The Equity Premium and the One Percent

Alexis Akira Toda†    Kieran Walsh‡

This version: March 18, 2016

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 the top 1% income earners in the U.S. rises above trend by one percentage point, subsequent one year market excess returns decline on average by about 3–5%. This negative relation is robust to (i) controlling for classic return predictors such as the price-dividend and consumption-wealth ratios and (ii) predicting out-of-sample. 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 (p. 175) the “enterpriser-borrower” with the “creditor, the salaried man, or the laborer,” emphasizing that the former class of society accelerates fluctuations

*We benefited from comments by Dan Cao, Vasco Carvalho, Peter Debaere, Graham Elliott, John Geanakoplos, Emilien Gouinbœufant, Jim Hamilton, Gordon Hanson, Fumio Hayashi, Toshiki Honda, Jiasun Li, 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, and 2015 UVa-Richmond Fed Jamboree. 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
in prices and production. Central to his theories of fluctuations were differences in preferences and wealth across people.

Following the seminal work of Lucas (1978), however, the “representative agent” consumption-based asset pricing models—which seem to allow no role for agent heterogeneity—have dominated the literature, at least until recently. Yet agent heterogeneity may (and is likely to) matter even if a representative agent exists: unless agents have very specific preferences that admit the Gorman (1953) aggregation (a knife-edge case, which is unlikely to hold in reality), the preferences of the representative agent will in general depend on the wealth distribution, as pointed out by Gollier (2001). Indeed, even with complete markets, the preferences of the representative agent are typically nonstandard when individual utilities do not reside within quite particular classes.

To see the intuition as to why the wealth distribution affects asset pricing, consider an economy consisting of people 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 static model with incomplete markets, heterogeneous CRRA utilities, and collinear (but different) endowments, we prove that there exists a unique equilibrium and that in this equilibrium increasing wealth concentration in the hands of stock holders (the more risk tolerant or optimistic agents) leads to a decline in the equity premium. We then explore the robustness of this link between prices and inequality by turning to an infinite horizon model with two assets, stocks and bonds, and a stark form of heterogeneity in risk aversion: one type of agent does not participate in risky stock markets. Exploiting log utility, we solve the model in closed-form (up to a single nonlinear equation with one unknown). Moreover, we prove that wealth redistribution from the bond holders to the stock and bond investors leads to a decline in the subsequent equity premium.

Although the inverse relationship between wealth concentration and risk premiums in the presence of heterogeneous risk aversion has been recognized at least since Dumas (1989) and recently emphasized by Garleanu and Panageas (2015), each of our models extends the existing results: our static model allows for incomplete markets, many agents with heterogeneous CRRA utilities, and arbitrary shocks; our infinite horizon model is in discrete time and allows for an arbitrary dividend (and thus consumption) growth process.

Second, we empirically explore the theoretical predictions. We find that when the income share of the top 1% income earners in the U.S. is above trend, the subsequent one year U.S. stock market equity premium is below average. That is, current inequality appears to forecast the subsequent risk premium of the U.S. stock market. Many heterogeneous agent general equilibrium models in both macroeconomics and finance predict a relationship between the concentration of income and asset prices (see Sections 1.1 and 2.3). We thus provide empirical support for a literature which has been subject to relatively little direct testing. Furthermore, the patterns we uncover are intuitive. In short, if
one believes top earners invest relatively more in risky assets (due to high risk tolerance or optimism), then it should not be surprising that in the data asset returns suffer as the rich get richer. Theorem 2.2 rigorously confirms this intuition.

More specifically, we employ regression analysis to establish the correlation between inequality and returns. 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 the top 1% income share rises above trend by one percentage point, subsequent one year market excess returns decline on average by about 3–5%, depending on the detrending method and 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 forecasters such as the price-dividend ratio (Shiller, 1981) and the consumption-wealth ratio (Letttau and Ludvigson, 2001). Our findings are also robust to the inclusion of macro control variables, such as GDP growth, and to a variety of detrending methods. 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% income share yields similar results. “cgdiff,” which we define to be the difference between the top 1% income share with and without capital gains income, also predicts poor subsequent returns, even when controlling for the price-dividend ratio.

Welch and Goyal (2008) show that excess return predictors suggested in the literature by and large perform poorly out-of-sample, possibly due to model instability, data snooping, or publication bias. 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, price-dividend ratio, or price-earnings ratio. That is, top income shares predict returns out-of-sample as well.

Given that in our regressions we lag the 1% share 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 marginal tax rates have an inverse relationship with top income shares (see, for example, Roine et al. (2009)), as an additional robustness check, we explore using tax changes as an instrument for inequality in predicting returns. We identify 7 periods in U.S. history (over 1915-2004) where top marginal income and estate tax rates were either trending upwards or downwards. Corroborating our regression analysis, 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. As industrial production growth is actually higher on average in our hike periods, we argue that these results are not driven by the expansionary/contractionary effects of fiscal policy.

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 models suggest 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). 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.

The seminal consumption-based asset pricing work of Lucas (1978), Breeden (1979), and Hansen and Singleton (1983) also abstracted from investor heterogeneity. They and others derived and tested analytic relationships between the marginal rate of substitution of a representative agent (with standard preferences) and asset prices. Despite the elegance and tractability of the representative agent/aggregation approach, it has failed to adequately explain the fluctuations of asset prices in the economy. Largely inspired by the limited empirical fit of the CAPM (in explaining the cross section of stock returns), the equity premium puzzle (Mehra and Prescott, 1985), and excess stock market volatility and related price-dividend ratio anomalies (Shiller, 1981), since the 1980s theorists have extended macro/finance general equilibrium models to consider meaningful investor heterogeneity. Such heterogeneous-agent models can be categorized into two groups.

In the first group, agents have identical standard (constant relative risk aversion) preferences but are subject to uninsured idiosyncratic risks. Although the models of this literature have had some success explaining returns in calibrated simulations, 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).

In the second group, markets are complete and agents have either heterogeneous

---

1See Geanakoplos and Shubik (1990) for a general and rigorous treatment of CAPM theory.
neous CRRA preferences or identical but non-homothetic preferences. 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 irrelevant 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 Găreleanu 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 Găreleanu 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) is the only one that explores this issue using income/wealth distribution data. However, his analysis is quite different from ours: his model relies on a “keeping up with the Joneses”-type consumption externality with incomplete markets. In contrast, we employ a standard general equilibrium model (a plain vanilla Arrow-Debreu model). Moreover, Johnson (2012) does not explore the ability of top income shares to predict market excess returns (our main result), and he detrends inequality differently from the way we do.

Lastly, our study is related to the empirical literature on heterogeneity in risk preferences. A number of recent papers have found that the wealthy have portfolios more heavily skewed towards risky assets, and many of these studies have concluded that the wealthy are relatively more risk tolerant, either due to declining relative risk aversion or innate heterogeneity in relative risk aversion. See, for example, Carroll (2002), Vissing-Jørgensen (2002), Campbell (2006), Bucciol and Miniaci (2011), or Calvet and Sodini (2014). This literature lends credibility to our premise that the rich invest more heavily in risky assets, likely due to higher risk tolerance.


\(^4\)Examples are Goller (2001) and Hatchondo (2008).

\(^5\)As we do, Basak and Cocco (1998) and Guvenen (2009) study cases with incomplete markets and agent heterogeneity. While both papers consider limited stock market participation, the former is in continuous time and the latter allows for heterogeneous Epstein-Zin preferences.
2 Wealth distribution and equity premium

In this section we present two models in which the wealth distribution across heterogeneous agents affects the equity premium. In Section 2.1, we consider a static model with incomplete markets and agents with heterogeneous risk aversion and beliefs. In Section 2.2, we consider an infinite horizon model with log utility and limited stock market participation. In either case, 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 Heterogeneous risk aversion and beliefs

Consider a standard general equilibrium model with incomplete markets consisting of a single good, \( I \) agents, \( J \) assets, and \( S \) states (Geanakoplos, 1990). Let \( e_i \in \mathbb{R}^S_{++} \) be the initial endowment of agent \( i \) 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 \( q = (q_1, \ldots, q_J)' \in \mathbb{R}^J \), agent \( i \)'s utility maximization problem is

\[
\begin{align*}
\text{maximize} & \quad U_i(x) \\
\text{subject to} & \quad q' y \leq 0, \quad x \leq e_i + Ay,
\end{align*}
\]

where \( U_i(x) \) is the utility function and \( y = (y_1, \ldots, y_J)' \in \mathbb{R}^J \) denotes the number of asset shares. A general equilibrium with incomplete markets (GEI) consists of an asset price \( q \in \mathbb{R}^J \), 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:

\[
U_i(x) = \begin{cases} 
\left( \sum_{s=1}^S \pi_s x_s^{1-\gamma_i} \right)^{1 \over \gamma_i}, & (\gamma_i \neq 1) \\
\exp \left( \sum_{s=1}^S \pi_s \log x_s \right), & (\gamma_i = 1)
\end{cases} \tag{2.1}
\]

where \( \gamma_i > 0 \) is agent \( i \)'s relative risk aversion coefficient and \( \pi_is \) is agent \( i \)'s subjective probability of state \( s \).

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

\[
\frac{1}{1-\gamma_i} U_i(x)^{1-\gamma_i} = \frac{1}{1-\gamma_i} \sum_{s=1}^S \pi_s x_s^{1-\gamma_i},
\]

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 given the negative result of Mantel (1976).

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 2–3 there exists a unique GEI. The equilibrium portfolio \((y_i)\) is the solution to the planner’s problem

\[
\begin{align*}
\text{maximize} & \quad \sum_{i=1}^{I} w_i \log U_i(e_i + Ay_i) \\
\text{subject to} & \quad \sum_{i=1}^{I} y_i = 0. \\
\end{align*}
\]

(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 proportional 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.

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 \( 1 = (1, \ldots, 1)' \in \mathbb{R}_+^S \).

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 assume \( q_1 = 1 \). Then the vector of gross stock returns is \( R := e/q_1 = e \), the aggregate endowment. The gross risk-free rate is \( R_f = 1/q_2 \). Since by Assumption 2 we have \( e_i = w_i e \), the initial wealth of agent \( i \) is \( w_i \) and the budget constraint is

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

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

7
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 \( 1 - \theta = \frac{y_2}{R_f w_i} \). Therefore the consumption vector satisfies

\[
x \leq c_i + Ay = (y_1 + w_i)e + y_21 = w_i(R\theta + R_f(1 - \theta)1).
\]

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

\[
\max_{\theta} E_i[u_i(R(\theta))],
\]

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 \( \{q_i(x_i), (y_i)\} \) be the unique GEI with corresponding portfolio \( (\theta_i) \). 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 lower equity premium.

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. The uniqueness result in Theorem 2.1 allows us to make this intuition rigorous.

The following propositions show that when agents have heterogeneous risk aversion or beliefs, the fraction of investment in the risky asset is ordered as risk tolerance or optimism.

**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 there are only two states indexed by \( s = u, d \), with \( e_u > e_d \). If \( \pi_{1d} > \pi_{2d} \), so agent 1 is more pessimistic, then \( \theta_1 < \theta_2 \).

Combining Theorem 2.2 together with 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, rising inequality should forecast declining excess returns.

**2.2 Limited market participation in infinite horizon**

In this section we exploit limited asset market participation and log utility to analytically derive the relationship between wealth concentration and excess
returns in an infinite horizon model. The model is a discrete-time counterpart of Basak and Cuoco [1998] with an arbitrary dividend process.

Consider an infinite horizon economy consisting of two agent types, 1 and 2. Each type has log utility with discount factor $0 < \beta < 1$. There are two assets, a stock in unit supply and a one period risk-free bond in zero net supply. We consider an extreme form of heterogeneity in risk aversion: while agent 1 may hold either asset, agent 2 may not buy or sell stock. That is, while the bondholder has curvature in his utility function and is willing to substitute consumption across time periods, he only invests in safe assets with nonrandom payoffs. We can interpret agent 2 as infinitely risk averse.

The dividend process for the stock is exogenous and is denoted by \( \{D_t\}_{t=0}^\infty \).

Let \( w_{it} \) be the wealth of agent \( i \) at time \( t \), \( W_t = w_{1t} + w_{2t} \) be aggregate wealth, and \( x_t = w_{1t}/W_t \) be the wealth share of agent 1. A sequential equilibrium consists of allocations (consumption and portfolio) and prices (stock price \( P_t \) and risk-free rate \( R_{f,t} \)) such that agents behave optimally taking prices as given and markets clear.

The following theorem characterizes the equilibrium in closed-form, up to a single state-by-state equation (part 2 of the theorem) for the interest rate.

**Theorem 2.5.** Suppose that $0 < \beta < 1$ and \( \sup_t \mathbb{E}[\log(D_{t+1}/D_t)] < \infty \). Then there exists a unique equilibrium. The equilibrium has the following properties.

1. The stock price is given by
   \[
   P_t = \beta \frac{1}{1-\beta} D_t
   \]
   and the stock return is
   \[
   R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{D_{t+1}}{1-\beta}. \]

2. The risk-free rate \( R_{f,t} \) satisfies
   \[
   \mathbb{E}_t \left[ \frac{(R_{t+1} - R_{f,t})x_t}{R_{t+1} + R_{f,t}(x_t - 1)} \right] = 0.
   \]
   The equity premium \( \mathbb{E}_t[R_{t+1}] - R_{f,t} \) is a decreasing function of agent 1’s wealth share \( x_t \).

3. Each agent’s wealth evolves according to
   \[
   w_{1,t+1} = \beta(R_{t+1}/x_t + R_{f,t}(1 - 1/x_t))w_{1t},
   \]
   \[
   w_{2,t+1} = \beta R_{f,t}w_{2t}.
   \]
   Agent 1’s wealth share evolves according to
   \[
   x_{t+1} = \frac{1 - R_{f,t}}{R_{f,t}/(x_t - 1)} (1 - x_t).
   \]

Theorem 2.5 confirms our intuition that concentration of wealth in the hands of the risk tolerent drives down the equity premium. In particular, Theorem 2.5 implies that a one-time, unanticipated shock redistributing from risky investors to bond holders would increase subsequent excess returns on average. One caveat for this example is that while the interest rate depends on the wealth distribution, the expected stock return does not. That is, the wealth distribution affects the equity premium only through the interest rate. Note, however, that this property is driven by the assumption of log utility, which ensures that the savings rate out of wealth is independent of the state and implies \( c = (1 - \beta)w \).

With general homothetic preferences, the savings rate depends on the state, which includes the wealth distribution.
Basak and Cuoco (1998) solve a limited participation model similar to ours in continuous-time. They show that with log utility, increasing the wealth share of the stock holder reduces the equity premium. Our result is exactly the same, but while they obtain this result with aggregate consumption growth following a geometric Brownian motion, in our discrete-time model the stochastic process for consumption is arbitrary. In summary, Theorem 2.5 shows that the Basak and Cuoco (1998) result also holds in a discrete time model with an essentially arbitrary underlying shock process.

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 falls and the equity premium rises. 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. 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 the 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 and incomplete markets.

Gârleanu and Panageas (2015) study a continuous-time overlapping generations endowment economy with two agent types with Epstein-Zin preferences. Unlike other papers on asset pricing models with heterogeneous preferences, all agent types survive in the long run due to birth/death, and also they 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.

3 Empirical link between inequality and equity premium

Thus far, we have theoretically analyzed models in which the extent of inequality across agents with different portfolios (due to, say, heterogeneous risk aversion or belief) is key in predicting returns. In particular, we found 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. Since income adds up to wealth, a positive income shock to the rich should negatively predict subsequent excess stock market returns.

In this section, consistent with our theory we show that there is a strong and robust inverse relationship between the top income share and subsequent excess stock market returns in the U.S. That is, current inequality appears to forecast the risk premium of the U.S. stock market (in Section 4 we uncover a similar pattern in the international data). In Section 3.3 we show that top income shares also predict returns out-of-sample.

3.1 Data

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 consider the top 0.1% share and the top 1% series that excludes capital gains income.

Figure 1a shows the top income shares with various cutoff points, the richest 0.1, 0.5, 1, 5, and 10%. All series look similar except for the level (the smallest pairwise correlation coefficient is 0.932, between top 0.1% and 10%). However, things are quite different when we break down the income shares by percentile. Figure 1b shows the income share of each group, the richest 0.1%, 0.1–0.5%, 0.5–1%, etc.

---


In view of the theoretical results in Section 2, top wealth share data are also relevant. However, the 1916–2000 top wealth 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 (2014) cover 1913-2012 but are estimates created by capitalizing income. Due to these limitations, we perform our analysis with top income share data.
0.5–1%, 1–5%, and 5–10%. Although all series seem to share a common U-shaped trend over the century, the behavior of these series around the trend is quite different. First, the top 0.5–1% share is very smooth. Second, the top 0–0.1% and 0.1–0.5% shares resemble each other and seem procyclical (move in the same direction as booms and busts), which is most apparent in the 1920s and post-1990. On the other hand, the top 1–5% and 5–10% resemble each other and seem countercyclical (move in the opposite direction as booms and busts).

To see this formally, Table 1 shows the pairwise correlation coefficients of the cyclical components of these series. The numbers range from 0.87 between the top 0.1, 0.1–0.5, and 0.5–1%, to -0.55 between the top 0.1 and 5–10%. Figure 2 shows the cyclical components of the top 1% and 1–10% shares. We can see that the two series are negatively correlated (correlation -0.32), and the top 1% series tend to rise (fall) during booms (busts). Within the context of the model in Section 2, this behavior can be explained if the richer agents are more risk tolerant or optimistic: since the top 1% is procyclical while the top 1–10% is countercyclical, the former (latter) is the natural stock (bond) holder. Thus, in bringing our theory to the data, we take the top 1%, rather than say the top 5% or 10%, as our dividing line between the types.

To bring our theory to the data, we use not the raw Piketty-Saez series but rather detrended, stationary versions. This is because while the inequality measures may change
for a variety of reasons, a slow-moving, predictable change should already be priced-in: what should matter for annual returns is the unanticipated, transitory component of inequality. Furthermore, imposing stationarity helps ensure the validity of standard error calculations and inference and prevent spurious regressions (Granger and Newbold, 1974).

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 (Roine et al., 2009). According to Figure 3, 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.

Specifically, for our baseline results we detrend the inequality measures in

---

10Stochastically detrending asset return predictors is in the tradition of Campbell (1991), for example, who removes a trend in the short-term interest rate before including it in stock return vector autoregressions.
two ways. First, we use the Kalman filter with an AR(1) cyclical component (see Appendix C for details). Second, we use “cgdiff,” which we define to be the top 1% income share including capital gains minus the top share without capital gains. cgdiff is stationary and is a measure of the component of the top 1% income share due to variation in capital gains. Since labor income is often set by contracts, it is more slow-moving and predictable than is capital income. Hence using cgdiff may help us isolate the shocks to inequality relevant for predicting excess returns. Figure 4 plots the top 1% series detrended by the AR(1) Kalman filter as well as cgdiff. We can see that the two series behave similarly (correlation 0.73).

![Figure 4: Time series plot of the cyclical component of the top 1% income share using the AR(1) Kalman filter and cgdiff (both demeaned).](image)

We calculate real annual one year U.S. stock market excess returns using the annual data from the website of Amit Goyal (the spreadsheet for Welch and Goyal (2008)). 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/E and P/D, which are the price-dividend and price-earnings ratios (in real terms) for the S&P 500. The spreadsheet also contains the Lettau-Ludvigson consumption-wealth ratio, commonly referred to as CAY, which spans the period 1945–2014. For presentation, we multiply CAY by 100.

Our other controls are GDP growth and, inspired by Lettau et al. (2008), we stochastically detrend with the Kalman filter instead of the Hodrick-Prescott (HP) filter, which is widely used, because the former is “one-sided” in the sense that the Kalman cycle estimate in year $t$ is based only on data up to year $t$. The HP filter, in contrast, 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 $t + 1$. Appendix B shows that our results are robust to a variety of detrending methods, including the HP filter with a smoothing parameter of 100 (which strengthens our results), the HP filter with smoothing parameter of 10, the one-sided HP filter, the moving average filter, and linear detrending.

11 We stochastically detrend with the Kalman filter instead of the Hodrick-Prescott (HP) filter, which is widely used, because the former is “one-sided” in the sense that the Kalman cycle estimate in year $t$ is based only on data up to year $t$. The HP filter, in contrast, 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 $t + 1$. Appendix B shows that our results are robust to a variety of detrending methods, including the HP filter with a smoothing parameter of 100 (which strengthens our results), the HP filter with smoothing parameter of 10, the one-sided HP filter, the moving average filter, and linear detrending.

12 The Phillips-Perron test (Phillips and Perron 1988) rejects a unit root in cgdiff at the 1% level.

13 http://www.hec.unil.ch/agoyal/
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) and span 1930–2014. We estimate consumption growth variance using an AR(1)-GARCH(1,1) model for consumption growth.

3.2 Regression analysis

Table 2 shows the results of regressions of one year \((t \to t+1)\) excess stock market returns on top share measures \((t)\), some classic return predictors \((t)\), and macro factors \((t)\). In column (1) we find that when the top 1% income share (January to December of year \(t\)) rises above trend by one percentage point, subsequent one year market excess returns (January to December of year \(t+1\)) decline on average by 2.8%. The coefficient is significant at the 5% level (using a Newey-West standard error), and the R-squared statistic is .04. It is clear, at least in sample, that the detrended top 1% share series forecasts the subsequent overall excess return on the stock market. Column (2) shows that the results are nearly identical using an AR(2) cyclical component. In column (4), we see that the 0.1% coefficient is -4.66 and significant at the 1% level with an R-squared of .05, so the relationship is even stronger for the top 0.1% income share.

In Table 3 and Table 4, we see that the inverse relationship between the top 1% share and subsequent excess returns is larger in magnitude and significant at the 1% level when we predict with \(cgdiff\) or use five year excess returns. Figures 5a and 5b show the corresponding scatter and time series plots for five year returns. The top 1% income share seems to predict subsequent five year excess returns well except around 1970 and the 1980s.

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 likely

14http://research.stlouisfed.org/fred2/
Table 2: Regressions of one year excess stock market returns on top income shares and other predictors

<table>
<thead>
<tr>
<th>Regressors (t)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>8.49</td>
<td>8.40</td>
<td>8.44</td>
<td>9.10</td>
<td>7.62</td>
<td>8.75</td>
<td>22.65</td>
<td>19.10</td>
<td>9.53</td>
<td>25.18</td>
</tr>
<tr>
<td></td>
<td>(1.76)</td>
<td>(1.76)</td>
<td>(1.74)</td>
<td>(1.73)</td>
<td>(2.36)</td>
<td>(1.53)</td>
<td>(13.35)</td>
<td>(10.86)</td>
<td>(1.88)</td>
<td>(15.44)</td>
</tr>
<tr>
<td>Top 1%</td>
<td>-2.82**</td>
<td></td>
<td></td>
<td></td>
<td>-4.11**</td>
<td>-4.69***</td>
<td>-2.22*</td>
<td>-2.60**</td>
<td>-4.23***</td>
<td>-4.86***</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td></td>
<td></td>
<td></td>
<td>(1.63)</td>
<td>(1.62)</td>
<td>(1.18)</td>
<td>(2.12)</td>
<td>(1.78)</td>
<td>(1.76)</td>
</tr>
<tr>
<td>Top 1% (p = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.64**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1% (no cg)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 0.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4.66***</td>
</tr>
<tr>
<td>Real GDP Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔCons. Growth Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(P/D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(P/E)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.04</td>
<td>.04</td>
<td>.02</td>
<td>.05</td>
<td>.06</td>
<td>.09</td>
<td>.05</td>
<td>.04</td>
<td>.13</td>
<td>.19</td>
</tr>
</tbody>
</table>

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 3: Regressions of one year excess stock market returns on cgdiff (top 1% – top 1% (no cg)) and other predictors

<table>
<thead>
<tr>
<th>Regressors (t)</th>
<th>Dependent Variable: ( t ) to ( t + 1 ) Excess Stock Market Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>(2.91)</td>
</tr>
<tr>
<td>cgdiff</td>
<td>-3.51***</td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
</tr>
<tr>
<td>cgdiff (top 0.1%)</td>
<td>-4.23**</td>
</tr>
<tr>
<td></td>
<td>(1.66)</td>
</tr>
<tr>
<td>Real GDP Growth</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
</tr>
<tr>
<td>( \Delta )Cons. Growth Variance</td>
<td>-10.14</td>
</tr>
<tr>
<td></td>
<td>(15.10)</td>
</tr>
<tr>
<td>log(P/D)</td>
<td>-0.71</td>
</tr>
<tr>
<td></td>
<td>(5.13)</td>
</tr>
<tr>
<td>log(P/E)</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>(4.26)</td>
</tr>
<tr>
<td>CAY</td>
<td>1.46*</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
</tr>
<tr>
<td>Sample</td>
<td>1913-</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.05</td>
</tr>
</tbody>
</table>

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). “no cg” refers to the series that excludes capital gains. cgdiff (top 0.1%) is calculated analogously, but with the top 0.1% income share. 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 4: Regressions of five year excess stock market returns on top income shares and other predictors

<table>
<thead>
<tr>
<th>Regressors (t)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-6.99</td>
<td>-6.79</td>
<td>7.04</td>
<td>7.64</td>
<td>7.91</td>
<td>7.64</td>
<td>20.24</td>
<td>14.38</td>
<td>7.84</td>
<td>28.01</td>
</tr>
<tr>
<td></td>
<td>(1.35)</td>
<td>(1.36)</td>
<td>(1.25)</td>
<td>(1.30)</td>
<td>(1.45)</td>
<td>(1.59)</td>
<td>(7.81)</td>
<td>(7.14)</td>
<td>(1.02)</td>
<td>(4.82)</td>
</tr>
<tr>
<td>Top 1%</td>
<td>-3.19***</td>
<td>-2.77**</td>
<td>-2.53**</td>
<td>-3.03***</td>
<td>-3.93***</td>
<td>-2.72***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(1.04)</td>
<td>(1.11)</td>
<td>(1.15)</td>
<td>(0.96)</td>
<td>(1.03)</td>
<td>(0.87)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1% (p = 2)</td>
<td></td>
<td>-2.82***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1% (no cg)</td>
<td>-5.71**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 0.1%</td>
<td>-4.97***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real GDP Growth</td>
<td>-0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.23)</td>
<td></td>
</tr>
<tr>
<td>∆Cons. Growth Variance</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.96</td>
<td>(8.00)</td>
</tr>
<tr>
<td></td>
<td>(5.42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(8.00)</td>
<td></td>
</tr>
<tr>
<td>log(P/D)</td>
<td>-4.11*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6.07***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.45)</td>
<td></td>
</tr>
<tr>
<td>log(P/E)</td>
<td>-2.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td>1.74***</td>
<td>1.62***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>1913-</td>
<td>1913-</td>
<td>1913-</td>
<td>1913-</td>
<td>1930-</td>
<td>1935-</td>
<td>1913-</td>
<td>1913-</td>
<td>1945-</td>
<td>1945-</td>
</tr>
<tr>
<td></td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
<td>-2010</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.26</td>
<td>.25</td>
<td>.20</td>
<td>.32</td>
<td>.18</td>
<td>.16</td>
<td>.30</td>
<td>.28</td>
<td>.52</td>
<td>.64</td>
</tr>
</tbody>
</table>

Note: Newey-West standard errors in parentheses ($k = 8$). ***, **, 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. Five year excess returns are annualized.
not lagged top shares. One might still worry that our results are driven by the Kalman filter. However, as we see in Appendix B, we get similar results with other filters. Table 12 in Appendix B repeats Table 2 but with the HP filter, yielding nearly uniformly larger and more significant effects. Table 11 shows that we get similar results using the one-sided HP filter, the 10 year moving average filter, and linear detrending, and Table 3 shows that our findings are robust to using cgdiff, which requires no trend estimate. Furthermore, as we see in Table 3, 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 .02 to .11. This is not the case, however, when using the HP filter or five year returns. Comparing columns (1) and (3) of Tables 12 and 4, we see that with the HP filter or five year returns excluding capital gains increases the magnitude of the association and maintains significance (at the 5% level).

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 market excess 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, and capital gains are the main avenue through which prices would determine inequality. Second, as we see in regressions (7) and (8) from Table 2, top shares predict excess returns even when controlling for the log price-dividend or price-earnings ratio. Including these controls does decrease the top shares coefficients slightly, but they remain large and significant. The P/D and P/E ratios are not significant after controlling for top income shares. Moreover, as we see in columns (5) and (6) of Table 3, controlling for P/D and P/E barely affects the result for cgdiff, which should be most susceptible to this bias.

In regressions (5), (6), (9), and (10) from Table 2, we also control for real GDP growth, consumption growth variance (Lettau et al. 2008) and Bansal et al. (2014) and CAY, which Lettau and Ludvigson (2001) show forecasts market excess returns. Including these controls does decrease the top shares coefficients slightly, but they remain large and significant. The P/D and P/E ratios are not significant after controlling for top income shares. When controlling for CAY, consumption growth variance, GDP growth, or all three and log(P/D) (column (10)), the 1% coefficient is between -4 and -5 and is significant at the 5% or 1% level. Again, the results are similar with cgdiff, five year returns, or the HP filter.

Our empirical analysis thus far has relied on detrending, which requires the researcher to take a stance on the underlying trend model. Do the raw data indicate a relationship between asset prices and the one percent? Figure 6 suggests that the answer is the yes. Over 1913-2014, both overall and within subsamples, there is a clear positive correlation between the top 1% income share (not detrended). We difference consumption growth variance, which appears nonstationary, and drop 1930-1934 to reduce the impact of the initial variance. Table 13 in Appendix B shows the pairwise correlations between the explanatory variables in Tables 2, 3, and 4.
trended) 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 detrending. Indeed, as we have shown, above trend inequality predicts subsequent excess returns even when using a simple, one-sided trend estimation method like the ten year moving average.

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

![Figure 6: 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 (+).](image)

### 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,

\[ y_{t+h} = \beta' x_t + \epsilon_{t+h}, \]

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, \( \epsilon_{t+h} \) is the error term, and \( \beta \) 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. Welch and Goyal (2008) consider the simplest possible case where \( x_{1t} \equiv 1 \) (constant) and \( x_{2t} \) consists of a single predictor.

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:

$$F = \frac{1}{\hat{\sigma}^2} \sum_{t=\lceil \rho T \rceil + 1}^{T} \left[ (y_{t+h} - \hat{y}_{t+h|t})^2 - (y_{t+h} - \hat{y}_{t+h|t})^2 \right], \quad (3.2)$$

where $\hat{\sigma}^2$ is a consistent estimator of $\text{Var}[\epsilon_{t+h}]$ (which we estimate from the sample average of the squared OLS residuals of (3.1) using the whole sample), $\hat{y}_{t+h|t} = \hat{\beta}_t'x_t$ ($\hat{y}_{t+h|t} = \hat{\beta}_{1t}'x_t$) 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 $\beta$ and $\beta_1$. 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 NULL model, we consider three choices, a constant, log price-dividend ratio ($\log(P/D)$), and log price-earnings ratio ($\log(P/E)$). 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. For the top income measure, we consider the top 1% series detrended with the Kalman filter with AR(1) cycle as well as the difference between the top 1% series with and without capital gains ($\text{cgdiff}$).

The choice of the proportion of the training sample, $\rho$, is necessarily subjective. Small $\rho$ leads to imprecise initial estimates of $\beta$, and large $\rho$ leads to the loss of power. Hence we simply report results for $\rho = 0.2, 0.3, 0.4$.

Table 5 shows the results. We can see that across specifications, the out-of-sample $F$ statistic is positive and significant. Results are stronger with $\text{cgdiff}$ than with the Kalman filter. (Note that since the asymptotic distribution of $F$ depends on the NULL model, the value of the $F$ statistic in Table 5 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. Figures 7 and 8 use the Kalman filter series and $\text{cgdiff}$, respectively. 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 except for when using $\log(P/D)$ in the NULL model, 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 6) but the stock market did not suffer. As for $\log(P/D)$, Welch and Goyal (2008) document that most of the prediction gains stem from the 1973–1975 Oil Shock, so it is not surprising that we see drops in Figures 7c and 8c around that period.

In summary, the top income series seem to predict returns out-of-sample as well, especially at the 1 year horizon.
Table 5: Out-of-sample performance of the top 1% series in predicting subsequent excess returns

<table>
<thead>
<tr>
<th>Top 1% Measure</th>
<th>Predictor in the NULL Model</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ρ</td>
<td>∅</td>
<td>(5yr)</td>
<td>log(P/D)</td>
<td>log(P/E)</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>3.16***</td>
<td>7.47***</td>
<td>0.98**</td>
<td>2.46**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0071)</td>
<td>(0.0155)</td>
<td>(0.0357)</td>
<td>(0.0118)</td>
</tr>
<tr>
<td>Kalman, p = 1</td>
<td>0.3</td>
<td>2.53**</td>
<td>3.30*</td>
<td>0.92**</td>
<td>1.88**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0115)</td>
<td>(0.0519)</td>
<td>(0.0462)</td>
<td>(0.0219)</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>2.13**</td>
<td>4.15**</td>
<td>0.97**</td>
<td>1.63**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0180)</td>
<td>(0.0416)</td>
<td>(0.0498)</td>
<td>(0.0258)</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>5.81***</td>
<td>7.38**</td>
<td>2.53**</td>
<td>4.13***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0005)</td>
<td>(0.0158)</td>
<td>(0.0110)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>4.73***</td>
<td>3.18**</td>
<td>2.30**</td>
<td>3.30***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0010)</td>
<td>(0.0594)</td>
<td>(0.0131)</td>
<td>(0.0039)</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>4.41***</td>
<td>5.45**</td>
<td>2.67***</td>
<td>3.38***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0014)</td>
<td>(0.0324)</td>
<td>(0.0083)</td>
<td>(0.0029)</td>
</tr>
</tbody>
</table>

Note: ρ = 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 NULL model in addition to a constant (∅ indicates no additional regressors). The column “Top 1% Measure” indicates the additional regressor in the ALTERNATIVE model. 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. Except for column (2), all regressions are 1-year horizon. Column (2) is five year horizon.

3.4 Using tax policy as an 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 (see, for example, Roine et al. (2009)) suggests that increases (decreases) in top marginal tax rates reduce (exacerbate) inequality. 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 or anti-deficit ideologies (de Rugy 2003a,b; Weinzierl and Werker 2009; Jacobson et al.; 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.

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 [6] shows how top tax rates, the top
(a) Constant.

(b) Constant (5 year returns).

(c) Constant, log(P/D).

(d) Constant, log(P/E).

Figure 7: Annual performance of the top 1% series (Kalman filter, $p = 1$) 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 includes the predictor variables specified in each subcaption. The ALTERNATIVE model also includes the detrended top 1% series (Kalman filter with AR(1) cycle). Except for Figure 7b, all regressions use 1-year excess returns. Figure 7b uses five year excess returns. Predictions start at $t = ⌊ρT⌋$, where $T$ is the sample size and $ρ = 0.2, 0.3, 0.4$.

1% income share, and Robert Shiller’s P/E10 ratio 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

Figure 8: Annual performance of the cgdiff top 1% series (difference between the raw series with or without capital gains) in predicting subsequent excess returns.

Note: see caption of Figure 7 for detailed explanations.

(annualized) of -2.88%. In tax cut periods, when top income shares rose, Shiller’s 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 6, 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 6, 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%) \(^{18}\) 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.

4 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.

4.1 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 (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, DN, 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 10, 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 Datasstream 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.

\(^{18}\)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 5 the 1% share strongly predicts excess returns even when controlling for GDP growth.
Table 6: Top 1% income share and stock prices during and after top tax rate change episodes

<table>
<thead>
<tr>
<th>Period</th>
<th>∆MTR</th>
<th>∆ETR</th>
<th>∆1%</th>
<th>ER₅yr</th>
<th>ER₅yr⁺t</th>
<th>%∆(P/E10)</th>
<th>%∆IP</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1915-1919</td>
<td>66</td>
<td>25</td>
<td>-0.36</td>
<td>2.78</td>
<td>2.54</td>
<td>-17.74</td>
<td>N/A</td>
<td>WWI</td>
</tr>
<tr>
<td>1921-1927</td>
<td>-48</td>
<td>-5</td>
<td>0.65</td>
<td>-8.65</td>
<td>-8.86</td>
<td>18.07</td>
<td>8.44</td>
<td>pro-growth</td>
</tr>
<tr>
<td>1931-1945</td>
<td>69</td>
<td>57</td>
<td>-0.27</td>
<td>8.61</td>
<td>4.70</td>
<td>3.86</td>
<td>7.18</td>
<td>WWII, budget balance</td>
</tr>
<tr>
<td>1963-1966</td>
<td>-21</td>
<td>0</td>
<td>0.19</td>
<td>-3.39</td>
<td>-1.00</td>
<td>-1.93</td>
<td>8.17</td>
<td>pro-growth</td>
</tr>
<tr>
<td>1980-1989</td>
<td>-42</td>
<td>-15</td>
<td>0.52</td>
<td>7.38</td>
<td>8.22</td>
<td>6.99</td>
<td>2.27</td>
<td>Reaganomics</td>
</tr>
<tr>
<td>1990-1994</td>
<td>11.6</td>
<td>0</td>
<td>-0.04</td>
<td>17.83</td>
<td>19.06</td>
<td>5.93</td>
<td>2.42</td>
<td>budget balance</td>
</tr>
<tr>
<td>2000-2004</td>
<td>-4.6</td>
<td>-7</td>
<td>-0.29</td>
<td>-6.86</td>
<td>-6.25</td>
<td>-8.01</td>
<td>0.12</td>
<td>pro-growth, stimulus</td>
</tr>
</tbody>
</table>

Across episode averages

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hikes</td>
<td>-0.24</td>
<td>9.74</td>
<td>8.77</td>
<td>0.00</td>
<td>6.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuts</td>
<td>0.36</td>
<td>-2.88</td>
<td>-1.97</td>
<td>6.05</td>
<td>4.37</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: MTR and ETR: top marginal income and estate tax rates (%). ∆1%: average per year change in top 1% income share including capital gains. ER₅yr: annualized five year average excess return (%), starting in final year of period. ER₅yr⁺t: ER₅yr, taxing interest and dividends at top marginal income rate. %∆(P/E10): average per year % change in Shiller’s P/E. %∆IP: average per year % change in the industrial production index. Sources: de Rugy (2003a,b), Weinzierl and Werker (2009), Jacobson et al., Romer and Romer (2010), Tax Foundation, IRS, and FRED.
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.

4.2 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 2 with stock returns instead of excess returns, the 1% coefficient is -2.20 with a Newey-West p-value of .085. With cgdiff and the HP filter, the coefficients are, respectively, -2.93 and -4.57 with p-values of .041 and .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 10, where we use cgdiff).

Table 7 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.

As a robustness check, Table 14 in Appendix E shows the panel regressions without time trends. The results are similar to the case with the linear time trend.

19 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 .31), Indonesia (p-value of .31), and South Africa (p-value of .05), all three of which have small sample size (≤ 12).

20 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 7: Country fixed effects panel regressions of one year stock market returns on top income shares

<table>
<thead>
<tr>
<th>Regressors</th>
<th>All</th>
<th>Advanced</th>
<th>IIPS</th>
<th>EME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1% (t)</td>
<td>-1.99**</td>
<td>-1.41*</td>
<td>-7.16*</td>
<td>-6.58**</td>
</tr>
<tr>
<td>Time Trend</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Country FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>790</td>
<td>699</td>
<td>106</td>
<td>91</td>
</tr>
<tr>
<td>R^2 (w,b)</td>
<td>(.01,.08)</td>
<td>(.00,.07)</td>
<td>(.05,.16)</td>
<td>(.05,.15)</td>
</tr>
</tbody>
</table>

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).

Table 8 is the same as Table 7 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 .16).

Table 8: Country fixed effects panel regressions of one year excess returns on top income shares

<table>
<thead>
<tr>
<th>Regressors</th>
<th>All</th>
<th>Advanced</th>
<th>IIPS</th>
<th>EME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1% (t)</td>
<td>-1.49†</td>
<td>-0.63</td>
<td>-1.42</td>
<td>-6.55**</td>
</tr>
<tr>
<td>Time Trend</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Country FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>660</td>
<td>569</td>
<td>72</td>
<td>91</td>
</tr>
<tr>
<td>R^2 (w,b)</td>
<td>(.00,.01)</td>
<td>(.00,.20)</td>
<td>(.01,.20)</td>
<td>(.05,.09)</td>
</tr>
</tbody>
</table>

Note: Clustered standard errors in parentheses, ***1%, **5%, *10%, †15%. R^2 (w,b): Within and between R-squared. Constants suppressed. †: p-value = .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).

In Tables 2, 7, 8, and 14 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).

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 9 repeats the regressions of Table 8 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 9: Country fixed effects panel regressions of one year excess returns on local and U.S. top income shares

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Dependent Variable: t to t + 1 Excess Market Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All†</td>
</tr>
<tr>
<td>Top 1% (t)</td>
<td>-1.62</td>
</tr>
<tr>
<td></td>
<td>(1.17)</td>
</tr>
<tr>
<td>U.S. Top 1% (t)</td>
<td>-3.37***</td>
</tr>
<tr>
<td></td>
<td>(1.02)</td>
</tr>
<tr>
<td>Time Trend</td>
<td>Yes</td>
</tr>
<tr>
<td>Country FE</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>616</td>
</tr>
<tr>
<td>$R^2$ (w,b)</td>
<td>(.02,.01)</td>
</tr>
</tbody>
</table>

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 10 we restrict the sample to the U.S., Japan, Canada, and Germany 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% 21. 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.

5 Concluding remarks

In this paper we built general equilibrium models 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.

21 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.
Table 10: Country fixed effects panel regressions of one year returns on local and U.S. cgdiff (top 1% − top 1% (no cg))

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Dependent Variable: $t$ to $t+1$ Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R$</td>
</tr>
<tr>
<td>cgdiff ($t$)</td>
<td>-2.19***</td>
</tr>
<tr>
<td>U.S. cgdiff ($t$)</td>
<td>-</td>
</tr>
<tr>
<td>Time Trend</td>
<td>No</td>
</tr>
<tr>
<td>Country FE</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>143</td>
</tr>
<tr>
<td>$R^2$ (w,b)</td>
<td>(.01,.04)</td>
</tr>
</tbody>
</table>

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).

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 market excess 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.

References


A Proof

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 $A = [e, A_2, \ldots, A_J]$. 

35
Let \( n_i = (w_i, 0, \ldots, 0)' \) 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 \( q'z \leq q'n_i \) and \( x \leq Az \). Therefore the utility maximization problem becomes equivalent to

\[
\text{maximize} \quad U_i(x) \\
\text{subject to} \quad q'z \leq q'n_i, \quad 0 \leq x \leq Az.
\]

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

\[
\text{maximize} \quad \sum_{i=1}^I w_i \log U_i(Az_i) \\
\text{subject to} \quad \sum_{i=1}^I z_i = n,
\]

where \( n = \sum_{i=1}^I n_i = (1, 0, \ldots, 0)' \) 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

\[
\nabla f(x) = \frac{1}{\Sigma} \begin{bmatrix} \pi_1 x_1^{1-\gamma} \\
\vdots \\
\pi_S x_S^{1-\gamma} \end{bmatrix},
\]

\[
\nabla^2 f(x) = -\frac{1-\gamma}{\Sigma^2} \begin{bmatrix} \pi_1 x_1^{1-\gamma} \\
\vdots \\
\pi_S x_S^{1-\gamma} \end{bmatrix}^T \begin{bmatrix} \pi_1 x_1^{1-\gamma} & \cdots & \pi_S x_S^{1-\gamma} \end{bmatrix}
\]

\[
+ \frac{1}{\Sigma} \text{diag} [-\gamma \pi_1 x_1^{-\gamma-1} \cdots -\gamma \pi_S x_S^{-\gamma-1}].
\]

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 \( h = (h_1, \ldots, h_S)' \) 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 \( u = (\cdots (\pi_s x_s^{-1})^{\frac{1}{2}} \cdots)' \) and \( v = (\cdots (\pi_s x_s^{-\gamma})^{\frac{1}{2}} h_s \cdots)' \). Then the above expression becomes

\[
h'[-\Sigma^2 \nabla^2 f(x)]h = (1-\gamma)(u \cdot v)^2 + \gamma \|u\|^2 \|v\|^2
\]

\[
= \gamma(\|u\|^2 \|v\|^2 - (u \cdot v)^2) + (u \cdot v)^2 \geq 0,
\]

36
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 \( k \|v\|^2 = 0 \), so \( k = 0 \) and therefore \( h = 0 \). Hence \( f = \log U \) is strictly concave.

\[ \text{Step 2. The planner’s problem } (2.2) \text{ has a unique solution.} \]

\[ \text{Proof. Let} \]

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

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 \( \Omega \) is nonempty, compact, and convex. Clearly \( \Omega \neq \emptyset \) because we can choose the initial endowment \( z_i = n_i \) and \( x_i = A n_i = \epsilon_i \). Since \( \Omega \) 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

\[ \sum_{i=1}^I x_i \leq \sum_{i=1}^I A z_i = A \sum_{i=1}^I z_i = A n = e. \]

Since \( x_i \geq 0 \) and \( e \gg 0 \), \( \Omega \) is bounded.

Let \( x = (x_i) \) be the unique maximizer of \( f \) on \( \Omega \). 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. \( \square \)

\[ \text{Step 3. } x = (x_i) \text{ is a GEI equilibrium allocation and the Lagrange multiplier to the planner’s problem gives the asset prices.} \]

\[ \text{Proof. Let} \]

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

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 \( A z_i \gg 0 \). Hence by the Karush-Kuhn-Tucker theorem and the chain rule, we have

\[ q' = w_i \frac{D U_i(A z_i) A}{U_i(A z_i)} \]

for all \( i \), where \( D U_i \) 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 \( \lambda \) and setting \( \lambda = 1 \), we have \( D U_i(x) x = U_i(x) \). Hence multiplying \( z_i \) from the right to \( (A.3) \), we get

\[ q' z_i = w_i \frac{D U_i(A z_i) A z_i}{U_i(A z_i)} = w_i. \]
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

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

so the budget constraint holds with equality. Furthermore, letting $\lambda_i = \frac{1}{wi}$, 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.

**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

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

Since $DU_i \gg 0$, $A = [e, A_2, \ldots, A_J]$, 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 $q'n = 1 \cdot 1 + q_2 \cdot 0 + \cdots + q_J \cdot 0 = 1$. 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

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

Substituting into (A.4), we obtain $q' = w_i D[\log U_i(Az_i)]$, which is precisely (A.5), 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. 

**A.2 Proof of Theorem 2.2 and Proposition 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

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

where $w$ is initial wealth. The following lemma is basic.

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

2. θ ≥ 0 according as \( E[R] ≥ R_f \).

3. Suppose \( E[R] > R_f \). If \( u \) exhibits decreasing relative risk aversion (DRRA), so \(-xu''(x)/u'(x)\) is decreasing, then \( \partial \theta / \partial w ≥ 0 \), 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 \( f(\theta) = E[u(R(\theta)w)] \). Then \( f'(\theta) = E[u'(R(\theta)w)(R-R_f)w] \) and \( f''(\theta) = E[u''(R(\theta)w)(R-R_f)^2 w^2] < 0 \), so \( f \) is strictly concave. Therefore the optimal \( \theta \) is unique (if it exists).

2. Since \( f'(\theta) = 0 \) and \( f'(0) = u'(R_f w)\), the result follows.

3. Dividing the first-order condition by \( w \), we obtain \( E[u'(R(\theta)w)(R-R_f)] = 0 \). 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 \( F_\theta = E[u''(R(\theta)w)(R-R_f)^2 w] < 0 \), it suffices to show \( F_w ≥ 0 \). Let \( γ(x) = -xu''(x)/u'(x) > 0 \) be the relative risk aversion coefficient. Then

\[
F_w = E[u''(R(\theta)w)(R-R_f)R(\theta)]
\]

\[
= -\frac{1}{w} E[γ(R(\theta)w)u'(R(\theta)w)(R-R_f)].
\]

Since \( E[R] > R_f \), by the previous result we have \( \theta > 0 \). Therefore \( R(\theta) = Rθ + R_f(1 - \theta) ≥ R_f \) according as \( R ≥ R_f \). Since \( u \) is DRRA, \( γ \) is decreasing, so \( γ(R(\theta)w) ≤ γ(R_f w) \) if \( R ≥ R_f \) (and reverse inequality if \( R ≤ R_f \)). Therefore

\[
γ(R(\theta)w)(R-R_f) ≤ γ(R_f w)(R-R_f)
\]

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

\[
wF_w = -E[γ(R(\theta)w)u'(R(\theta)w)(R-R_f)]
\]

\[
≥ -E[γ(R_f w)u'(R(\theta)w)(R-R_f)] = 0,
\]

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, \( θ_i ≥ 0 \) according as \( E[\theta_i] ≥ R_f \).

Suppose that \( θ_1 < θ_2 \) and we transfer some wealth \( \epsilon > 0 \) from agent 1 to 2. Let \( θ'_2 \) be the new portfolio of agent \( i \). The change in agent 1 and 2’s demand in the risky asset is

\[
Δ = (w_1 - \epsilon)θ'_1 + (w_2 + \epsilon)θ'_2 - (w_1θ_1 + w_2θ_2)
\]

\[
= w_1(θ'_1 - θ_1) + w_2(θ'_2 - θ_2) + \epsilon(θ'_2 - θ_1).
\]

Suppose that the risk-free rate does not change. Since agents have CRRA preferences, we have \( θ'_i = θ_i \), so \( Δ = \epsilon(θ_2 - θ_1) > 0 \). Since agents \( i > 2 \) are unaffected unless the risk-free rate changes, there is a positive excess demand in the risky asset.
Regard $\theta_i$ as a function of the risk-free rate $R_f$. By the maximum theorem, $\theta_i$ is continuous, and so is the aggregate demand. Since $\theta_i < 0$ if $R_f > E_i[R]$ by Lemma A.1, the aggregate excess demand of the risky asset becomes negative as $R_f \to \max_i E_i[R]$. Therefore by the intermediate value theorem, there exists an equilibrium risk-free rate higher than the original one. Since by Theorem 2.1 the equilibrium is unique, in the new equilibrium the risk-free rate is higher, and hence the equity premium is lower.

**Lemma A.2.** Consider two agents indexed by $i = 1, 2$ with common beliefs. Let $w_i(x)$, $\gamma_i(x) = -x u''_i(x)/u'_i(x)$, and $\theta_i$ be the initial wealth, utility function, relative risk aversion, and the optimal portfolio of agent $i$. Suppose that $\gamma_1(w_1) > \gamma_2(w_2)$ for all $x$, so agent 1 is more risk averse than agent 2. Then

$$E[R] > R_f \implies \theta_2 > \theta_1 > 0,$$

$$E[R] < R_f \implies \theta_2 < \theta_1 < 0,$$

so the less risk averse agent invests more aggressively.

**Proof.** Since $\gamma_1(w_1) > \gamma_2(w_2)$, we have

$$\frac{d}{dx} \left( \frac{u'_2(w_2)}{u'_1(w_1)} \right) = \frac{w_2 u'_2 u''_1 - u'_2 w_1 u''_1}{(u'_1)^2} = \frac{1}{w_1} \left( \gamma_1(w_1) - \gamma_2(w_2) \right) > 0,$$

so $u'_2(w_2)/u'_1(w_1)$ is increasing. Suppose $E[R] > R_f$. By Lemma A.1 we have $\theta_1 > 0$. Then $R(\theta_1) \geq R_f$ according as $R \geq R_f$. Since $u'_2(w_2)/u'_1(w_1)$ is increasing (and positive), we have

$$\frac{u'_2(R(\theta_1)w_2)}{u'_1(R(\theta_1)w_1)} (R - R_f) > \frac{u'_2(R_f w_2)}{u'_1(R_f w_1)} (R - R_f)$$

always (except when $R = R_f$). Multiplying both sides by $u'_1(R(\theta_1)w_1) > 0$ and taking expectations, we get

$$E[u'_2(R(\theta_1)w_2)(R - R_f)] = E \left[ \frac{u'_2(R(\theta_1)w_2)}{u'_1(R(\theta_1)w_1)} u'_1(R(\theta_1)w_1)(R - R_f) \right]$$

$$> E \left[ \frac{u'_2(R_f w_2)}{u'_1(R_f w_1)} u'_1(R(\theta_1)w_1)(R - R_f) \right]$$

$$= \frac{u'_2(R_f w_2)}{u'_1(R_f w_1)} E \left[ u'_1(R(\theta_1)w_1)(R - R_f) \right] = 0,$$

where the last equality uses the first-order condition for agent 1. Letting $f_2(\theta) = E[u_2(R(\theta)w_2)]$, 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 $E[R] < R_f$ is analogous.

**Proof of Proposition 2.3.** Since agents have common beliefs, we have $\theta_i \geq 0$ for all $i$ if $E[R] \geq R_f$. Since the stock is in positive supply, in equilibrium we must have $E[R] > R_f$. Therefore by Lemma A.2 if $\gamma_1 > \cdots > \gamma_f$, we have $0 < \theta_1 < \cdots < \theta_f$. 

| 40 |
Proof of Proposition 2.4. Let \( u(x) \) be the common CRRA utility function of agents 1 and 2. By the first-order condition, we have

\[
\sum_{s=u,d} \pi_{is} u'(R_f + \theta_i X_s) X_s = 0,
\]

where \( X_s = R_s - R_f \) denotes the excess return in state \( s \). Since \( u' > 0, s = u, d \), and \( e_u > e_d \), it must be that \( X_u > 0 > X_d \). Suppose \( \theta_1 \geq \theta_2 \). Since \( X_u > 0 \), we obtain

\[
R_f + \theta_1 X_u \geq R_f + \theta_2 X_u.
\]

Applying \( u'(\cdot) \) to both sides (which is a decreasing function) and multiplying by \( X_u > 0 \), we obtain

\[
0 < u'(R_f + \theta_1 X_u) X_u \leq u'(R_f + \theta_2 X_u) X_u.
\]

By repeating the same argument for the down state, we obtain

\[
u'(R_f + \theta_1 X_d) X_d \leq u'(R_f + \theta_2 X_d) X_d < 0.
\]

If \( \pi_{1d} > \pi_{2d} \), then \( \pi_{1u} < \pi_{2u} \) because there are only two states. Since \( u' > 0 \) and \( X_u > 0 > X_d \), we obtain

\[
0 = \sum_{s=u,d} \pi_{1s} u'(R_f + \theta_1 X_s) X_s < \sum_{s=u,d} \pi_{2s} u'(R_f + \theta_1 X_s) X_s
\]

\[
\leq \sum_{s=u,d} \pi_{2s} u'(R_f + \theta_2 X_s) X_s = 0,
\]

which is a contradiction. Therefore \( \theta_1 < \theta_2 \).

A.3 Proof of Theorem 2.5

First we characterize the equilibrium, assuming existence.

Individual problem. Let \( R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} \) be the gross return on stock (to be determined in equilibrium). Let \( V_{it}(w) \) be the value function of agent \( i \) at time \( t \). Since utility is logarithmic, we can guess that

\[
V_{it}(w) = a_{it} + \beta \left\{ \log R_{t+1} + \theta_1 (1 - \theta_1) (w - c) \right\},
\]

for some bounded random variable \( a_{it} \) (that does not affect the agent’s behavior).

The Bellman equation for agent 1 is

\[
V_{i1}(w) = \max_{c, \theta_1} \{ \log c + \beta \mathbb{E}_t [V_{1,t+1}((R_{t+1} \theta_1 + R_f)(1 - \theta_1))(w - c)] \},
\]

where \( \theta_1 \) is the fraction of agent 1’s wealth invested in the stock. Substituting the guess into the Bellman equation and taking the first-order condition with respect to consumption, we get \( c = (1 - \beta)w \). The optimal portfolio problem becomes

\[
\max_{\theta_1} \mathbb{E}_t [(R_{t+1} \theta_1 + R_f)(1 - \theta_1)].
\]
Suppressing the \( t \) subscript and using prime (\( ' \)) for time \( t + 1 \) variables, the first-order condition is

\[
E \left[ \frac{R' - R_f}{R'\theta_1 + R_f(1 - \theta_1)} \right] = 0.
\]

For agent 2, there is no portfolio choice \((\theta_2 = 1 \text{ always})\) and the consumption rule is \( c = (1 - \beta)w \).

**Market clearing** By the market clearing for the good, we have \( D = c_1 + c_2 = (1 - \beta)(w_1 + w_2) = (1 - \beta)W \). Since the only asset in positive supply is the stock, we have \( W = P + D \). Therefore \( D = (1 - \beta)(P + D) \iff P = \frac{\beta}{1-\beta}D \); the stock price depends only on the current dividend. Hence the stock return

\[
R' = \frac{P' + D'}{P} = \frac{\beta}{1-\beta}D' + \frac{D'}{\beta D} = \frac{D'}{\beta D}
\]
is exogenous.

Since agents save at rate \( \beta \) out of wealth, agent 1 invests a dollar amount of \((1 - \beta)w_1(1 - \theta_1)\) in the bond. Similarly, agent 2 invests \((1 - \beta)w_2\). Since the bond is in zero net supply, by market clearing we have

\[
(1 - \beta)w_1(1 - \theta_1) + (1 - \beta)w_2 = 0 \iff x(1 - \theta_1) + (1 - x) = 0 \iff \theta_1 = 1/x.
\]

Substituting into the first-order condition of agent 1’s portfolio problem, we obtain

\[
0 = E \left[ \frac{R' - R_f}{R'/x + R_f(1 - 1/x)} \right] = E \left[ \frac{(R' - R_f)x}{R' + R_f(x - 1)} \right].
\]

**(A.5)**

**Equilibrium** Given the history up to \( t \), the risk-free rate is determined by \((A.5)\), where \( R' = D'/\beta D \) is the stock return. By the budget constraints, we obtain

\[
w'_1 = \beta(R'/x + R_f(1 - 1/x))w_1,
\]
\[
w'_2 = \beta R_f w_2.
\]

Since \( W' = \beta R'W \), it follows that

\[
x' = \frac{w'_1}{W'} = \frac{R'/x + R_f(1 - 1/x)}{R'}x = 1 + \frac{R_f}{R'}(x - 1),
\]
which are the equations of motion.

Next, we show the existence and uniqueness of equilibrium. To this end, it suffices to show that \((A.5)\) has a unique solution \( R_f \). Let \( F(x, R_f) = E \left[ \frac{(R' - R_f)x}{R' + R_f(x - 1)} \right] \). Then \( F(x, 0) = x > 0 \), and letting \( R_f \to \frac{\text{int} R'}{1-\beta} \), we have \( F(x, R_f) = -\infty \). Clearly \( F(x, R_f) \) is continuous in \( R_f \). Therefore there exists
an equilibrium. To show uniqueness, note that

\[
\frac{\partial}{\partial R_f} F(x, R_f) = E \left[ \frac{-x(R' + R_f(x - 1)) - (R' - R_f)x(x - 1)}{(R' + R_f(x - 1))^2} \right]
\]

\[
= E \left[ \frac{-R'x^2}{(R' + R_f(x - 1))^2} \right] < 0,
\]

so \( F(x, R_f) \) is decreasing in \( R_f \).

Finally, we show that the equity premium \( E_t[R_{t+1} - R_{f,t}] \) is decreasing in \( x_t \). Since \( R_{t+1} = \frac{\partial W_{t+1}}{\partial W_t} \) is exogenous, it suffices to show that \( R_f \) is increasing in \( x \). By the implicit function theorem, we have \( \partial R_f/\partial x = -F_x/F_{R_f} \). Since \( F_{R_f} < 0 \) by the previous proof, it suffices to show that \( F_x > 0 \). Now

\[
\frac{\partial}{\partial x} F(x, R_f) = E \left[ \frac{(R' - R_f)(R' + R_f(x - 1)) - (R - R_f)xR_f}{(R' + R_f(x - 1))^2} \right]
\]

\[
= E \left[ \frac{(R' - R_f)^2}{(R' + R_f(x - 1))^2} \right] > 0,
\]

so the claim is proved.

**B Robustness of predictability**

Tables 12 and 11 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 13 shows the pairwise correlations between the explanatory variables used in Section 3.2.

As described in Section 3.2, Table 12 repeats the analysis of Table 2 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 2 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.

Table 11 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

22The technical condition \( \sup_t E[\log(D_{t+1}/D_t)] < \infty \) guarantees the transversality condition \( \lim_{t \to \infty} t \beta^t E[V_t(w_{it})] \leq 0 \). To see this, since \( W_t = P_t + D_t \) and \( P_t = \frac{1}{1+\beta} D_t \), we have \( W_t = \frac{1}{1+\beta} D_t \), so in particular \( W_{t+1}/W_t = D_{t+1}/D_t \). Taking logs, expectations, and summing over \( t \), it follows that \( E[\log W_t] - \log W_0 = \sum E[\log(D_{t+1}/D_t)] \). If \( \sup_t E[\log D_{t+1}/D_t] \leq M < \infty \), then \( E[\log W_t] \leq \log W_0 + Mt \). Since \( V_t(w_{it}) = a_{it} + \frac{2}{\beta} \log w_{it} \), \( a_{it} \) is bounded, \( 0 < \beta < 1 \), and \( w_{it} \leq W_t \), the transversality condition holds.
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 what extent then are the top income shares correlated with classic return predictors?

Table 11: Regressions of one year excess stock market returns on top income shares (using different trend estimates)

<table>
<thead>
<tr>
<th>Dependent Variable: t to t + 1 Excess Stock Market Return Regressors (t)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>7.97</td>
<td>8.16</td>
<td>8.78</td>
</tr>
<tr>
<td>(1.89)</td>
<td>(1.68)</td>
<td>(1.87)</td>
<td></td>
</tr>
<tr>
<td>Top 1% (one-sided HP)</td>
<td>-3.63*</td>
<td>-3.11***</td>
<td></td>
</tr>
<tr>
<td>(1.87)</td>
<td>(0.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1% (linear detrending)</td>
<td>-1.97**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>1936-2014</td>
<td>1913-2014</td>
<td>1922-2014</td>
</tr>
<tr>
<td>R²</td>
<td>.04</td>
<td>.09</td>
<td>.04</td>
</tr>
</tbody>
</table>

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.

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

\[
(1 - L)^2 g_t = \epsilon_t, \quad \epsilon_t \sim i.i.d. \ N(0, \sigma^2_\epsilon), \tag{C.2a}
\]

\[
\varphi(L) u_t = w_t, \quad w_t \sim i.i.d. \ N(0, \sigma^2_w), \tag{C.2b}
\]

44
Table 12: Regressions of one year excess stock market returns on top income shares and other predictors (using HP filter)

<table>
<thead>
<tr>
<th>Regressors (t)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>8.10</td>
<td>8.10</td>
<td>8.12</td>
<td>8.11</td>
<td>5.59</td>
<td>7.57</td>
<td>24.54</td>
<td>22.97</td>
<td>7.74</td>
<td>28.97</td>
</tr>
<tr>
<td></td>
<td>(1.73)</td>
<td>(1.72)</td>
<td>(1.73)</td>
<td>(1.70)</td>
<td>(2.34)</td>
<td>(1.59)</td>
<td>(13.26)</td>
<td>(11.30)</td>
<td>(1.67)</td>
<td>(14.50)</td>
</tr>
<tr>
<td>Top 1%</td>
<td>-4.21***</td>
<td>-5.22***</td>
<td>-5.54***</td>
<td>-3.93***</td>
<td>-4.24***</td>
<td>-4.03***</td>
<td>-4.21***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.24)</td>
<td>(1.70)</td>
<td>(1.50)</td>
<td>(1.18)</td>
<td>(1.26)</td>
<td>(1.46)</td>
<td>(1.46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1% (HP param.= 10)</td>
<td></td>
<td>-5.89***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1% (no cg)</td>
<td></td>
<td>-4.40**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 0.1%</td>
<td></td>
<td></td>
<td>-6.00***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real GDP Growth</td>
<td></td>
<td>0.60*</td>
<td></td>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.36)</td>
<td></td>
<td></td>
<td>(0.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔCons. Growth Variance</td>
<td>-16.28</td>
<td>-16.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.90)</td>
<td></td>
<td></td>
<td>(15.76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(P/D)</td>
<td></td>
<td>-5.02</td>
<td></td>
<td></td>
<td>-6.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.89)</td>
<td></td>
<td></td>
<td>(4.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(P/E)</td>
<td></td>
<td>-5.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.59***</td>
<td>1.50**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.58)</td>
<td>(0.70)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>1913-</td>
<td>1913-</td>
<td>1913-</td>
<td>1913-</td>
<td>1930-</td>
<td>1935-</td>
<td>1913-</td>
<td>1913-</td>
<td>1945-</td>
<td>1945-</td>
</tr>
<tr>
<td>R(^2)</td>
<td>.09</td>
<td>.10</td>
<td>.03</td>
<td>.09</td>
<td>.10</td>
<td>.13</td>
<td>.10</td>
<td>.10</td>
<td>.14</td>
<td>.20</td>
</tr>
</tbody>
</table>

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 13: Pairwise correlations between explanatory variables

<table>
<thead>
<tr>
<th></th>
<th>Top 1%</th>
<th>Top 1% (p = 2)</th>
<th>Top 1% (no cg)</th>
<th>Top 0.1%</th>
<th>cgdiff</th>
<th>cgdiff (0.1%)</th>
<th>%ΔRGDP</th>
<th>ΔCGV</th>
<th>log(P/D)</th>
<th>log(P/E)</th>
<th>CAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1%</td>
<td>1.00</td>
<td>0.94*</td>
<td>0.87*</td>
<td>0.95*</td>
<td>0.73*</td>
<td>0.68*</td>
<td>0.12</td>
<td>-0.22</td>
<td>0.37*</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td>Top 1% (p = 2)</td>
<td></td>
<td>1.00</td>
<td>0.83*</td>
<td>0.91*</td>
<td>0.78*</td>
<td>0.76*</td>
<td>0.07</td>
<td>-0.16</td>
<td>0.48*</td>
<td>0.23*</td>
<td>0.16</td>
</tr>
<tr>
<td>Top 1% (no cg)</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.83*</td>
<td>0.41*</td>
<td>0.39*</td>
<td>0.03</td>
<td>-0.19</td>
<td>0.26*</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>Top 0.1%</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.70*</td>
<td>0.67*</td>
<td>0.13</td>
<td>-0.20</td>
<td>0.33*</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>cgdiff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.99*</td>
<td>0.01</td>
<td>-0.08</td>
<td>0.68*</td>
<td>0.44*</td>
<td>0.17</td>
</tr>
<tr>
<td>cgdiff (0.1%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.67*</td>
<td>0.99*</td>
<td>-0.05</td>
<td>0.70*</td>
<td>0.46*</td>
<td>0.46</td>
</tr>
<tr>
<td>%ΔRGDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.21</td>
<td>-0.10</td>
</tr>
<tr>
<td>ΔCGV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td>0.74*</td>
<td>1.00</td>
</tr>
<tr>
<td>log(P/D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(P/E)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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% 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. Top 0.1% is defined analogously. The top shares series are detrended with the Kalman filter (p = 1) unless otherwise noted. “cgdiff” is top 1% minus top 1% (no cg), neither detrended. cgdiff (top 0.1%) is 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.
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
\[
\begin{bmatrix}
g_t \\
g_{t-1}
\end{bmatrix} =
\begin{bmatrix}
2 & -1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
g_{t-1} \\
g_{t-2}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_t \\
0
\end{bmatrix},
\]
(C.3a)
\[
y_t = \phi_1 y_{t-1} + g_t - \phi_1 g_{t-1} + w_t.
\]
(C.3b)
Letting \( \xi_t = (g_t, g_{t-1}')', \quad v_t = (\epsilon_t, 0)', \quad x_t = y_{t-1}, \quad A = \phi_1, \quad F =
\begin{bmatrix}
2 & -1 \\
1 & 0
\end{bmatrix}, \quad \text{and}\]
\[
H = \begin{bmatrix} 1 & -\phi_1 \end{bmatrix},
\]
(C.3) reduces to
\[
\xi_t = F \xi_{t-1} + v_t,
\]
(C.4a)
\[
y_t = Ax_t + H \xi_t + w_t.
\]
(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)
   Local Currency Deposit Rate 30-59 day deposit rate.

2. Australia (AUS)
   Local Currency Deposit Rate 1972-2011.

3. Canada (CAN)
   1% Income Share LAD series post-1995.
   Local Currency Deposit Rate 90 day deposit rate. 1971-2011.

4. China (CHN)
   Local Currency Deposit Rate 1 year deposit rate.

5. Colombia (COL)

6. Denmark (DNM)
1% Income Share “Adults” series.

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)

9. Germany (GER)
Local Currency Deposit Rate 3 month deposit rate, 1978-2003.
Price Index GDP deflator pre-1991.

10. India (INI)
Local Currency Deposit Rate Bank discount rate from Haver.

11. Indonesia (INO)
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)

14. Japan (JPN)
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)
Local Currency Deposit Rate 3 month deposit rate.
Price Index blahblah

17. Mauritius (MAU)
Local Currency Deposit Rate 3 month deposit rate.

18. Netherlands (NET)

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)
Local Currency Deposit Rate 1979-2010.

21. Portugal (POR)
Local Currency Deposit Rate 180-360 day deposit rate, 1990-2000.

22. South Africa (SAF)
Local Currency Deposit Rate 88-91 day deposit.

23. Singapore (SIN)
Local Currency Deposit Rate 3 month deposit rate, 1977-2013.
Price Index blahbla

24. Spain (SPA)
Coverage 1982-2013.
Local Currency Deposit Rate 6-12 month deposit rate, 1982-2013.

25. Sweden (SWE)
Local Currency Deposit Rate 1970-2006.

26. Switzerland (SWI)
Local Currency Deposit Rate 3 month deposit rate,
Price Index 1982-2011.

27. Taiwan (TAI)
Local Currency Deposit Rate Missing.
Price Index CPI, Datastream.

28. United Kingdom (UNK)
Local Currency Deposit Rate 90 day T-bill rate.

29. United States (USA)
Local Currency Deposit Rate 3 month T-bill rate.

E Additional international results
Table 14: Country fixed effects panel regressions of one year stock returns on top income shares

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Dependent Variable: t to t + 1 Stock Market Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1% (t)</td>
<td>-1.45** (0.67) -1.05* (0.56) -5.42** (1.47) -5.53* (2.45)</td>
</tr>
<tr>
<td>Time Trend</td>
<td>No No No No</td>
</tr>
<tr>
<td>Country FE</td>
<td>Yes Yes Yes Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>790 699 106 91</td>
</tr>
<tr>
<td>$R^2$ (w,b)</td>
<td>(.01,.11) (.00,.06) (.05,.06) (.04,.18)</td>
</tr>
</tbody>
</table>

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).