causation, but it is more evidence in favor of our theory and suggests that our empirical results are not simply artifacts of the construction of `cgdiff`.

In summary, the data appear consistent with our theory that an increasing concentration of income decreases the market risk premium.

---

[17] Tables 15 and 16 in Appendix B repeat Table 1 but with the Kalman and HP filters, respectively, yielding a comparable relationship between inequality and returns. Table 17 shows that we get similar results using the one-sided HP filter, the 10 year moving average filter, and linear detrending. Furthermore, as we see in regression (3) from Table 15, which uses the Kalman filter, when excluding capital gains, the top 1% income share coefficient actually strengthens from -2.82 to -3.78. If returns were strongly affecting lagged inequality, excluding capital gains would likely *mitigate* the result. While the top 1% coefficient is larger in magnitude without capital gains, removing capital income increases the Newey-West standard error and pushes the p-value from 0.02 to 0.11. This is not the case, however, when using the HP filter. Comparing columns (1) and (3) of Table 16, we see that with the HP filter excluding capital gains increases the magnitude of the association *and* maintains significance (at the 5% level).

[18] We difference consumption growth variance, which appears nonstationary, and drop 1930-1934 to reduce the impact of the initial variance.

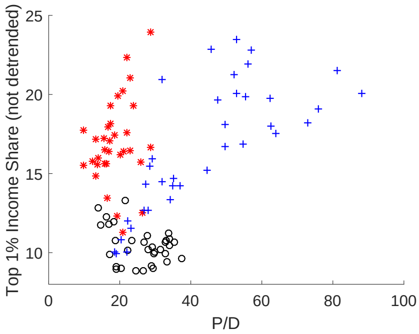


Figure 5: Top 1% income share (not detrended) vs. price-dividend ratio (in real terms) for the S&P500. 1913-1945 (*), 1946-1978 (o), and 1979-2014 (+).

3.3 Out-of-sample predictions

So far, we have seen that the current top income share predicts future excess stock market returns in-sample. However, Welch and Goyal (2008) have shown that the predictors suggested in the literature by and large perform poorly out-of-sample, possibly due to model instability, data snooping, or publication bias. In this section, we explore the ability of the top income share to predict excess stock market returns out-of-sample.

Consider the predictive regression model for the equity premium,

$$y_{t+h} = \beta' x_t + \epsilon_{t+h}, \tag{3.1}$$

where h is the forecast horizon (typically $h = 1$), y_{t+h} is the year t to $t + h$ excess stock market return, x_t is the vector of predictors, ϵ_{t+h} is the error term, and β is the population OLS coefficient. Suppose that the predictors can be divided into two groups, so $x_t = (x_{1t}, x_{2t})$ and $\beta = (\beta_1, \beta_2)$ accordingly. In this section we are interested in whether the variables x_{2t} are useful in predicting y_{t+h}, that is, we want to test $H_0 : \beta_2 = 0$. We call the model with $\beta_2 = 0$ the NULL model and the one with $\beta_2 \neq 0$ the ALTERNATIVE.

To evaluate the performance of the ALTERNATIVE model against the null, following McCracken (2007) and Hansen and Timmermann (2015) we consider the following out-of-sample F statistic:

$$F = \frac{1}{\hat{\sigma}_\epsilon^2} \sum_{t=\lfloor \rho T \rfloor + 1}^{T} \left[(y_{t+h} - \hat{y}^N_{t+h|t})^2 - (y_{t+h} - \hat{y}^A_{t+h|t})^2 \right], \tag{3.2}$$

where $\hat{\sigma}_\epsilon^2$ is a consistent estimator of $\mathrm{Var}[\epsilon_{t+h}]$ (which we estimate from the sample average of the squared OLS residuals of (3.1) using the whole sample), $\hat{y}^A_{t+h|t} = \hat{\beta}_t' x_t$ ($\hat{y}^N_{t+h|t} = \hat{\beta}_{1t}' x_t$) is the predicted value of y_{t+h} based on x_t using the ALTERNATIVE (NULL) model (here $\hat{\beta}_t, \hat{\beta}_{1t}$ are the OLS estimator of (3.1) using data only up to time t), T is the sample size, and $0 < \rho < 1$ is the proportion of observations set aside for initial estimation of β and β_1. Theorems 3 and 4 of Hansen and Timmermann (2015) show that under the null ($H_0 : \beta_2 = 0$), the asymptotic distribution of F is a weighted sum of the difference of independent $\chi^2(1)$ variables.

For the regressors in the ALTERNATIVE model, following Welch and Goyal (2008), we consider the simplest possible case where $x_{1t} \equiv 1$ (constant) and x_{2t} consists of a single predictor. For the predictor x_{2t}, we consider `cgdiff` and valuation ratios (log(P/D) and log(P/E)). The reason is that (i) since the top income series is at annual frequency, the sample size is already small at around 100 (1913 to 2014), so we cannot afford to use variables that are available only in shorter samples (*e.g.*, CAY) for performing out-of-sample predictions, and (ii) since Welch and Goyal (2008) find that most predictor variables suggested in the literature are poor, there is no point in comparing many variables. The choice of the proportion of the training sample, ρ, is necessarily subjective. Small ρ leads to imprecise initial estimates of β, and large ρ leads to the loss of power. Hence we simply report results for $\rho = 0.2, 0.3, 0.4$. Table 3 shows the results.

Table 3: Out-of-sample performance of the top 1% series in predicting subsequent 1-year excess returns

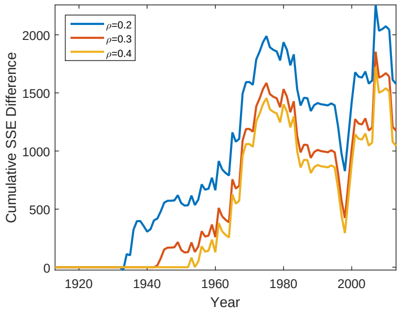
	Predictor in the ALTERNATIVE Model				
ρ	`cgdiff`(1%)	`cgdiff`(10%)	`cgdiff`(0.1%)	log(P/D)	log(P/E)
0.2	4.51***	7.03***	2.90***	0.54*	-0.00
	(0.0019)	(0.0006)	(0.0088)	(0.0755)	(0.1205)
0.3	3.36***	4.94***	1.89**	1.02*	0.89*
	(0.0057)	(0.0019)	(0.0222)	(0.0604)	(0.0573)
0.4	2.99**	5.25***	1.41**	0.60*	0.89*
	(0.0073)	(0.0012)	(0.0402)	(0.0959)	(0.0664)

Note: $\rho = 0.2, 0.3, 0.4$ is the proportion of observations set aside to compute an initial OLS estimate. Columns correspond to the predictors included in the ALTERNATIVE model in addition to a constant ($\emptyset$ indicates no additional regressors). The numbers in the table are the out-of-sample F statistic computed by (3.2). p-values (in parentheses) are computed by simulating 10,000 realizations from the asymptotic distribution based on Hansen and Timmermann (2015) (one sided). ***, **, and * indicate significance at 1%, 5%, and 10% levels.

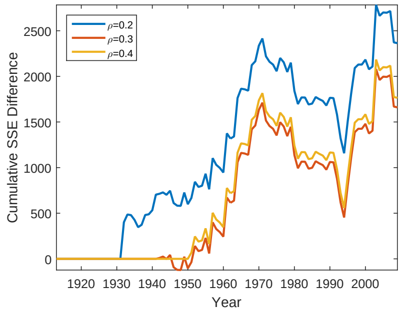
According to Table 3, we can see that across specifications, the out-of-sample F statistic is positive and significant when we use `cgdiff`, while it is insignificant or marginally significant for log(P/D) or log(P/E). (Note that since the asymptotic distribution of F depends on the NULL model, the relationship between the F statistic in Table 3 and the p-values are not necessarily monotonic across models.)

To see this result graphically, in the spirit of Welch and Goyal (2008), we plot the difference in the cumulative sum of squared errors (the numerator of (3.2)) over the prediction period in Figure 6. The vertical axis is the cumulative sum for the NULL model minus the ALTERNATIVE, so a positive value favors the ALTERNATIVE. We can see that for all `cgdiff` specifications, the plots roughly monotonically increase up to 1980, decrease until 1990, and then increase again. This result is not surprising, since 1980s was a time when income inequality increased (see Figure 1a and Table 4) but the stock market did not suffer. On the other hand, the log(P/D) and log(P/E) specifications deteriorate after 1970. This finding is consistent with Welch and Goyal (2008), who document that most of the prediction gains stem from the 1973–1975 Oil Shock.

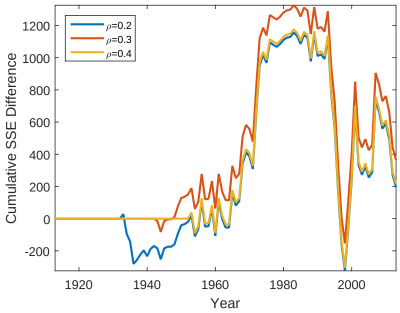
In summary, the top income series seem to predict returns out-of-sample.



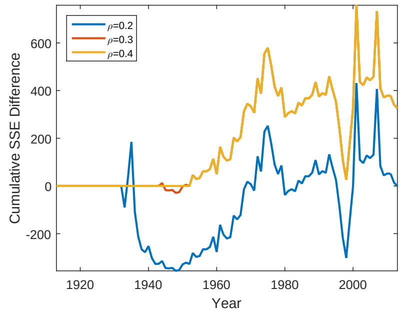
(a) cgdiff(1%).



(b) cgdiff(10%).



(c) log(P/D).



(d) log(P/E).

Figure 6: Annual performance in predicting subsequent excess returns.

Note: The figures plot the out-of-sample performance of annual predictive regressions. The vertical axis is the cumulative squared prediction errors of the NULL model minus the cumulative squared prediction error of the ALTERNATIVE model (hence a positive value favors the ALTERNATIVE). The NULL model uses only a constant. The ALTERNATIVE model includes the predictor variables specified in each subcaption. Predictions start at $t = \lfloor \rho T \rfloor$, where T is the sample size and $\rho = 0.2, 0.3, 0.4$.

4 Using tax policy as instrument

The top 1% income share is an endogenous variable in the macroeconomy. While in Section 3.2 we showed that top income shares are not simply proxying for GDP growth, volatility, the consumption/wealth ratio, or the level of the stock market in explaining subsequent returns, it is difficult to rule out the possibility that omitted variables are leading to endogeneity bias.

Fortunately, research on inequality (Roine et al., 2009; Kaymak and Poschke, 2016) suggests that increases (decreases) in top marginal tax rates reduce (exacerbate) inequality. Indeed, the Piketty-Saez series appear to exhibit a U-shaped trend over the century, which might be due to the change in the marginal income tax rates. According to Figure 7, the marginal tax rate for the highest income earners increased from about 25% to 90% over the period 1930–1945 and started to decline in the 1960s, reaching about 40% in the 1980s. Thus the marginal tax rate exhibits an inverse U-shape that seems to coincide with the trend in the Piketty-Saez series.

Furthermore, top tax rate changes are the result of Congressional bills, which generally take years to pass and usually stem from wars or pro-long-term growth

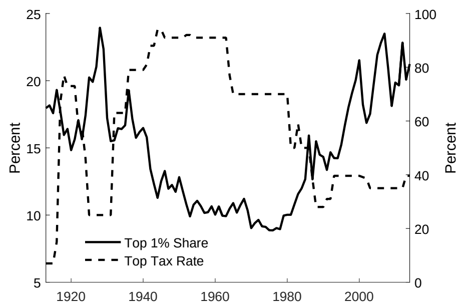


Figure 7: Top 1% income share including capital gains (left axis) and top marginal tax rate (right axis), 1913-2014. Source: IRS.

or anti-deficit ideologies (de Rugy, 2003a,b; Jacobson et al., 2007; Weinzierl and Werker, 2009; Romer and Romer, 2010). Therefore, while alterations in top tax rates impact inequality, their timing and justification are likely not the result of financial market fluctuations. Provided top tax rate changes have a muted effect on returns, except via inequality, they can serve as an instrument for top income shares. We address this "excludability" condition below.

In Section 4.1, we describe how inequality, stock prices, and subsequent returns fluctuated in tax cut and tax hike episodes. The data yield a narrative consistent with a causal link between inequality and subsequent excess returns. In Section 4.2, we calculate GMM estimates using changes in top estate and capital gains tax rates as instruments. The results suggest both that `cgdiff` actually reflects inequality (vs. capital gains timing) and that rising inequality *causes* low subsequent excess returns.

4.1 Tax change episodes and returns

We examine how periods of changing tax rates have affected top income shares, stock prices, and subsequent returns. We identify seven periods in U.S. history in which top income tax and estate rates were either rising or falling. Each period starts the year before the first tax change became effective and ends the year after the last change. Table 4 shows how top tax rates, the top 1% income share, and Robert Shiller's P/E10 ratio[19] evolved over each of these periods and provides the five year excess return starting in the final year of the period.

Each of the three tax increase periods (1915-1919, 1931-1945, and 1990-1994) was accompanied by a decline in the 1% income share (-0.24% per year, averaging across the periods, or around -1.44% for a typical 6 year episode). And, in line with our theory, each period was followed by 5 years of positive excess returns on average. The five year average excess returns (annualized) starting in 1919, 1945, and 1994 were, respectively, 2.78%, 8.61%, and 17.83%. In contrast, the tax cut periods (1921-1927, 1963-1966, 1980-1989, and 2000-2004) led to an increase in the top 1% share of 0.36% per year on average (2.16% for a 6 year episode) and an average subsequent five year excess return (annualized) of -2.88%. In tax cut periods, when top income shares rose, Shiller's

[19]Shiller (2000) argues that P/E10, the ratio of the S&P 500 index price to the 10 year average of earnings, is a measure of stock market over-valuation. The series is available at `http://www.econ.yale.edu/~shiller/data.htm`

P/E increased on average by 6.05% per year. In tax hike periods, the P/E ratio was flat on average. In summary, tax cut periods have been associated with increasing concentration of income, rising stock prices, and low subsequent excess stock returns. Tax hike periods have been times of falling inequality, low stock price growth, and higher subsequent excess returns.

However, to interpret these excess return fluctuations as the result of redistribution from the taxation of the rich, one must believe that top tax rates do not affect returns in other ways. Since we are looking at pre-tax returns, one possibility is that tax rate shocks directly impact returns by changing the after-tax dividend yield. To address this concern, we also consider after-tax returns, applying the top marginal tax rate to both dividends and interest. In Table 4, we see that doing so has an only negligible effect on five year returns: intuitively, most of the variation in excess returns stems from stock price movements and not from dividends or interest, the components impacted by income taxes.

A second "excludability" concern is that top tax rate changes may stimulate or contract the overall economy. Perhaps our tax changes are simply proxying for economic growth, which can affect stock and bond markets. For example, a tax cut could stimulate household income/demand, leading to higher stock prices and lower subsequent returns. In Table 4, however, we see that average per year growth in U.S. industrial production was actually higher on average in hike periods than in cut periods (6.12% vs. 4.37%).[20] Indeed, while industrial production boomed during the 20's and 60's tax cuts, it was stagnant during the early 2000's cuts. Conversely, average growth was a reasonable 2.42% during the 90's tax increases and a strong 7.18% on average over 1931-1945.

[20]We use industrial production because, unlike GDP, it almost spans our entire sample. FRED does not provide industrial production for the 1915-1919 period. Also, recall that in Table 15 the 1% share strongly predicts excess returns even when controlling for GDP growth.

Table 4: Top 1% income share and stock prices during and after top tax rate change episodes

Period	ΔMTR	ΔETR	Δ1%	ER$_{5yr}$	ER$_{5yr}^{a.t.}$	%Δ(P/E10)	%ΔIP	Reason
1915-1919	66	25	-0.36	2.78	2.54	-17.74	N/A	WWI
1921-1927	-48	-5	0.65	-8.65	-8.86	18.07	8.44	pro-growth
1931-1945	69	57	-0.27	8.61	4.70	3.86	7.18	WWII, budget balance
1963-1966	-21	0	0.19	-3.39	-1.00	-1.93	8.17	pro-growth
1980-1989	-42	-15	0.52	7.38	8.22	6.99	2.27	Reaganomics
1990-1994	11.6	0	-0.04	17.83	19.06	5.93	2.42	budget balance
2000-2004	-4.6	-7	-0.29	-6.86	-6.25	-8.01	0.12	pro-growth, stimulus
Across episode averages								
Hikes			-0.24	9.74	8.77	0.00	6.12	
Cuts			0.36	-2.88	-1.97	6.05	4.37	

Note: MTR and ETR: top marginal income and estate tax rates (%). Δ1%: average per year change in top 1% income share including capital gains. ER$_{5yr}$: annualized five year average excess return (%), starting in final year of period. ER$_{5yr}^{a.t.}$: ER$_{5yr}$, taxing interest and dividends at top marginal income rate. %Δ(P/E10): average per year % change in Shiller's P/E. %ΔIP: average per year % change in the industrial production index. Sources: de Rugy (2003a,b), Jacobson et al. (2007), Weinzierl and Werker (2009), Romer and Romer (2010),Tax Foundation, IRS, and FRED.

4.2 Instrumental variables regressions using changes in top estate tax rates

In this section we formally address the causality from inequality to the equity premium by instrumental variables regressions. So far we have assumed that `cgdiff` is a measure of inequality due to variation in capital income, but other interpretations are possible. For example, `cgdiff` may be varying due to the timing of realizing capital gains.

To address this issue, let `cgdiff` in year t be denoted by x_t, and suppose that it can be decomposed as

$$x_t = \alpha + x_{1t} + x_{2t},$$

where α is a constant and x_{1t}, x_{2t} are zero mean variables that reflect inequality and timing (an incentive to realize capital gains), respectively. Consider the model

$$R_{t+1} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_{t+1}, \tag{4.1}$$

where R_{t+1} is the excess stock return from year t to $t+1$. (For notational simplicity we are omitting additional control variables, but it is straightforward to include them.) We are interested in testing $\beta_1 = 0$. The problem is that x_{1t}, x_{2t} are not observed separately.

To identify β_1, suppose that there is an instrument z_{1t} for x_{1t}, so (i) z_{1t} is exogenous (uncorrelated with ε_{t+1}), (ii) z_{1t} is correlated with x_{1t}, and, furthermore, (iii) z_{1t} is uncorrelated with x_{2t}. Then it follows that

$$\begin{aligned} 0 = \mathrm{E}[z_{1t}\varepsilon_{t+1}] &= \mathrm{E}[z_{1t}(R_{t+1} - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t})] \\ &= \mathrm{E}[z_{1t}(R_{t+1} - \beta_0 - \beta_1 (x_t - \alpha - x_{2t}) - \beta_2 x_{2t})] \\ &= \mathrm{E}[z_{1t}(R_{t+1} - \alpha_1 - \beta_1 x_t)], \end{aligned} \tag{4.2}$$

where $\alpha_1 = \beta_0 + \alpha\beta_1$ and we have used $\mathrm{E}[z_{1t}x_{2t}] = 0$. Therefore even if the true inequality measure x_{1t} is unobserved, we can identify the coefficient of interest β_1 by exploiting the moment condition (4.2).

In line with our findings in Section 4.1, both Piketty and Saez (2003) and Piketty (2003) argue that income inequality should decline in response to expansion of progressive estate taxation: capital gains comprise a substantial portion of the income of the rich, and high estate taxes decrease the ability and incentive to amass wealth in financial assets. Thus, increasing the top estate tax rate should disproportionately reduce the wealth of the very rich and subsequently mitigate capital gains income inequality, which is driven by inequality in asset holdings. On the other hand, since estate taxes apply to both realized and unrealized capital gains, it is unlikely that estate taxes affect the timing of realizing capital gains beyond their incentive effects. Therefore current and lagged changes in the estate tax rates are a good candidate for an instrument. The first stage regressions in Table 5 confirm this hypothesis: contemporaneous and lagged changes in the top estate tax rate significantly explain a substantial portion of the variation in `cgdiff` (and the 10% and 0.1% analogs).

Table 5 suggests that changes in top estate tax rates can instrument for `cgdiff` in explaining excess returns. Whether one believes this instrument can test causation depends on if lagged changes in estate tax rates are excludable or not. One concern is that estate tax cuts stimulate the economy and thus

Table 5: Regressions of `cgdiff` (top income share − top income share (no cg)) on contemporaneous and lagged changed in top estate tax rates

	Dependent Variable: `cgdiff` (t)		
Regressors	0.1%	1%	10%
Constant	1.39	1.96	1.82
ΔETR_t	-0.04***	-0.05***	-0.04***
ΔETR_{t-1}	-0.02**	-0.03*	-0.03*
ΔETR_{t-2}	-0.07***	-0.08***	-0.06***
ΔETR_{t-3}	-0.05***	-0.06***	-0.05***
R^2	0.29	0.27	0.22

Note: the table shows regressions of `cgdiff` on lagged changes in top estate tax rates (ETR). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants) according to Newey-West standard errors ($k = 4$). For $y = 0.1\%, 1\%, 10\%$ `cgdiff` is the difference between the top y% income shares, with and without capital gains. Sample: 1913-2014. Sources: Tax Foundation and IRS.

stock market returns. Another concern is that even if estate tax rates only affect inequality, inequality may simply be proxying for the level of stock market, which we already know predicts returns. To control for these possibilities, we allow `cgdiff`, industrial production growth, and log(P/E) to be endogenous and instrument all three with contemporaneous and three lags of the change in the top estate tax rate (ΔETR for $t, t-1, t-2, t-3$) as well as the lagged price-earnings ratio ($\log(\text{P/E})_{t-1}$).[21] Table 6 shows the results of GMM estimation of the moment condition (4.2) (including industrial production growth and log(P/E) as controls).

Including industrial production growth and log(P/E) as endogenous regressors and using contemporaneous and three lags of changes in the top estate tax rate as well as the lagged log(P/E) as instruments, `cgdiff` is significant at the 5% level in predicting subsequent excess returns. This relationship holds regardless of whether we use the top 0.1%, 1%, or 10% income share in constructing `cgdiff`.

Our theory suggests that inequality predicts returns, but it does not say anything about the timing of realizing capital gains. Can we identify the coefficient β_2 in (4.1)? Suppose that there is an additional instrument z_{2t} for x_{2t} that is uncorrelated with x_{1t}. By the same argument as the derivation of (4.2), we can show that the moment condition

$$\mathrm{E}[z_{2t}(R_{t+1} - \alpha_2 - \beta_2 x_t)] = 0 \tag{4.3}$$

holds, where $\alpha_2 = \beta_0 + \alpha\beta_2$. What would be a good candidate for z_{2t}? Rational agents have an incentive to realize (delay) capital gains if they expect the capital gains tax rate to increase (decrease). Since tax rates in year $t+1$ are announced in year t, we can use the change in the maximum capital gains tax rate from year t to $t+1$, ΔCGTR_{t+1}, as an instrument z_{2t} for the timing component of `cgdiff`, x_{2t}. Table 7 adds ΔCGTR_{t+1} to the first-stage regressions displayed in Table 5. As we predicted, the change in top capital gains tax rates from

[21] Industrial production growth (t) is significantly correlated with ΔETR for $t, t-1$; $\log(\text{P/E})_t$ is significantly correlated with $\log(\text{P/E})_{t-1}$. Hence the rank condition for identification holds.

Table 6: Instrumental variables GMM estimates of the effect of `cgdiff` (top income share − top income share (no cg)), industrial production growth, and log(P/E) on one year excess stock market returns

Regressors (t)	Dependent Variable: t to $t+1$ Excess Market Return (1)	(2)	(3)
Constant	16.23 (26.97)	23.55 (26.35)	37.19 (27.07)
`cgdiff`(0.1%)	-14.39** (5.95)		
`cgdiff`(1%)		-11.26** (4.79)	
`cgdiff`(10%)			-14.89** (6.31)
%ΔIP	-1.61*** (0.54)	-1.57*** (0.52)	-1.54*** (0.49)
log(P/E)	6.45 (11.48)	4.56 (11.36)	1.44 (11.50)
J statistic	0.52 ($p = 0.77$)	0.57 ($p = 0.75$)	0.64 ($p = 0.72$)

Note: the table shows the results of two-step GMM estimation of the moment condition (4.2) (including industrial production growth and log(P/E) as controls). The initial weighting matrix is identity, and the second stage one is Newey-West ($k = 4$). Newey-West standard errors are in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). For $y = 0.1\%, 1\%, 10\%$ `cgdiff` is the difference between the top y% income shares, with and without capital gains. %ΔIP is the annual % change in the industrial production index. P/E is the S&P500 price-earnings ratios. The instruments are a constant, changes in the top estate tax rate (ΔETR for $t, t-1, t-2, t-3$), and the lagged price-earnings ratio ($\log(\mathrm{P/E})_{t-1}$). Sample: 1913-2014. Sources: Tax Foundation, IRS, and FRED.

year t to $t+1$ have positive and significant relationship with year t `cgdiff`. Current and lagged changes in estate tax rates, however, continue to have a strong inverse association with `cgdiff`. As rising capital gains and estate tax rates should, all else equal, discourage wealth accumulation amongst the rich, the positive coefficient on $\Delta\mathrm{CGTR}_{t+1}$ is likely reflecting the timing component of `cgdiff` (x_{2t}): when the rich expect capital gains taxes to rise, they move forward the realization of capital gains, which causes `cgdiff` to rise.

Thus, in Table 8 we jointly estimate the moment conditions (4.2) and (4.3) (including industrial production growth and log(P/E) as controls) by multiple equation GMM using the instruments

$$z_{1t} = (1, \Delta\mathrm{ETR}_t, \Delta\mathrm{ETR}_{t-1}, \Delta\mathrm{ETR}_{t-2}, \Delta\mathrm{ETR}_{t-3}, \log(\mathrm{P/E})_{t-1})',$$
$$z_{2t} = (1, \Delta\mathrm{CGTR}_{t+1})',$$

respectively. The coefficients are positive and insignificant for the timing components (x_{2t}) identified by changes in future capital gains tax rates. The inequality components (x_{1t}), however, have negative and significant coefficients. This is true regardless of whether we use the top 0.1%, 1%, or 10% income share, and

Table 7: Regressions of `cgdiff` (top income share – top income share (no cg)) on contemporaneous and lagged changed in top estate tax rates and the one-period-ahead change in the captial gains tax rate

	Dependent Variable: `cgdiff` (t)		
Regressors	0.1%	1%	10%
Constant	1.39	1.96	1.82
ΔETR_t	-0.04***	-0.05***	-0.04***
ΔETR_{t-1}	-0.04***	-0.05***	-0.04**
ΔETR_{t-2}	-0.06***	-0.08***	-0.05***
ΔETR_{t-3}	-0.06***	-0.07***	-0.04***
ΔCGTR_{t+1}	0.03***	0.04***	0.03***
R^2	0.34	0.31	0.26

Note: the table shows regressions of `cgdiff` on lagged changes in top estate tax rates (ETR) and the one-period-ahead change in the maximum capital gains tax rate (CGTR). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants) according to Newey-West standard errors ($k = 4$). For $y = 0.1\%, 1\%, 10\%$ `cgdiff` is the difference between the top y% income shares, with and without capital gains. Sample: 1913-2014. Sources: Tax Foundation and IRS.

suggests that the causal effect of `cgdiff` on subsequent excess returns is driven by inequality rather than by the timing of capital gains realization.

In summary, our finding that rising top income shares lead to low subsequent excess returns is robust to instrumenting inequality with changes in estate tax rates, even when controlling for economic growth and the level of the stock market. Introducing one-period-ahead capital gains tax rate changes as an additional instrument, we are able to separately identify how the inequality and timing components of `cgdiff` impact returns. The predictive power of `cgdiff` established in Section 3 appears driven by the inequality component.

5 International evidence

Thus far, we have shown that in the U.S. shocks to the concentration of income are associated with large and significant declines in subsequent excess returns on average. We have also provided a theoretical explanation for this pattern: if the rich are relatively more risk tolerant, when their wealth share rises relative aggregate demand for risky assets increases, which in equilibrium leads to a decline in the equity premium. Our theoretical argument, however, is not specific to the U.S. Therefore, we can test our theory by seeing whether or not this pattern holds internationally. In this section, we employ cross country fixed effects panel regressions and show that outside of the U.S. there also appears to be an inverse relationship between inequality and subsequent excess returns.

5.1 Data

We consider 29 countries, for the time period 1969-2013, spanning the continents: Americas (Argentina, Canada, Colombia, U.S.), Europe (Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and U.K.), Africa (Mauritius and South Africa), Asia

Table 8: Instrumental variables multiple equation GMM estimates of the effect of `cgdiff` (top income share − top income share (no cg)), industrial production growth, and log(P/E) on one year excess stock market returns

Regressors (t)	Dependent Variable: t to $t+1$ Excess Market Return (1)	(2)	(3)
Constant (α_1)	16.23	23.55	37.19
	(26.97)	(26.35)	(27.07)
Constant (α_2)	-20.47	-15.58	-8.15
	(41.12)	(41.58)	(42.32)
`cgdiff`(0.1%) (inequality, β_1)	-14.39**		
	(5.95)		
`cgdiff`(0.1%) (timing, β_2)	13.57		
	(11.50)		
`cgdiff`(1%) (inequality, β_1)		-11.26**	
		(4.79)	
`cgdiff`(1%) (timing, β_2)		9.55	
		(8.86)	
`cgdiff`(10%) (inequality, β_1)			-14.89**
			(6.31)
`cgdiff`(10%) (timing, β_2)			10.66
			(10.75)
%ΔIP	-1.61***	-1.57***	-1.54***
	(0.54)	(0.52)	(0.49)
log(P/E)	6.45	4.56	1.44
	(11.48)	(11.36)	(11.50)
J statistic	0.52	0.57	0.64
	($p = 0.77$)	($p = 0.75$)	($p = 0.72$)

Note: the table shows the results of two-step multiple equation GMM estimation of the moment conditions (4.2) and (4.3) (including industrial production growth and log(P/E) as controls). The initial weighting matrix is identity, and the second stage one is Newey-West ($k = 4$). Newey-West standard errors are in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). For $y = 0.1\%, 1\%, 10\%$ `cgdiff` is the difference between the top y% income shares, with and without capital gains. x_{1t} and x_{2t} refer to, respectively, the inequality and timing components of `cgdiff`. %ΔIP is the annual % change in the industrial production index. P/E is the S&P500 price-earnings ratios. The instruments for moment condition (4.2) are a constant, changes in the top estate tax rate (ΔETR for $t, t-1, t-2, t-3$), and the lagged price-earnings ratio ($\log(\mathrm{P/E})_{t-1}$). The instruments for moment condition (4.3) are a constant and $\Delta\mathrm{CGTR}_{t+1}$ (the one-period-ahead change in the maximum capital gains tax rate). Sample: 1913-2014. Sources: Tax Foundation, IRS, and FRED.

(China, India (INI), Japan, Singapore, South Korea, Malaysia, and Taiwan), and Oceania (Australia, Indonesia (INO), and New Zealand). Due to missing data points for some countries, we have around 100-800 observations, depending on the regions included. In the regressions below, we divide the countries into the following groups: Advanced Economies ("Advanced") (AUS, CAN, DNM, FIN, FRA, GER, IRE, ITA, JPN, KOR, NET, NOR, NZL, POR, SIN, SPA, SWE, SWI, TAI, UNK, and USA), IIPS (IRE, ITA, POR, and SPA), and EME

(ARG, CHN, COL, INI, INO, MAL, MAU, and SAF).

Our panel data on inequality are from Alvaredo et al. (2015). To be consistent across countries, we use top 1% income shares excluding capital gains in most specifications. The one exception is Table 12, in which we consider `cgdiff`, the difference between the top 1% share with and without capital gains income. Due to data limitations, this restricts our sample to Canada, Germany, Japan, and the U.S. See Appendix D for country-specific details on top income shares.

To calculate annual stock returns (end-of-period) we acquire from Datastream the MSCI total return indexes in local currency. To convert returns into local real terms, we deflate the stock indexes by local CPI (or GDP deflator when CPI is unavailable), which we obtain from Haver's IMF data. See Appendix D for country-specific details on stock market and price indexes.

Given the liquidity and safety of U.S. Treasuries, T-bill returns provide a standard and relatively uncontroversial measure of the risk-free rate in the U.S. In markets outside of the U.S., especially emerging ones where government and private sector default are not uncommon, it is not immediately obvious how to measure the risk-free rate. To make the definition of excess returns relatively consistent across countries, we use the Haver/IMF "deposit rate" series (in most cases), which is, depending on the country, the savings rate offered on one to twenty-four month deposits. Specifically, we take the year t safe return to be the average of annualized rates quoted in January to September of that year. Local nominal rates are converted into real terms by local CPI (or GDP deflator when CPI is unavailable). See Appendix D for more details.

5.2 International regression results

In Section 3, we showed that income concentration is inversely related to subsequent *excess* returns. However, quantitatively, this result was really about stock returns. Indeed, redoing column (1) of Table 1 with stock returns instead of excess returns, the 1% coefficient is -2.93 with a Newey-West p-value of 0.041. With the Kalman and HP filter, the coefficients are, respectively, -2.20 and -4.57 with p-values of 0.085 and 0.002. Also, with none of our top income share measures do we find a significant relationship between inequality and risk-free rates in the U.S. Furthermore, due to the limited availability of similar interest rates across countries, using stock returns instead of excess returns substantially expands the sample size. In light of these facts and because of the nebulous nature of international risk-free rates, we first present the international results for stock market returns without netting out an interest rate.

Another difference from our U.S. analysis in Section 3 is that in the post-1969 sample there is no obvious U-shape for top income shares, which simplifies handling the potentially nonstationary nature of inequality. In this section, we simply include a linear time trend as one of regressors (except in Table 12, where we use `cgdiff`).

Table 9 presents the panel regression results for both the whole sample and different regions. First, we see in the column "All" that when including all countries a one percentage point increase above trend in the top income share is associated with a subsequent decline in stock market returns of 2% on average. The coefficient is significant at the 5% level with standard errors clustered by country (results are similar without clustering). Columns "IIPS" and "EME" show that this inverse relationship is even stronger when we restrict the sample to the "GIIPS" (without Greece) or the emerging market economies. The pattern is weaker in the more advanced economies.[22]

Table 9: Country fixed effects panel regressions of one year stock market returns on top income shares

	Dependent Variable: t to $t+1$ Stock Market Return			
Regressors	All	Advanced	IIPS	EME
Top 1% (t)	-1.99**	-1.41^{+}	-7.16*	-6.58**
	(0.92)	(0.85)	(3.01)	(2.25)
Time Trend	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Obs.	790	699	106	91
R^2 (w,b)	(.01,.08)	(.00,.07)	(.05,.16)	(.05,.15)

Note: Clustered standard errors in parentheses, ***1%, **5%, *10%, $^{+}$15%. R^2 (w,b): Within and between R-squared. Constants suppressed. Top 1% is the pre-tax share of income going to the top 1% of earners (excluding capital gains). The column headings refer to the countries included (see the main text for details). Sample: 1969-2013 (see Appendix D for country details).

As a robustness check, Table 19 in Appendix E shows the panel regressions without time trends. The results are similar to the case with the linear time trend.[23]

Table 10 is the same as Table 9 except with *excess* returns (using real deposit rates) as the dependent variable. For "EME" countries, the results are essentially unchanged. Including all countries, the 1% coefficient falls in magnitude slightly to -1.49 but remains significant at the 10% level without clustering standard errors (with country clustering, the p-value is 0.16).

In Tables 1, 9, 10, and 19 we see that the relationship between inequality and returns is most apparent in the U.S. and emerging markets. One potential explanation for this finding is variation in the degree of stock market home bias. In either very large markets (such as the U.S.) or relatively closed ones (such as emerging markets), our theory suggests that local inequality should impact domestic stock markets. In small open markets, however, foreigners own a substantial fraction of the domestic stock markets and mitigate the role of local inequality. Indeed, according to measures in Mishra (2015), many of our "EME" countries (such as India, Indonesia, Colombia, and Malaysia) exhibit some of the highest degrees of home bias, while most of our "Advanced" and "IIPS" members are in the bottom half of countries ranked by home bias. Averaging his measures, Italy, the Netherlands, Singapore, Portugal, and Norway have the lowest home bias, and the Philippines, India, Turkey, Indonesia, and Pakistan have the highest (with Colombia and Malaysia close behind).

[22] Does including the time trend mitigate potential nonstationarity? The answers appears to be yes: the Phillips-Perron test (Phillips and Perron, 1988) rejects the presence of a unit root in the fitted residuals for each country (at least at the 5% level) except in Argentina (p-value of 0.31), Indonesia (p-value of 0.31), and South Africa (p-value of 0.052), all three of which have small sample size (≤ 12).

[23] And, somewhat surprisingly, the unit root tests on the residuals have the same results as with the inclusion of the time trend: we only fail to reject a unit root in Argentina, Indonesia, and South Africa, all of which are short time series.

Table 10: Country fixed effects panel regressions of one year excess returns on top income shares

Regressors	Dependent Variable: t to $t+1$ Excess Market Return			
	All	Advanced	IIPS	EME
Top 1% (t)	$-1.49^{\dagger}$	-0.63	-1.42	-6.55**
	(1.04)	(0.91)	(4.56)	(2.69)
Time Trend	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Obs.	660	569	72	91
R^2 (w,b)	(.00,.01)	(.00,.20)	(.01,.20)	(.05,.09)

Note: Clustered standard errors in parentheses, ***1%, **5%, *10%. R^2 (w,b): Within and between R-squared. Constants suppressed. †: p-value = 0.16, significant at 10% level without clustering. Top 1% is the pre-tax share of income going to the top 1% of earners (excluding capital gains). The column headings refer to the countries included (see the main text for details). Sample: 1969-2013 (see Appendix D for country details).

While local inequality appears less important in small and open financial markets, inequality amongst global investors should still impact excess returns in these markets. Table 11 repeats the regressions of Table 10 but also includes the U.S. 1% share as a proxy for global investor inequality. As conjectured, the U.S. 1% share has a large and significant inverse correlation with subsequent excess returns for the "Advanced" and "IIPS" groups (small open economies), and the local 1% share is significant for emerging markets (relatively closed economies).

Table 11: Country fixed effects panel regressions of one year excess returns on local and U.S. top income shares

Regressors	Dependent Variable: t to $t+1$ Excess Market Return			
	All†	Advanced†	IIPS	EME
Top 1% (t)	-1.62	-0.65	-2.53	-6.16*
	(1.17)	(1.07)	(4.20)	(2.88)
U.S. Top 1% (t)	-3.37***	-2.67***	-6.44**	-11.96
	(1.02)	(0.73)	(1.74)	(8.32)
Time Trend	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Obs.	616	525	72	91
R^2 (w,b)	(.02,.01)	(.01,.13)	(.04,.00)	(.09,.10)

Note: Clustered standard errors in parentheses, ***1%, **5%, *10%. †: excluding U.S. R^2 (w,b): Within and between R-squared. Constants suppressed. Top 1% is the pre-tax share of income going to the top 1% of earners (excluding capital gains). The column headings refer to the countries included (see the main text for details). Sample: 1969-2013 (see Appendix D for country details).

Lastly, in Table 12 we restrict the sample to the U.S., Japan, Canada, and Germany (the only countries for which `cgdiff` is available) and use `cgdiff` to predict returns. In the first two columns, we see that when `cgdiff` rises by one percentage point, both subsequent one year returns and excess returns significantly fall by about 2%.[24] In the last two columns, we exclude the U.S. and include U.S. `cgdiff` as a regressor. The coefficient for U.S. `cgdiff` is negative but not significant, while local `cgdiff` remains similar in magnitude and significance.

Table 12: Country fixed effects panel regressions of one year returns on local and U.S. `cgdiff` (top 1% − top 1% (no cg))

	Dependent Variable: t to $t+1$ Return			
Regressors	R	$R-R_f$	$R^\dagger$	$R-R_f^\dagger$
`cgdiff` (t)	-2.19***	-1.81*	-2.24**	-1.25*
	(0.31)	(0.70)	(0.56)	(0.32)
U.S. `cgdiff` (t)			-2.87	-2.86
			(1.55)	(1.36)
Time Trend	No	No	No	No
Country FE	Yes	Yes	Yes	Yes
Obs.	143	135	99	91
R^2 (w,b)	(.01,.04)	(.01,.70)	(.04,.48)	(.03,.63)

Note: Clustered standard errors in parentheses, ***1%, **5%, *10%. The column heading R ($R-R_f$) denotes that the dependent variable is the stock market return (excess stock market return). Countries included: CAN, JPN, GER, USA. †: excluding U.S. R^2 (w,b): Within and between R-squared. Constants suppressed. "`cgdiff`" is top 1% minus top 1% (no cg), neither detrended, where top 1% is the pre-tax share of income going to the top 1% of earners (including capital gains). "no cg" refers to the series that excludes capital gains. Sample: 1969-2013 (see Appendix D for country details).

6 Concluding remarks

In this paper we built a general equilibrium model with agents that are heterogeneous in both wealth and attitudes towards risk or beliefs. We proved that the concentration of wealth/income drives down the subsequent equity premium. Our model is a mathematical formulation of Irving Fisher's narrative that booms and busts are caused by changes in the relative wealth of the rich (the "enterpriser-borrower") and the poor (the "creditor, the salaried man, or the laborer"). Consistent with our theory, we found that the income/wealth distribution is closely connected with stock market returns. When the rich are richer than usual the stock market subsequently performs poorly, both in- and out-of-sample.

Could one exploit the predictive power of top income shares to beat the market on average? The answer is probably no since the top income share—which comes from tax return data—is calculated with a substantial lag. One would receive the inequality update too late to act on its asset pricing information. However, our analysis provides a novel positive explanation of excess market returns over time. We conclude, as decades of macro/finance theory have suggested, that stock market fluctuations are intimately tied to the distribution of wealth, income, and assets.

[24]Unlike in the previous international regressions, when we restrict the sample to U.S., Japan, Canada, and Germany and use `cgdiff`, the 1% coefficients are only significant at the 10% level when clustering by country.

References

Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. The world top incomes database, 2015. URL `http://topincomes.parisschoolofeconomics.eu`.

Andrew Ang and Geert Bekaert. Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707, 2007. doi:10.1093/rfs/hhl021.

Kenneth J. Arrow. *Aspects of the Theory of Risk-Bearing*. Yrjö Jahnssonin Säätiö, 1965.

Pierluigi Balduzzi and Tong Yao. Testing heterogeneous-agent models: An alternative aggregation approach. *Journal of Monetary Economics*, 54(2): 369–412, March 2007. doi:10.1016/j.jmoneco.2005.08.021.

Ravi Bansal, Dana Kiku, Ivan Shaliastovich, and Amir Yaron. Volatility, the macroeconomy, and asset prices. *Journal of Finance*, 69(6):2471–2511, December 2014. doi:10.1111/jofi.12110.

Suleyman Basak and Domenico Cuoco. An equilibrium model with restricted stock market participation. *Review of Financial Studies*, 11(2):309–341, 1998. doi:10.1093/rfs/11.2.309.

Harjoat S. Bhamra and Raman Uppal. Asset prices with heterogeneity in preferences and beliefs. *Review of Financial Studies*, 27(2):519–580, 2014. doi:10.1093/rfs/hht051.

Lawrence Blume and David Easley. If you're so smart, why aren't you rich? belief selection in complete and incomplete markets. *Econometrica*, 74(4): 929–966, July 2006. doi:10.1111/j.1468-0262.2006.00691.x.

Jacob Boudoukh, Matthew Richardson, and Robert F. Whitelaw. The myth of long-horizon predictability. *Review of Financial Studies*, 21(4):1577–1605, 2008. doi:10.1093/rfs/hhl042.

Alon Brav, George M. Constantinides, and Christopher C. Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110(4):793–824, August 2002. doi:10.1086/340776.

Alessandro Bucciol and Raffaele Miniaci. Household portfolio and implicit risk preference. *Review of Economics and Statistics*, 93(4):1235–1250, 2011. doi:10.1162/REST_a_00138.

Laurent E. Calvet and Paolo Sodini. Twin picks: Disentangling the determinants of risk-taking in household portfolios. *Journal of Finance*, 69(2): 867–906, April 2014. doi:10.1111/jofi.12125.

John Y. Campbell. Household finance. *Journal of Finance*, 61(4):1553–1604, August 2006. doi:10.1111/j.1540-6261.2006.00883.x.

John Y. Campbell and Robert J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228, 1988. doi:10.1093/rfs/1.3.195.

John Y. Campbell and Samuel B. Thomson. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531, 2008. doi:10.1093/rfs/hhm055.

Sean D. Campbell, Stefanos Delikouras, Danling Jiang, and George M. Korniotis. The human capital that matters: Expected returns and high-income households. *Review of Financial Studies*, 29(9):2523–2563, 2016. doi:10.1093/rfs/hhw048.

Christopher D. Carroll. Portfolios of the rich. In Luigi Guiso, Michael Haliassos, and Tullio Jappelli, editors, *Household Portfolios*, chapter 10, pages 389–430. MIT Press, Cambridge, MA, 2002.

Yeung Lewis Chan and Leonid Kogan. Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy*, 110(6):1255–1285, December 2002. doi:10.1086/342806.

John S. Chipman. Homothetic preferences and aggregation. *Journal of Economic Theory*, 8(1):26–38, May 1974. doi:10.1016/0022-0531(74)90003-9.

John H. Cochrane. The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575, 2008. doi:10.1093/rfs/hhm046.

Timothy Cogley. Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics*, 49(2): 309–334, March 2002. doi:10.1016/S0304-3932(01)00106-4.

George M. Constantinides. Intertemporal asset pricing with heterogeneous agents without demand aggregation. *Journal of Business*, 55(2):253–267, April 1982.

George M. Constantinides and Darrell Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104(2):219–240, April 1996.

Jakša Cvitanić, Elyès Jouini, Semyon Malamud, and Clotilde Napp. Financial markets equilibrium with heterogeneous agents. *Review of Finance*, 16(1): 285–321, 2012. doi:10.1093/rof/rfr018.

Veronique de Rugy. Tax rates and tax revenue: The Mellon income tax cuts of the 1920s. *Cato Institute Tax and Budget Bulletin*, (13), February 2003a. URL `http://object.cato.org/sites/cato.org/files/pubs/pdf/tbb-0302-13.pdf`.

Veronique de Rugy. High taxes and high budget deficits: The Hoover-Roosevelt tax increases of the 1930s. *Cato Institute Tax and Budget Bulletin*, (14), March 2003b. URL `http://object.cato.org/sites/cato.org/files/pubs/pdf/tbb-0303-14.pdf`.

Bernard Dumas. Two-person dynamic equilibrium in the capital market. *Review of Financial Studies*, 2(2):157–188, 1989. doi:10.1093/rfs/2.2.157.

Eugene F. Fama and Kenneth R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25, October 1988. doi:10.1016/0304-405X(88)90020-7.

Irving Fisher. *Introduction to Economic Science*. Macmillan, NY, 1910.

Nicolae Gârleanu and Stavros Panageas. Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing. *Journal of Political Economy*, 123(3):670–685, June 2015. doi:10.1086/680996.

John Geanakoplos. An introduction to general equilibrium with incomplete asset markets. *Journal of Mathematical Economics*, 19(1-2):1–38, 1990. doi:10.1016/0304-4068(90)90034-7.

John Geanakoplos and Martin Shubik. The capital asset pricing model as a general equilibrium with incomplete markets. *Geneva Papers on Risk and Insurance Theory*, 15(1):55–71, 1990. doi:10.1007/BF01498460.

John Geanakoplos and Kieran Walsh. Uniqueness and stability of equilibrium in economies with two goods. 2016.

Christian Gollier. Wealth inequality and asset pricing. *Review of Economic Studies*, 68(1):181–203, January 2001. doi:10.1111/1467-937X.00165.

William M. Gorman. Community preference fields. *Econometrica*, 21(1):63–80, January 1953. doi:10.2307/1906943.

Clive W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121–130, May 1981. doi:10.1016/0304-4076(81)90079-8.

Clive W. J. Granger and Paul Newbold. Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120, July 1974. doi:10.1016/0304-4076(74)90034-7.

Daniel L. Greenwald, Martin Lettau, and Sydney C. Ludvigson. Origins of stock market fluctuations. 2016.

Michael Haliassos and Carol C. Bertaut. Why do so few hold stocks? *Economic Journal*, 105(432):1110–1129, September 1995. doi:10.2307/2235407.

James D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

Peter Reinhard Hansen and Allan Timmermann. Equivalence between out-of-sample forecast comparisons and Wald statistics. *Econometrica*, 83(6): 2485–2505, November 2015. doi:10.3982/ECTA10581.

Chiaki Hara, James Huang, and Christoph Kuzmics. Representative consumer's risk aversion and efficient risk-sharing rules. *Journal of Economic Theory*, 137 (1):652–672, November 2007. doi:10.1016/j.jet.2006.11.002.

Juan Carlos Hatchondo. A quantitative study of the role of wealth inequality on asset prices. *Federal Reserve Bank of Richmond Economic Quarterly*, 94 (1):73–96, 2008.

John Heaton and Deborah J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy*, 104(3):443–487, June 1996.

Robert J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5 (3):357–386, 1992. doi:10.1093/rfs/5.3.351.

Darien B. Jacobson, Brian G. Raub, and Barry W. Johnson. The estate tax: Ninety years and counting. *SOI Bulletin*, 27(1):118–128, Summer 2007. URL `https://www.irs.gov/pub/irs-soi/ninetyestate.pdf`.

Timothy C. Johnson. Inequality risk premia. *Journal of Monetary Economics*, 59(6):565–580, October 2012. doi:10.1016/j.jmoneco.2012.06.008.

Barış Kaymak and Markus Poschke. The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics*, 77:1–25, February 2016. doi:10.1016/j.jmoneco.2015.10.004.

Timothy J. Kehoe. Uniqueness and stability. In Alan Kirman, editor, *Elements of General Equilibrium Analysis*, chapter 3, pages 38–87. Wiley-Blackwell, 1998.

Narayana R. Kocherlakota and Luigi Pistaferri. Asset pricing implications of Pareto optimality with private information. *Journal of Political Economy*, 117(3):555–590, June 2009. doi:10.1086/599761.

Wojciech Kopczuk and Emmanuel Saez. Top wealth shares in the United States, 1916–2000: Evidence from estate tax returns. *National Tax Journal*, 57(2): 445–487, 2004.

Martin Lettau and Sydney C. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849, June 2001. doi:10.1111/0022-1082.00347.

Martin Lettau and Sydney C. Ludvigson. Measuring and modeling variation in the risk-return trade-off. In Yacine Aït-Sahalia and Lars Peter Hansen, editors, *Handbook of Financial Econometrics*, volume 1, chapter 11, pages 617–690. Elsevier, Amsterdam, 2010. doi:10.1016/B978-0-444-50897-3.50014-6.

Martin Lettau, Sydney C. Ludvigson, and Jessica A. Wachter. The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies*, 21(4):1653–1687, 2008. doi:10.1093/rfs/hhm020.

John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37, February 1965a. doi:10.2307/1924119.

John Lintner. Security prices, risk, and maximal gains from diversification. *Journal of Finance*, 20(4):587–615, December 1965b. doi:10.1111/j.1540-6261.1965.tb02930.x.

Francis A. Longstaff and Jiang Wang. Asset pricing and the credit market. *Review of Financial Studies*, 25(11):3169–3215, 2012. doi:10.1093/rfs/hhs086.

Sydney C. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. In George M. Constantinides, Milton Harris, and René M. Stultz, editors, *Handbook of the Economics of Finance*, volume 2, chapter 12, pages 799–906. Elsevier, Amsterdam, 2013. doi:10.1016/B978-0-44-459406-8.00012-3.

N. Gregory Mankiw. The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics*, 17(1):211–219, September 1986. doi:10.1016/0304-405X(86)90012-7.

Rolf R. Mantel. Homothetic preferences and community excess demand functions. *Journal of Economic Theory*, 12(2):197–201, April 1976. doi:10.1016/0022-0531(76)90073-9.

Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, March 1952. doi:10.1111/j.1540-6261.1952.tb01525.x.

Michael W. McCracken. Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, 140(2):719–752, October 2007. doi:10.1016/j.jeconom.2006.07.020.

Anil V. Mishra. Measures of equity home bias puzzle. *Journal of Empirical Finance*, 34:293–312, December 2015. doi:10.1016/j.jempfin.2015.08.001.

Charles R. Nelson and Myung J. Kim. Predictable stock returns: The role of small sample bias. *Journal of Finance*, 48(2):641–661, June 1993. doi:10.1111/j.1540-6261.1993.tb04731.x.

Peter C. B. Phillips and Ji Hyung Lee. Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264, December 2013. doi:10.1016/j.jeconom.2013.04.011.

Peter C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988. doi:10.1093/biomet/75.2.335.

Thomas Piketty. Income inequality in France, 1901–1998. *Journal of Political Economy*, 111(5):1004–1042, October 2003. doi:10.1086/376955.

Thomas Piketty and Emmanuel Saez. Income inequality in the United States, 1913–1998. *Quarterly Journal of Economics*, 118(1):1–41, February 2003. doi:10.1162/00335530360535135.

David E. Rapach and Guofu Zhou. Forecasting stock returns. In Graham Elliott and Allan Timmerman, editors, *Handbook of Economic Forecasting*, volume 2, chapter 6, pages 328–383. Elsevier, 2013. doi:10.1016/B978-0-444-53683-9.00006-2.

David E. Rapach, Jack K. Strauss, and Guofu Zhou. Out-of-sample equity premium prediction: Combination forechasts and links to the real economy. *Review of Financial Studies*, 23(2), 821-862 2010. doi:10.1093/rfs/hhp063.

Jesper Roine, Jonas Vlachos, and Daniel Waldenström. The long-run determinants of inequality: What can we learn from top income data? *Journal of Public Economics*, 93(7-8):974–988, August 2009. doi:10.1016/j.jpubeco.2009.04.003.

Christina D. Romer and David H. Romer. The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, 100(3):763–801, June 2010. doi:10.1257/aer.100.3.763.

Emmanuel Saez and Gabriel Zucman. Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, 131(2):519–578, May 2016. doi:10.1093/qje/qjw004.

Alvaro Sandroni. Do markets favor agents able to make accurate prediction? *Econometrica*, 68(6):1303–1341, November 2000. doi:10.1111/1468-0262.00163.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, September 1964. doi:10.1111/j.1540-6261.1964.tb02865.x.

Robert J. Shiller. *Irrational Exuberance*. Princeton University Press, 2000.

Robert F. Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, December 1999. doi:10.1016/S0304-405X(99)00041-0.

Kjetil Storesletten, Chris I. Telmer, and Amir Yaron. Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics*, 10(4):519–548, October 2007. doi:10.1016/j.red.2007.02.004.

James Tobin. Liquidity preference as behavior towards risk. *Review of Economics and Statistics*, 25(2):65–86, February 1958. doi:10.2307/2296205.

Alexis Akira Toda and Kieran Walsh. The double power law in consumption and implications for testing Euler equations. *Journal of Political Economy*, 123(5):1177–1200, October 2015. doi:10.1086/682729.

Alexis Akira Toda and Kieran James Walsh. Fat tails and spurious estimation of consumption-based asset pricing models. *Journal of Applied Econometrics*, 2016a. doi:10.1002/jae.2564.

Alexis Akira Toda and Kieran James Walsh. Edgeworth box economies with multiple equilibria. *Economic Theory Bulletin*, 2016b. doi:10.1007/s40505-016-0102-3.

Rossen Valkanov. Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics*, 68(2):201–232, May 2003. doi:10.1016/S0304-405X(03)00065-5.

Annette Vissing-Jørgensen. Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. NBER Working Paper, 2002. URL `http://www.nber.org/papers/w8884`.

Jessica A. Wachter and Motohiro Yogo. Why do household portfolio shares rise in wealth? *Review of Financial Studies*, 23(11):3929–3965, 2010. doi:10.1093/rfs/hhq092.

Jiang Wang. The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics*, 41(1):75–110, May 1996. doi:10.1016/0304-405X(95)00854-8.

Matthew C. Weinzierl and Eric D. Werker. Barack Obama and the Bush tax cuts (A). Case 709-037, Harvard Business School, January 2009.

Ivo Welch and Amit Goyal. A compehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008. doi:10.1093/rfs/hhm014.

A Proofs

A.1 Proof of Theorem 2.1

Since by Assumption 3 the aggregate endowment e is spanned by the column vectors of A, without loss of generality we may assume $A = [e, A_2, \dots, A_J]$. Let $n_i = (w_i, 0, \dots, 0)'$ be the vector of initial endowment of assets. Then by Assumption 2 we have $e_i = w_i e = An_i$. Letting $z = y + n_i$, the budget constraint becomes $q'z \le q'n_i$ and $x \le Az$. Therefore the utility maximization problem becomes equivalent to

$$\begin{aligned} &\text{maximize} && U_i(x) \\ &\text{subject to} && q'z \le q'n_i, 0 \le x \le Az. \end{aligned} \tag{A.1}$$

Similarly, the planner's problem (2.2) is equivalent to

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{I} w_i \log U_i(Az_i) \\ &\text{subject to} && \sum_{i=1}^{I} z_i = n, \end{aligned} \tag{A.2}$$

where $n = \sum_{i=1}^{I} n_i = (1, 0, \dots, 0)'$ is the vector of aggregate endowment of assets.

Step 1. $\log U_i(x)$ *is strictly concave.*

Proof. Let us suppress the i subscript and define

$$f(x) = \log U(x) = \begin{cases} \frac{1}{1-\gamma} \log \left(\sum_{s=1}^{S} \pi_s x_s^{1-\gamma} \right), & (\gamma \neq 1) \\ \sum_{s=1}^{S} \pi_s \log x_s. & (\gamma = 1) \end{cases}$$

If $\gamma = 1$, then f is clearly strictly concave. If $\gamma \neq 1$, let $\Sigma = \sum_{s=1}^{S} \pi_s x_s^{1-\gamma}$. Then by simple algebra we have

$$\begin{aligned} \nabla f(x) =& \frac{1}{\Sigma} \begin{bmatrix} \pi_1 x_1^{-\gamma} \\ \vdots \\ \pi_S x_S^{-\gamma} \end{bmatrix}, \\ \nabla^2 f(x) =& -\frac{1-\gamma}{\Sigma^2} \begin{bmatrix} \pi_1 x_1^{-\gamma} \\ \vdots \\ \pi_S x_S^{-\gamma} \end{bmatrix} \begin{bmatrix} \pi_1 x_1^{-\gamma} & \cdots & \pi_S x_S^{-\gamma} \end{bmatrix} \\ &+ \frac{1}{\Sigma} \operatorname{diag} \begin{bmatrix} -\gamma \pi_1 x_1^{-\gamma-1} & \cdots & -\gamma \pi_S x_S^{-\gamma-1} \end{bmatrix}. \end{aligned}$$

To show that $\nabla^2 f(x)$ is negative definite, it suffices to show that $-\Sigma^2\nabla^2 f(x)$ is positive definite. To this end, let $h = (h_1, \dots, h_S)'$ be any vector. Then

$$h'[-\Sigma^2\nabla^2 f(x)]h = (1-\gamma)\left(\sum_{s=1}^{S} \pi_s x_s^{-\gamma} h_s\right)^2 + \gamma\left(\sum_{s=1}^{S} \pi_s x_s^{1-\gamma}\right)\left(\sum_{s=1}^{S} \pi_s x_s^{-\gamma-1} h_s^2\right).$$

Define $u, v \in \mathbb{R}^S$ by $u = (\cdots (\pi_s x_s^{1-\gamma})^{\frac{1}{2}} \cdots)'$ and $v = (\cdots (\pi_s x_s^{-\gamma-1})^{\frac{1}{2}} h_s \cdots)'$. Then the above expression becomes

$$\begin{aligned} h'[-\Sigma^2\nabla^2 f(x)]h &= (1-\gamma)(u \cdot v)^2 + \gamma \|u\|^2 \|v\|^2 \\ &= \gamma(\|u\|^2 \|v\|^2 - (u \cdot v)^2) + (u \cdot v)^2 \geq 0, \end{aligned}$$

where we have used the Cauchy-Schwarz inequality. Equality occurs when u, v are collinear and $u \cdot v = 0$. Since $u \neq 0$, this is true if and only if $v = ku$ for some k and $k\|u\|^2 = 0$, so $k = 0$ and therefore $h = 0$. Hence $f = \log U$ is strictly concave. □

Step 2. The planner's problem (2.2) *has a unique solution.*

Proof. Let

$$\Omega = \left\{ x = (x_i) \in \mathbb{R}_+^{SI} \,\middle|\, (\exists z = (z_i))(\forall i) x_i \leq A z_i, \sum_{i=1}^{I} z_i = n \right\}$$

be the set of all feasible consumption allocations. Then the planner's problem (A.2) is equivalent to maximizing $f(x) = \sum_{i=1}^{I} w_i \log U_i(x_i)$ subject to $x \in \Omega$. Clearly f is continuous, and by the previous step strictly concave. Therefore to show the uniqueness of the solution, it suffices to show that Ω is nonempty, compact, and convex. Clearly $\Omega \neq \emptyset$ because we can choose the initial endowment $z_i = n_i$ and $x_i = A n_i = e_i$. Since Ω is defined by linear inequalities and equations, it is closed and convex. If $x \in \Omega$, by definition we can take $z = (z_i)$ such that $x_i \leq A z_i$ for all i and $\sum_{i=1}^{I} z_i = n$. Then

$$\sum_{i=1}^{I} x_i \leq \sum_{i=1}^{I} A z_i = A \sum_{i=1}^{I} z_i = An = e.$$

Since $x_i \geq 0$ and $e \gg 0$, Ω is bounded.

Let $x = (x_i)$ be the unique maximizer of f on Ω. Since f is strictly increasing, we have $x_i = A z_i$ for some $z = (z_i)$ such that $\sum_{i=1}^{I} z_i = n$. If there is another such $z' = (z_i')$, then $A z_i = A z_i' \iff A(z_i - z_i') = 0$. Since by assumption A has full column rank, we have $z_i - z_i' = 0 \iff z_i = z_i'$. Therefore the planner's problem (A.2) has a unique solution. □

Step 3. $x = (x_i)$ is a GEI equilibrium allocation and the Lagrange multiplier to the planner's problem gives the asset prices.

Proof. Let

$$L(z, q) = \sum_{i=1}^{I} w_i \log U_i(A z_i) + q' \left(n - \sum_{i=1}^{I} z_i\right)$$

be the Lagrangian of the planner's problem (A.2). By the previous step, a unique solution $z = (z_i)$ exists. Furthermore, since U_i satisfies the Inada condition, it must be $Az_i \gg 0$. Hence by the Karush-Kuhn-Tucker theorem and the chain rule, we have

$$q' = w_i \frac{DU_i(Az_i)A}{U_i(Az_i)} \tag{A.3}$$

for all i, where DU_i denotes the $(1 \times S)$ Jacobi matrix of the function U_i. Since U_i is homogeneous of degree 1, for all $x \gg 0$ and $\lambda > 0$ we have $U_i(\lambda x) = \lambda U_i(x)$. Differentiating both sides with respect to λ and setting $\lambda = 1$, we have $DU_i(x)x = U_i(x)$. Hence multiplying z_i from the right to (A.3), we get

$$q' z_i = w_i \frac{DU_i(Az_i)Az_i}{U_i(Az_i)} = w_i.$$

Adding across i and using the complementary slackness condition, we get

$$q'n = q' \sum_{i=1}^{I} z_i = \sum_{i=1}^{I} w_i = 1.$$

Therefore

$$q' z_i = w_i = w_i q' n = q'(w_i n) = q' n_i,$$

so the budget constraint holds with equality. Furthermore, letting $\lambda_i = \frac{1}{w_i}$, by (A.3) we obtain $D[\log U_i(Az_i)] = \lambda_i q'$, which is the first-order condition of the utility maximization problem (A.1) after taking the logarithm. Since $\log U_i$ is concave, z_i solves the utility maximization problem. Since $\sum_{i=1}^{I} z_i = n$, the asset markets clear, so $\{q, (x_i), (z_i)\}$ is a GEI. $\square$

Step 4. The GEI is uniquely given as the solution to the planner's problem (A.2).

Proof. Let $\{q, (x_i), (z_i)\}$ be a GEI. By the first-order condition to the utility maximization problem, there exists a Lagrange multiplier $\lambda_i \geq 0$ such that

$$\lambda_i q' = D[\log U_i(Az_i)] = \frac{DU_i(Az_i)A}{U_i(Az_i)}. \tag{A.4}$$

Since $DU_i \gg 0$, $A = [e, A_2, \dots, A_J]$, and $e \gg 0$, comparing the first element of (A.4), we have

$$\lambda_i q_1 = \frac{DU_i(Az_i)e}{U_i(Az_i)} > 0.$$

Therefore $\lambda_i > 0$ and $q_1 > 0$. By scaling the price vector if necessary, we may assume $q_1 = 1$ and hence $q'n = 1 \cdot 1 + q_2 \cdot 0 + \dots + q_J \cdot 0 = 1$. Multiplying z_i to (A.4) from the right and using $DU_i(x)x = U_i(x)$ and the complementary slackness condition, we have

$$\lambda_i q' n_i = \lambda_i q' z_i = \frac{DU_i(Az_i)Az_i}{U_i(Az_i)} = 1 \iff \frac{1}{\lambda_i} = q' n_i = w_i q' n = w_i.$$

Substituting into (A.4), we obtain $q' = w_i D[\log U_i(Az_i)]$, which is precisely (A.3), the first-order condition of the planner's problem (A.2) with Lagrange multiplier q. Since (z_i) is feasible and the objective function is strictly concave, (z_i) is the unique solution to the planner's problem. $\square$

A.2 Proof of Theorem 2.2 and Propositions 2.3, 2.4

Let u be a general von Neumann-Morgenstern utility function with $u' > 0$ and $u'' < 0$. In Theorem 2.2, we have $u(x) = \frac{1}{1-\gamma} x^{1-\gamma}$ or $u(x) = \log x$, but most of the following results do not depend on the particular functional form. Then a typical agent's optimal portfolio problem is

$$\max_{\theta} \mathrm{E}[u(R(\theta)w)],$$

where w is initial wealth. The following lemma is basic (*e.g.*, Arrow, 1965).

Lemma A.1. *Let everything be as above and θ be the optimal portfolio. Then the following is true.*

1. *θ is unique.*
2. *$\theta \gtreqless 0$ according as $\mathrm{E}[R] \gtreqless R_f$.*
3. *Suppose $\mathrm{E}[R] > R_f$. If u exhibits decreasing relative risk aversion (DRRA), so $-xu''(x)/u'(x)$ is decreasing, then $\partial\theta/\partial w \geq 0$,* i.e., *the agent invests comparatively more in the risky asset as he becomes richer. The opposite is true if u exhibits increasing relative risk aversion (IRRA).*

Proof. 1. Let $f(\theta) = \mathrm{E}[u(R(\theta)w)]$. Then $f'(\theta) = \mathrm{E}[u'(R(\theta)w)(R-R_f)w]$ and $f''(\theta) = \mathrm{E}[u''(R(\theta)w)(R-R_f)^2 w^2] < 0$, so f is strictly concave. Therefore the optimal θ is unique (if it exists).

2. Since $f'(\theta) = 0$ and $f'(0) = u'(R_f w)w(\mathrm{E}[R] - R_f)$, the result follows.

3. Dividing the first-order condition by w, we obtain $\mathrm{E}[u'(R(\theta)w)(R-R_f)] = 0$. Let $F(\theta, w)$ be the left-hand side. Then by the implicit function theorem we have $\partial\theta/\partial w = -F_w/F_\theta$. Since $F_\theta = \mathrm{E}[u''(R(\theta)w)(R - R_f)^2 w] < 0$, it suffices to show $F_w \geq 0$. Let $\gamma(x) = -xu''(x)/u'(x) > 0$ be the relative risk aversion coefficient. Then

$$\begin{aligned} F_w &= \mathrm{E}[u''(R(\theta)w)(R - R_f)R(\theta)] \\ &= -\frac{1}{w} \mathrm{E}[\gamma(R(\theta)w)u'(R(\theta)w)(R - R_f)]. \end{aligned}$$

Since $\mathrm{E}[R] > R_f$, by the previous result we have $\theta > 0$. Therefore $R(\theta) = R\theta + R_f(1 - \theta) \gtreqless R_f$ according as $R \gtreqless R_f$. Since u is DRRA, γ is decreasing, so $\gamma(R(\theta)w) \leq \gamma(R_f w)$ if $R \geq R_f$ (and reverse inequality if $R \leq R_f$). Therefore

$$\gamma(R(\theta)w)(R - R_f) \leq \gamma(R_f w)(R - R_f)$$

always. Multiplying both sides by $-u'(R(\theta)w) < 0$ and taking expectations, we obtain

$$\begin{aligned} wF_w &= -\mathrm{E}[\gamma(R(\theta)w)u'(R(\theta)w)(R - R_f)] \\ &\geq -\mathrm{E}[\gamma(R_f w)u'(R(\theta)w)(R - R_f)] = 0, \end{aligned}$$

where the last equality uses the first-order condition. □

Proof of Theorem 2.2. Let θ_i be the optimal portfolio of agent i. By Lemma A.1, $\theta_i \gtrless 0$ according as $\mathrm{E}_i[R] \gtrless R_f$.

Suppose that $\theta_1 < \theta_2$ and we transfer some wealth $\epsilon > 0$ from agent 1 to 2. Let θ_i' be the new portfolio of agent i. The change in agent 1 and 2's demand in the risky asset is

$$\begin{aligned}\Delta &= (w_1 - \epsilon)\theta_1' + (w_2 + \epsilon)\theta_2' - (w_1\theta_1 + w_2\theta_2)\\ &= w_1(\theta_1' - \theta_1) + w_2(\theta_2' - \theta_2) + \epsilon(\theta_2' - \theta_1').\end{aligned}$$

Suppose that the stock price q_1 does not change. Since agents have CRRA preferences, we have $\theta_i' = \theta_i$, so $\Delta = \epsilon(\theta_2 - \theta_1) > 0$. Since agents $i > 2$ are unaffected unless the risk-free rate changes, there is a positive excess demand in the risky asset.

Regard θ_i as a function of the stock price q_1. By the maximum theorem, θ_i is continuous, and so is the aggregate demand. Since $\theta_i < 0$ if $R_f > \mathrm{E}_i[R]$ by Lemma A.1, and $R = e/q_1$, the aggregate excess demand of the risky asset becomes negative when $q_1 > \mathrm{E}_i[e]/R_f$. Therefore by the intermediate value theorem, there exists an equilibrium stock price higher than the original one. Since by Theorem 2.1 the equilibrium is unique, in the new equilibrium the stock price is higher, and hence the equity premium is lower. □

Lemma A.2. *Consider two agents indexed by $i = 1, 2$ with common beliefs. Let w_i, $u_i(x)$, $\gamma_i(x) = -xu_i''(x)/u_i'(x)$, and θ_i be the initial wealth, utility function, relative risk aversion, and the optimal portfolio of agent i. Suppose that $\gamma_1(w_1x) > \gamma_2(w_2x)$ for all x, so agent 1 is more risk averse than agent 2. Then*

$$\begin{aligned}\mathrm{E}[R] > R_f &\implies \theta_2 > \theta_1 > 0,\\ \mathrm{E}[R] < R_f &\implies \theta_2 < \theta_1 < 0,\end{aligned}$$

so the less risk averse agent invests more aggressively.

Proof. Since $\gamma_1(w_1x) > \gamma_2(w_2x)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u_2'(w_2x)}{u_1'(w_1x)}\right) = \frac{w_2u_2''u_1' - u_2'w_1u_1''}{(u_1')^2} = \frac{1}{x}\frac{u_2'}{u_1'}(\gamma_1(w_1x) - \gamma_2(w_2x)) > 0,$$

so $u_2'(w_2x)/u_1'(w_1x)$ is increasing. Suppose $\mathrm{E}[R] > R_f$. By Lemma A.1, we have $\theta_1 > 0$. Then $R(\theta_1) \gtrless R_f$ according as $R \gtrless R_f$. Since $u_2'(w_2x)/u_1'(w_1x)$ is increasing (and positive), we have

$$\frac{u_2'(R(\theta_1)w_2)}{u_1'(R(\theta_1)w_1)}(R - R_f) > \frac{u_2'(R_fw_2)}{u_1'(R_fw_1)}(R - R_f)$$

always (except when $R = R_f$). Multiplying both sides by $u_1'(R(\theta_1)w_1) > 0$ and taking expectations, we get

$$\begin{aligned}\mathrm{E}[u_2'(R(\theta_1)w_2)(R - R_f)] &= \mathrm{E}\left[\frac{u_2'(R(\theta_1)w_2)}{u_1'(R(\theta_1)w_1)}u_1'(R(\theta_1)w_1)(R - R_f)\right]\\ &> \mathrm{E}\left[\frac{u_2'(R_fw_2)}{u_1'(R_fw_1)}u_1'(R(\theta_1)w_1)(R - R_f)\right]\\ &= \frac{u_2'(R_fw_2)}{u_1'(R_fw_1)}\,\mathrm{E}\left[u_1'(R(\theta_1)w_1)(R - R_f)\right] = 0,\end{aligned}$$

where the last equality uses the first-order condition for agent 1. Letting $f_2(\theta) = \mathrm{E}[u_2(R(\theta)w_2)]$, the above inequality shows that $f_2'(\theta_1) > 0$. Since $f_2(\theta)$ is concave and $f_2'(\theta_2) = 0$ by the first-order condition, we have $\theta_2 > \theta_1$.

The case $\mathrm{E}[R] < R_f$ is analogous. □

Proof of Proposition 2.3. Since agents have common beliefs, we have $\theta_i \gtreqless 0$ for all i if $\mathrm{E}[R] \gtreqless R_f$. Since the stock is in positive supply, in equilibrium we must have $\mathrm{E}[R] > R_f$. Therefore by Lemma A.2, if $\gamma_1 > \cdots > \gamma_I$, we have $0 < \theta_1 < \cdots < \theta_I$. □

Proof of Proposition 2.4. Let $u(x)$ be the common CRRA utility function of agents 1 and 2, and

$$f_i(\theta) = \mathrm{E}_i[u(R(\theta))] = \sum_{s=1}^{S} \pi_{is} u(R_f + X_s \theta)$$

be the objective function of agent i, where $X_s = R_s - R_f$ denotes the excess return in state s. By the first-order condition, we have

$$f_i'(\theta_i) = \sum_{s=1}^{S} \pi_{is} u'(R_f + X_s \theta_i) X_s = 0. \tag{A.5}$$

Letting q be the stock price, since $R_s = e_s/q$ and $e_1 < \cdots < e_S$, we have $X_1 < \cdots < X_S$. Since $\pi_{is} > 0$ and $u' > 0$, by (A.5), it must be $X_1 < 0 < X_S$. Let $s^* = \max\{s \mid X_s < 0\}$ be the best state with negative excess returns. Clearly $1 \le s^* < S$.

Using the definition of the likelihood ratio $\lambda_s = \pi_{1s}/\pi_{2s}$, by (A.5) we obtain

$$0 = f_1'(\theta_1) = \sum_{s=1}^{S} \pi_{1s} u'(R_f + X_s \theta_1) X_s = \lambda_{s^*} \sum_{s=1}^{S} \frac{\lambda_s}{\lambda_{s^*}} \pi_{2s} u'(R_f + X_s \theta_1) X_s.$$

Since by assumption the likelihood ratio λ_s is monotonically decreasing, we have $\lambda_s/\lambda_{s^*} \ge (\le)\ 1$ for $s \le (\ge)\ s^*$. Furthermore, since beliefs are heterogeneous, either $\lambda_1/\lambda_{s^*} > 1$ or $\lambda_S/\lambda_{s^*} < 1$ (or both). Combined with $X_1 < 0 < X_S$ and $X_s < (\ge)\ 0$ for $s \le (\ge)\ s^*$, it follows that

$$\begin{aligned} 0 &= \lambda_{s^*} \sum_{s=1}^{S} \frac{\lambda_s}{\lambda_{s^*}} \pi_{2s} u'(R_f + X_s \theta_1) X_s \\ &< \lambda_{s^*} \sum_{s=1}^{S} \pi_{2s} u'(R_f + X_s \theta_1) X_s = \lambda_{s^*} f_2'(\theta_1), \end{aligned}$$

where the inequality is due to the fact that replacing $\lambda_s/\lambda_{s^*} \ge (\le)\ 1$ by 1 for $s \le (\ge)\ s^*$ makes the term less negative (more positive), and the inequality is strict for $s = 1$ or $s = S$. Therefore $f_2'(\theta_1) > 0$, and since f_2 is strictly concave and $f_2'(\theta_2) = 0$, we obtain $\theta_1 < \theta_2$. □

B Robustness of predictability

As described in Section 3.2, Tables 13 and 14 show that the inverse relationship between the top 1% share and subsequent excess returns is larger in magnitude and significant at the 1% and 5% levels, respectively, when we predict with `cgdiff`(10%) or `cgdiff`(0.1%). Tables 15–17 explore the robustness (with respect to detrending) of the result that when the top income share is above trend, subsequent one year excess returns are significantly below average. Table 18 shows the pairwise correlations between the explanatory variables used in Section 3.2.

As described in Section 3.2, Table 15 repeats the analysis of Table 1 but with the Kalman filter with an AR(1) cyclical component as discussed in Appendix C, which is one-sided (the cycle estimate in year t is based only on data up to year t). The results are roughly the same as in Table 1. Table 16 is similar to Table 15 but with the HP filter with a smoothing parameter of 100, which is standard for annual frequencies. The one difference is that column (2) uses the HP filter with a smoothing parameter of 10, whereas column (2) of Table 15 considers the AR(2) Kalman filter. With the exception of the 1945-2014 specifications including CAY (regressions (9) and (10)), the HP results are stronger and more significant, with top share coefficients ranging from around -4 to -6 and most p-values below 1%. When including CAY, the 1% coefficient is roughly the same in both the Kalman and HP specifications.[25]

Table 17 explores other detrending techniques. In column (1), we use the one-sided HP filter with a smoothing parameter of 100. The one-sided HP filter detrends each data point by applying the filter only to the previous data. In column (2), we estimate and remove two linear trends, a downward one pre-1977 and an upward one post-1977. Each case gives a slightly stronger result than in our baseline regression but a slightly weaker result than with the two-sided HP filter. Finally, in column (3) we estimate the trend using a ten year moving average. Compared with the AR(1) Kalman filter, this method, which is also one-sided, yields a slightly weaker but still significant relationship between inequality and subsequent excess returns.

As we saw in Section 3.2, controlling for the price-dividend (or price-earnings ratio) mitigates to a small degree the estimated effect of inequality on subsequent excess returns. But, in the post-1944 sample, when controlling for the price-dividend ratio, CAY, and the other macro factors, the 1% coefficient increases in magnitude (from -2.82 to -4.86) and becomes significant at the 1% level. However, because the rich hold more stock than do the poor, high prices and the resulting capital gains likely have some direct impact on the top income shares. To see this point, Table 18 shows the correlations between top income shares and classic return predictors. The only control variables significantly correlated with the top share measures are log(P/D) and log(P/E). This relationship is consistent both with the idea that rising income concentration pushes up stock prices and that the rich are disproportionately exposed to stocks.

[25]In contrast to the Kalman filter, the HP filter uses past, current, and future data to obtain a smooth trend, thereby potentially introducing a look-ahead bias. For example, since the rich are likely to be more exposed to the stock market, when the stock market goes up at year $t+1$, the rich will be richer than usual. But then the trend in the top income share will shift upwards, and the year t deviation of the top income share will be lower. Therefore the low income share at year t may spuriously predict a high stock return at $t+1$. This might explain the generally stronger results in Table 16.

Table 13: Regressions of one year excess stock market returns on `cgdiff` (top 10% − top 10% (without capital gains)) and other predictors

	Dependent Variable: t to $t+1$ Excess Market Return						
Regressors (t)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	17.89	16.66	18.29	19.08	24.09	18.75	23.17
	(3.47)	(4.53)	(4.01)	(14.36)	(9.78)	(4.27)	(14.23)
`cgdiff`(10%)	-5.46***	-5.36***	-5.67***	-5.34**	-5.05***	-5.46***	-5.36**
	(1.69)	(1.97)	(2.05)	(2.19)	(1.83)	(1.94)	(2.25)
Real GDP Growth		0.32					0.02
		(0.40)					(0.71)
ΔCons. Growth Variance			-12.59				-26.88*
			(15.07)				(16.13)
log(P/D)				-0.43			-1.58
				(4.84)			(4.59)
log(P/E)					-2.58		
					(3.86)		
CAY						1.53**	1.60**
						(0.73)	(0.73)
Sample	1917-	1930-	1935-	1917-	1917-	1945-	1945-
	-2014	-2014	-2014	-2014	-2014	-2014	-2014
R^2	0.08	0.08	0.10	0.08	0.09	0.17	0.20

Note: Newey-West standard errors in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). "`cgdiff`" is top 10% minus top 10% (no cg), neither detrended, where top 10% is the pre-tax share of income going to the top 10% of earners (including capital gains). Consumption growth volatility is from an $AR(1) - GARCH(1,1)$ model. P/D and P/E are the S&P500 price-dividend and price-earnings ratios. CAY is the consumption/wealth ratio.

Table 14: Regressions of one year excess stock market returns on `cgdiff` (top 0.1%−top 0.1% (without capital gains)) and other predictors

	Dependent Variable: t to $t+1$ Excess Market Return						
Regressors (t)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	13.44	12.93	13.20	17.07	16.13	13.25	27.16
	(2.69)	(3.82)	(3.25)	(16.41)	(11.29)	(3.61)	(15.31)
`cgdiff`(0.1%)	-4.23**	-4.19**	-3.94*	-3.80	-4.01**	-3.60	-2.11
	(1.66)	(2.07)	(2.07)	(2.55)	(1.86)	(2.23)	(2.41)
Real GDP Growth		0.24					-0.20
		(0.41)					(0.81)
ΔCons. Growth Variance			-8.52				-22.07
			(15.24)				(16.82)
log(P/D)				-1.27			-4.69
				(5.58)			(4.60)
log(P/E)					-1.11		
					(4.39)		
CAY						1.46**	1.42*
						(0.73)	(0.79)
Sample	1913-	1930-	1935-	1913-	1913-	1945-	1945-
	-2014	-2014	-2014	-2014	-2014	-2014	-2014
R^2	0.05	0.05	0.05	0.05	0.05	0.10	0.13

Note: Newey-West standard errors in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). "`cgdiff`" is top 0.1% minus top 0.1% (no cg), neither detrended, where top 0.01% is the pre-tax share of income going to the top 0.1% of earners (including capital gains). Consumption growth volatility is from an $AR(1) - GARCH(1,1)$ model. P/D and P/E are the S&P500 price-dividend and price-earnings ratios. CAY is the consumption/wealth ratio.

Table 15: Regressions of one year excess stock market returns on top income shares and other predictors (using Kalman filter)

	Dependent Variable: t to $t+1$ Excess Market Return									
Regressors (t)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	8.49	8.40	8.44	9.10	7.62	8.75	22.65	19.10	9.53	25.18
	(1.76)	(1.76)	(1.74)	(1.73)	(2.36)	(1.53)	(13.35)	(10.86)	(1.88)	(15.44)
Top 1%	-2.82**				-4.11**	-4.69***	-2.22*	-2.60**	-4.23**	-4.86***
	(1.21)				(1.63)	(1.62)	(1.18)	(2.12)	(1.78)	(1.76)
Top 1% ($p=2$)		-2.64**								
		(1.15)								
Top 1% (no cg)			-3.78							
			(2.36)							
Top 0.1%				-4.66***						
				(1.58)						
Real GDP Growth					0.37					0.35
					(0.37)					(0.79)
ΔCons. Growth Variance						-17.11				-30.05*
						(16.43)				(17.18)
log(P/D)							-4.34			-4.98
							(3.94)			(4.23)
log(P/E)								-3.98		
								(3.88)		
CAY									1.85**	1.92***
									(0.75)	(0.71)
Sample	1913- -2014	1913- -2014	1913- -2014	1913- -2014	1930- -2014	1935- -2014	1913- -2014	1913- -2014	1945- -2014	1945- -2014
R^2	0.04	0.04	0.02	0.05	0.06	0.09	0.05	0.04	0.13	0.19

Note: Newey-West standard errors in parentheses ($k=4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). Unless otherwise stated, Top 1% is the pre-tax share of income going to the top 1% of earners (including capital gains). "no cg" refers to the series that excludes capital gains. The top shares series are detrended with the Kalman filter ($p=1$) unless otherwise noted. Consumption growth volatility is from an $AR(1)-GARCH(1,1)$ model. P/D and P/E are the S&P500 price-dividend and price-earnings ratios. CAY is the consumption/wealth ratio.

Table 16: Regressions of one year excess stock market returns on top income shares and other predictors (using HP filter)

	Dependent Variable: t to $t+1$ Excess Market Return									
Regressors (t)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	8.10	8.10	8.12	8.11	5.59	7.57	24.54	22.97	7.74	28.97
	(1.73)	(1.72)	(1.73)	(1.70)	(2.34)	(1.59)	(13.26)	(11.30)	(1.67)	(14.50)
Top 1%	-4.21***				-5.22***	-5.54***	-3.93***	-4.24***	-4.03***	-4.21***
	(1.24)				(1.70)	(1.50)	(1.18)	(1.26)	(1.46)	(1.46)
Top 1% (HP param.= 10)		-5.89***								
		(1.75)								
Top 1% (no cg)			-4.40**							
			(2.18)							
Top 0.1%				-6.00***						
				(1.53)						
Real GDP Growth					0.60*					0.02
					(0.36)					(0.72)
ΔCons. Growth Variance						-16.28				-27.81*
						(15.90)				(15.76)
log(P/D)							-5.02			-6.38
							(3.89)			(4.12)
log(P/E)								-5.57		
								(4.06)		
CAY									1.59***	1.50**
									(0.58)	(0.70)
Sample	1913- -2014	1913- -2014	1913- -2014	1913- -2014	1930- -2014	1935- -2014	1913- -2014	1913- -2014	1945- -2014	1945- -2014
R^2	0.09	0.10	0.03	0.09	0.10	0.13	0.10	0.10	0.14	0.20

Note: Newey-West standard errors in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). Unless otherwise stated, Top 1% is the pre-tax share of income going to the top 1% of earners (including capital gains). "no cg" refers to the series that excludes capital gains. The top shares series are detrended with the HP filter (smoothing parameter of 100 unless otherwise stated). Consumption growth volatility is from an $AR(1) - GARCH(1,1)$ model. P/D and P/E are the S&P500 price-dividend and price-earnings ratios. CAY is the consumption/wealth ratio.

Table 17: Regressions of one year excess stock market returns on top income shares (using different trend estimates)

Dependent Variable: t to $t+1$ Excess Market Return			
Regressors (t)	(1)	(2)	(3)
Constant	7.97 (1.89)	8.16 (1.68)	8.78 (1.87)
Top 1% (one-sided HP)	-3.63* (1.87)		
Top 1% (linear detrending)		-3.11*** (0.98)	
Top 1% (10 year MA trend)			-1.97** (0.96)
Sample	1936- -2014	1913- -2014	1922- -2014
R^2	0.04	0.09	0.04

Note: Newey-West standard errors in parentheses ($k = 4$). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). Top 1% is the pre-tax share of income going to the top 1% of earners (including capital income). The one-sided HP filter uses a smoothing parameter of 100. The MA trend is a 10 year moving average. To linearly detrend, we impose a downward time trend before 1977 and an upward one after.

Table 18: Pairwise correlations between explanatory variables

	Top 1% Kal.	`cgdiff` (10%)	Top 1% (no cg)	Top 0.1% Kal.	`cgdiff` (1%)	`cgdiff` (0.1%)	%ΔRGDP	ΔCGV	log(P/D)	log(P/E)	CAY
Top 1% (Kal.)	1.00										
`cgdiff`(10%)	0.73*	1.00									
Top 1% (no cg)	0.87*	0.44*	1.00								
Top 0.1% (Kal.)	0.95*	0.75*	0.83*	1.00							
`cgdiff`(1%)	0.73*	0.97*	0.41*	0.70*	1.00						
`cgdiff`(0.1%)	0.68*	0.93*	0.39*	0.67*	0.99*	1.00					
%ΔRGDP	0.12	0.06	0.03	0.13	0.01	-0.04	1.00				
ΔCGV	-0.22	-0.12	-0.19	-0.20	-0.08	-0.05	-0.01	1.00			
log(P/D)	0.37*	0.62*	0.26*	0.33*	0.68*	0.70*	-0.04	0.00	1.00		
log(P/E)	0.16	0.39*	0.08	0.11	0.44*	0.46*	-0.21	-0.04	0.74*	1.00	
CAY	0.18	-0.05	0.17	0.17	-0.10	-0.13	-0.16	0.11	-0.21	-0.09	1.00

Note: This table shows annual frequency time series correlations for the explanatory variables used in Section 3.2. * indicates significance at the 5% level. Top 1% (Kal.) is the pre-tax share of income going to the top 1% of earners (including capital gains), detrended with the Kalman filter ($p = 1$). Top 0.1% (Kal.) is defined analogously. "no cg" refers to the Kalman filtered series that excludes capital gains. "`cgdiff`" is top 1% minus top 1% (no cg), neither detrended. `cgdiff`(0.1%) and `cgdiff`(10%) are calculated analogously. Consumption growth volatility (CGV) is from an $AR(1) - GARCH(1,1)$ model. %ΔRGDP is the percentage change in real GDP. P/D and P/E are the S&P500 price-dividend and price-earnings ratios. CAY is the consumption/wealth ratio. The samples are 1913-2014 for the top share series and price ratios, 1930-2014 for GDP, 1935-2014 for consumption volatility, 1945-2014 for CAY.

C Kalman filter

This appendix explains how we detrend the top income/wealth share using the Kalman filter.

Let y_t be the observed top income/wealth share data at time t. Let

$$y_t = g_t + u_t, \tag{C.1}$$

where g_t is the trend and u_t is the cyclical component. We conjecture that the trend is an I(2) process, and the cycle is an AR(p) process, so

$$(1-L)^2 g_t = \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ N(0, \sigma_\epsilon^2), \tag{C.2a}$$

$$\phi(L) u_t = w_t, \qquad w_t \sim i.i.d.\ N(0, \sigma_w^2), \tag{C.2b}$$

where L is the lag operator and

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

is the lag polynomial for the autoregressive process. For concreteness, assume $p = 1$ so $\phi(z) = 1 - \phi_1 z$. Then (C.1) and (C.2) can be written as

$$\begin{bmatrix} g_t \\ g_{t-1} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} g_{t-1} \\ g_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}, \tag{C.3a}$$

$$y_t = \phi_1 y_{t-1} + g_t - \phi_1 g_{t-1} + w_t. \tag{C.3b}$$

Letting $\xi_t = (g_t, g_{t-1})'$, $v_t = (\epsilon_t, 0)'$, $x_t = y_{t-1}$, $A = \phi_1$, $F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$, and $H = \begin{bmatrix} 1 & -\phi_1 \end{bmatrix}$, (C.3) reduces to

$$\xi_t = F \xi_{t-1} + v_t, \tag{C.4a}$$

$$y_t = A x_t + H \xi_t + w_t. \tag{C.4b}$$

(C.4a) is the state equation and (C.4b) is the observation equation of the state space model. We can then estimate the model parameters $\phi_1, \sigma_\epsilon^2, \sigma_w^2$ as well as the trend $\{g_t\}$ by maximum likelihood: see Chapter 13 of Hamilton (1994) for details. The extension to a general AR(p) model is straightforward.

D International data

Unless otherwise noted, the top income share series is the "Top 1% income share" excluding capital gains from Alvaredo et al. (2015) (see also their documentation), the price index is the Haver/IMF CPI, and the interest rate is the Haver "Deposit Rate" series.

1. **Argentina (ARG)**

   Coverage 1998-2005.

   Local Currency Deposit Rate 30-59 day deposit rate.

2. **Australia (AUS)**

   Coverage 1970-2011.

   Local Currency Deposit Rate 1972-2011.

3. **Canada (CAN)**

   Coverage 1970-2011.

   1% Income Share LAD series post-1995.

   Local Currency Deposit Rate 90 day deposit rate. 1971-2011.

4. **China (CHN)**

   Coverage 1993-2004.

   Local Currency Deposit Rate 1 year deposit rate.

5. **Colombia (COL)**

   Coverage 1994-2011.

6. **Denmark (DNM)**

   Coverage 1971-1973, 1975-2011.

   1% Income Share "Adults" series.

   Local Currency Deposit Rate 1980-2002.

7. **Finland (FIN)**

   Coverage 1988-2010.

   1% Income Share "Tax data" series pre-1993 and "IDS" 1993–. We average the two for 1990-1992.

   Local Currency Deposit Rate 23 month deposit rate, 1988-2005.

8. **France (FRA)**

   Coverage 1970-2010.

9. **Germany (GER)**

   Coverage 1972, 1975, 1978, 1981, 1984, 1986, 1990, 1993, 1996, 1999, 2002-2009.

   Local Currency Deposit Rate 3 month deposit rate, 1978-2003.

   Price Index GDP deflator pre-1991.

10. **India (INI)**

    Coverage 1993-2000.

    Local Currency Deposit Rate Bank discount rate from Haver.

11. **Indonesia (INO)**

    Coverage 1988, 1991, 1994, 1997, 1999-2005.

    Local Currency Deposit Rate 3 months deposit rate.

12. **Ireland (IRE)**

    Coverage 1988-2010.

    Local Currency Deposit Rate 1988-2006.

    Price Index http://www.cso.ie

13. **Italy (ITA)**

    Coverage 1975-1996, 1999-2010.

    Local Currency Deposit Rate 1983-2004.

14. **Japan (JPN)**

    Coverage 1970-2011.

    Local Currency Deposit Rate 3 month deposit rate.

15. **South Korea (KOR)**

    Coverage 1996-2013.

Local Currency Deposit Rate 1 year deposit rate.

16. **Malaysia (MAL)**

    Coverage 1989, 1994-1996, 2001-2004, 2006, 2010-2013.

    Local Currency Deposit Rate 3 month deposit rate.

    Price Index blabla

17. **Mauritius (MAU)**

    Coverage 2003-2009, 2011-2012.

    Local Currency Deposit Rate 3 month deposit rate.

18. **Netherlands (NET)**

    Coverage 1971, 1974, 1976, 1978, 1982, 1986, 1990-2013.

19. **New Zealand (NZL)**

    Coverage 1988-2012.

    1% Income Share "Adults" series.

    Local Currency Deposit Rate 6 month deposit rate, 1990-2012.

20. **Norway (NOR)**

    Coverage 1970-2012.

    Local Currency Deposit Rate 1979-2010.

21. **Portugal (POR)**

    Coverage 1990-2006.

    Local Currency Deposit Rate 180-360 day deposit rate, 1990-2000.

22. **South Africa (SAF)**

    Coverage 1993-1994, 2003-2012.

    1% Income Share Pre-1990, "Married Couples and Single Adults" series. Post-1990, "Adults" series.

    Local Currency Deposit Rate 88-91 day deposit.

23. **Singapore (SIN)**

    Coverage 1970-1992, 1994-2013.

    Local Currency Deposit Rate 3 month deposit rate, 1977-2013.

    Price Index blabla

24. **Spain (SPA)**

    Coverage 1982-2013.

    Local Currency Deposit Rate 6-12 month deposit rate, 1982-2013.

25. **Sweden (SWE)**

    Coverage 1970-2013.

    Local Currency Deposit Rate 1970-2006.

26. **Switzerland (SWI)**

    Coverage 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996-2011.

    Local Currency Deposit Rate 3 month deposit rate,

    Price Index 1982-2011.

27. **Taiwan (TAI)**

    Coverage 1988-2014.

Local Currency Deposit Rate Missing.

Price Index CPI, Datastream.

28. **United Kingdom (UNK)**

Coverage 1970-2013.

1% Income Share Pre-1990, "Married Couples and Single Adults" series. Post-1990, "Adults" series.

Local Currency Deposit Rate 90 day T-bill rate.

Price Index GDP deflator pre-1988.

29. **United States (USA)**

Coverage 1970-2013.

Local Currency Deposit Rate 3 month T-bill rate.

E Additional international results

Table 19: Country fixed effects panel regressions of one year stock returns on top income shares

	Dependent Variable: t to $t+1$ Stock Market Return			
Regressors	All	Advanced	IIPS	EME
Top 1% (t)	-1.45**	-1.05*	-5.42**	-5.53*
	(0.67)	(0.56)	(1.47)	(2.45)
Time Trend	No	No	No	No
Country FE	Yes	Yes	Yes	Yes
Obs.	790	699	106	91
R^2 (w,b)	(.01,.11)	(.00,.06)	(.05,.06)	(.04,.18)

Note: Clustered standard errors in parentheses, ***1%, **5%, *10%. R^2 (w,b): Within and between R-squared. Constants suppressed. Top 1% is the pre-tax share of income going to the top 1% of earners (excluding capital gains). The column headings refer to the countries included (see the main text for details). Sample: 1969-2013 (see Appendix D for country details).