

The Equity Premium and the One Percent*

Alexis Akira Toda^{†1} and Kieran James Walsh^{‡2}

¹Department of Economics, University of California, San Diego

²Darden School of Business, University of Virginia

First draft: March 2014
This version: May 24, 2018

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, inequality should predict subsequent excess stock market returns. Consistent with our theory, we find that when the income share of top earners in the U.S. rises, subsequent one year excess market returns significantly decline. This negative relation is robust to (i) controlling for classic return predictors such as the price-dividend and consumption-wealth ratios, (ii) predicting out-of-sample, and (iii) instrumenting with changes in estate tax rates. Cross-country panel regressions suggest that the inverse relation between domestic inequality and returns also holds outside of the U.S., with stronger results in relatively closed economies than in ones with low home bias (in which U.S. inequality predicts returns).

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

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

*We benefited from comments by Daniel Andrei, Lint Barrage, Brendan Beare, Dan Cao, Vasco Carvalho, Peter Debaere, Graham Elliott, Nicolae Gârleanu, John Geanakoplos, Émilien Gouin-Bonenfant, Jim Hamilton, Gordon Hanson, Fumio Hayashi, Toshiki Honda, George Korniotis, Jiasun Li, Sydney Ludvigson, Semyon Malamud, Larry Schmidt, Allan Timmermann, Frank Warnock, Amir Yaron, 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, Washington State, Yale, Yokohama National University, 2014 Australasian Finance and Banking Conference, 2014 Northern Finance Association Conference, 2015 Econometric Society World Congress, 2015 ICMAIF, 2015 Midwest Macro, 2015 SED, 2015 UVa-Richmond Fed Jamboree, 2017 AFA, and 2017 ICEF. We especially thank Snehal Banerjee, Daniel Greenwald, Stavros Panageas, and Jessica Wachter for detailed comments. Earlier drafts of this paper were circulated with the title “Asset Pricing and the One Percent.”

[†]Email: atoda@ucsd.edu.

[‡]Email: walshk@darden.virginia.edu.

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 in prices and production. Central to his theory of fluctuations were differences in preferences and wealth across people.

To see the intuition as to why the wealth distribution affects asset pricing, consider an economy consisting of investors with different attitudes towards risk or beliefs about future dividends. In this economy, equilibrium risk premia 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 premia must fall to counterbalance the new demand of these agents. In this paper, we establish both the theoretical and empirical links between inequality and asset prices.

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

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

For a number of reasons including the apparent high persistence of top income and wealth shares, we employ a stationary component of inequality, “KGR” (capital gains ratio), which we define to be the difference between the top 1% income share with and without realized capital gains income, divided by the bottom 99% income share. We use data and theory to argue that KGR is a reasonable proxy for capital wealth and income inequality. Regressions of the year t to $t + 1$ excess return on the year t top 1% income share indicate a

strong and significant negative correlation: when KGR rises by one percentage point, subsequent one year excess market returns decline on average by about 2–4%, depending on the controls included (Table 2). Overall, our evidence suggests that the top 1% income share is not simply a proxy for the price level, which previous research shows correlates with subsequent returns, or for aggregate consumption factors: the top 1% income share predicts excess returns even after we control for some classic return predictors such as the price-dividend ratio (Fama and French, 1988) and the consumption-wealth ratio (Lettau and Ludvigson, 2001). Our findings are also robust to the inclusion of macro control variables, such as GDP growth. Using five year excess returns or the top 0.1% or 10% income share also yields similar results, although the predictability is really due to the top 1% (Table 4).

The empirical literature on return prediction is not without controversy. While many papers find evidence for return predictability,¹ others are more skeptical (Ang and Bekaert, 2007), and some point out econometric issues such as small sample bias when regressors are persistent (Nelson and Kim, 1993; Stambaugh, 1999) and problems with overlapping data (Valkanov, 2003; Boudoukh et al., 2008). In an influential study, Welch and Goyal (2008) show that excess return predictors suggested in the literature by and large perform poorly out-of-sample. How does the top 1% share fare out-of-sample? Using the methodologies of McCracken (2007) and Hansen and Timmermann (2015), we show that including the top 1% as a predictor significantly decreases out-of-sample forecast errors relative to using the historical mean excess return (Table 5). That is, top income shares predict returns out-of-sample as well.

Our finding that inequality predicts future returns is consistent with our theory, robust to the inclusion of controls and the construction of KGR, and thus interesting in its own right. But, does more inequality actually *cause* lower returns, and is KGR actually reflecting inequality? To address causality, we use tax rate changes as an instrument. Since contemporaneous and lagged changes in top estate tax rates explain a substantial portion of the variation in KGR (Table 6), we estimate the effect of inequality on returns using generalized method of moments (GMM) with instrumental variables (Table 7). Including KGR, industrial production growth, and the log price-earnings ratio as endogenous explanatory variables and using lags of top estate tax rate changes and the log price-earnings ratio as instruments, top income shares are still significant in predicting excess returns. This finding addresses another concern, which is that part of the variation in KGR is not due to inequality but rather from the timing of realizing capital gains or other omitted variables. Including one-year-ahead changes in capital gains tax rates as an additional instrument, we separately identify how the timing and inequality components predict returns (Table 9). The coefficient on the inequality component is negative and significant, while the timing coefficient is insignificant.

Further suggestive evidence that KGR reflects inequality comes from the

¹Classic examples are the price-dividend ratio (Campbell and Shiller, 1988; Fama and French, 1988; Hodrick, 1992; Cochrane, 2008) and the consumption-wealth ratio (Lettau and Ludvigson, 2001). Campbell and Thomson (2008) suggest that many economic variables predict returns by imposing weak restrictions such as a nonnegative equity premium. Rapach et al. (2010) show that instead of using a single predictive regression model, combining forecasts significantly decreases the out-of-sample forecast errors. See Lettau and Ludvigson (2010) and Rapach and Zhou (2013) for reviews on forecasting stock returns.

portfolios of the rich. Using the wealth composition estimates from Saez and Zucman (2016), we show that rising KGR is associated with subsequent increases in the share of equities owned by the rich (Table 10). The share of bonds, however, does not significantly change, which is consistent with our theory that rising inequality increases relative demand for risky assets.

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 by one percentage point, subsequent one year market returns significantly decline on average by 1%. However, this effect is not uniform across countries. Our theory suggests that for relatively “closed” economies such as emerging markets with high levels of investment home bias, the domestic top 1% share should matter for asset pricing because domestic agents account for a substantial proportion of the universe of investors. However, for small open economies with low home bias, the inequality amongst global investors (proxied by U.S. KGR) should matter because domestic agents comprise only a small fraction of investors. Consistent with our theory, we find that the interaction terms between top income shares and home bias measures significantly predict stock returns. In an economy with complete home bias, a one percentage point increase in the top 1% income share is associated with a subsequent 2.8% decline in stock market returns. In a small open economy (no home bias), a one percentage point increase in U.S. KGR is associated with a subsequent decline in stock market returns of 4.7%.

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. The original capital asset pricing model (CAPM) actually allowed for substantial heterogeneity in endowments and risk preferences across investors. (See Sharpe (1964), Lintner (1965a,b), and Geanakoplos and Shubik (1990) for a general and rigorous treatment.) However, their form of quadratic or mean-variance preferences admitted aggregation and obviated the role of the wealth distribution. Largely inspired by the limited empirical fit of the CAPM and asset pricing puzzles that arise in representative-agent models, since the 1980s theorists have extended macro/finance models to consider meaningful investor heterogeneity. Such heterogeneous-agent models fall into two groups.

In the first group, agents have identical standard (constant relative risk aversion) preferences but are subject to uninsured idiosyncratic risks.² Although the models of this literature have had some success in explaining returns in calibrations, the empirical results (based on consumption panel data) are mixed and may even be spuriously caused by the heavy tails in the cross-sectional consumption distribution (Toda and Walsh, 2015, 2017b). In the second group, markets are complete and agents have either heterogeneous 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”

²Examples are Mankiw (1986), Constantinides and Duffie (1996), Heaton and Lucas (1996), Brav et al. (2002), Cogley (2002), Balduzzi and Yao (2007), Storesletten et al. (2007), Kocherlakota and Pistaferri (2009), and Constantinides and Ghosh (2017), among many others. See Ludvigson (2013) for a review.

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.

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 and Hatchondo (2008) also calibrate the model and find that the effect of wealth inequality on the equity premium is small.

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 share of the less risk averse agent increases, then the risk-free rate goes up and the equity premium goes down. Although this prediction is similar to ours, he imposes an assumption on endogenous variables (see his equation (8)).

Following Dumas (1989), a large theoretical literature has studied the asset pricing implication of preference heterogeneity under complete markets.³ 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 first welfare theorem, for each equilibrium we can find Pareto weights such that the consumption allocation is the solution to the planner's problem, since in general the Pareto weights depend on the initial wealth distribution, changing the wealth distribution will change the Pareto weights, and consequently the asset prices. But in general it is hard to predict how the Pareto weights change. Second, even if we can predict how the Pareto weights change, there is the possibility of multiple equilibria. In such cases the comparative statics often go in the opposite direction depending on the choice of the equilibrium. Thus our results are quite different since we prove the uniqueness of equilibrium and derive comparative statics with respect to the initial wealth distribution.

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 they also solve the model without appealing to a planner's 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 different since we prove more general comparative statics results (though in two period models).

Our model is also related to the work on limited asset market participation, such as Basak and Cuoco (1998), Guvenen (2009), Chien et al. (2011, 2012),

³Examples are Wang (1996), Chan and Kogan (2002), Hara et al. (2007), Cvitanic et al. (2012), Longstaff and Wang (2012), Bhamra and Uppal (2014), and the references therein.

and Chabakauri (2013, 2015). In these papers some agents do not participate in certain asset markets or face portfolio constraints, which affects the asset prices beyond the heterogeneity in preferences or beliefs. In our model we also have hand-to-mouth laborers, but since they do not participate in *any* asset market, we prove that their presence affects only the risk-free rate and not the equity premium (Theorem 2.2).

Although the wealth distribution theoretically affects asset prices, there are few empirical papers that directly document this connection. To the best of our knowledge, Johnson (2012) and Campbell et al. (2016) are the only ones that explore this issue. Using incomplete markets models, they show that top income shares or top income growth innovations are cross-sectional asset pricing factors. However, they do not explore the ability of top income shares to predict excess market returns (our main empirical result).⁴

Lastly, our study is related to the findings of Greenwald et al. (2016), who identify innovations to wealth ($e_{a,t}$) that explain much of the variation in the stock market and significantly predict low subsequent excess returns. In an equilibrium model, they show that $e_{a,t}$ captures the risk tolerance of a representative stockholder. Interestingly, there is substantial correlation between $e_{a,t}$ and our inequality predictor variable KGR.⁵ Since in heterogeneous risk aversion models without aggregation, rising wealth concentration can effectively decrease the risk aversion of the corresponding representative stockholder/planner, an interpretation of $e_{a,t}$ is that it reflects the wealth share of relatively risk tolerant stockholders vs. more risk averse ones.

2 Wealth distribution and equity premium

In this section we present a theoretical model in which the wealth distribution across heterogeneous agents affects the equity premium. In Section 2.1, we consider a static model with agents that have heterogeneous but homothetic preferences and prove the uniqueness of equilibrium. In Section 2.2, we prove in a two period model that shifting wealth from a bondholder to a stockholder pushes down the equity premium. Appendix A contains all proofs. Appendix B analyzes an infinite horizon extension of this model.

2.1 Uniqueness of equilibrium

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

⁴Campbell et al. (2016) do explore market return prediction in their online appendix, but they uncover no relationship between the income of the rich and subsequent stock returns. Our findings are different likely because they use income instead of the income share and since they detrend top income linearly.

⁵We thank Daniel Greenwald for discovering this.

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

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

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

We make the following assumptions.

Assumption 1 (Homothetic, convex preferences). *For all i , $U_i : \mathbb{R}_+^S \rightarrow \mathbb{R}$ is continuous, strictly quasi-concave, homogeneous of degree 1, differentiable on \mathbb{R}_{++}^S , and $\nabla U_i(x) \gg 0$ with the Inada condition $\partial U_i(x)/\partial x_s \rightarrow \infty$ as $x_s \rightarrow 0$.*

Assumption 1 is standard in applied works. For example, the following constant relative risk aversion (CRRA) utility satisfies this assumption:

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

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

Assumption 2 (Tradability of endowments). *Agents' endowments are tradable: for all i , e_i is spanned by the column vectors of A .*

Assumption 2 holds, for example, if markets are complete (A is the identity matrix), but it does not need to be so. Under this assumption, since there exists $y_i \in \mathbb{R}^J$ such that $e_i = Ay_i$, by redefining e_i to be zero and n_i to be $n_i + y_i$, without loss of generality we may assume $e_i = 0$, i.e., agents are endowed only with assets.

Assumption 3 (Collinear endowments). *Agents have collinear endowments: letting $n = \sum_{i=1}^I n_i$ be the aggregate endowment of assets, we have $n_i = w_i n$, where $w_i > 0$ is the wealth share of agent i , so $\sum_{i=1}^I w_i = 1$. Furthermore, $An \gg 0$.*

Since $e_i = 0$ by assumption, the aggregate endowment of goods is An . Hence the assumption $An \gg 0$ simply says that aggregate endowment is positive. While the collinearity assumption is strong, it is indispensable in order to guarantee the uniqueness of equilibrium: Mantel (1976) shows that if we drop collinear endowments, then even with homothetic preferences "anything goes" for the aggregate excess demand function, and hence there may be multiple equilibria.⁶ With multiple equilibria, comparative statics may go in opposite directions, depending on the choice of equilibrium.

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

⁶See Toda and Walsh (2017a) for concrete examples of multiple equilibria with canonical two-agent, two-state economies.

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

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

Letting

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

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

Chipman (1974) shows that under complete markets, heterogeneous homothetic preferences, and collinear endowments, aggregation is possible and hence the equilibrium is unique. Our Theorem 2.1 is a stronger result since we prove the same for incomplete markets and we also obtain a complete characterization of the equilibrium portfolio as a solution to a planner’s problem. 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.⁷

2.2 Comparative statics

So far we assumed that there is no consumption at $t = 0$, but we can obtain similar results to Theorem 2.1 with consumption at $t = 0$ by interpreting $t = 0$ as a new “state” denoted by $s = 0$. Furthermore, we can introduce an agent type that does not participate in the asset market. Let e_s ($s = 0, 1, \dots, S$) be the aggregate endowment in state s . Suppose that there is a hand-to-mouth agent $i = 0$ (whom we call the laborer) that is endowed with goods but does not trade assets. Let $1 - \alpha_t$ be the fraction of aggregate income earned by the laborer at time t . Then the endowment of agent $i \geq 1$ (whom we call the capitalist) in state s is $\alpha_t w_i e_s$, where $t = 0$ if $s = 0$ and $t = 1$ if $s \geq 1$. In order to derive implications for asset pricing, we specialize the economy in Section 2.1 as follows.

Assumption 4 (Epstein-Zin with unit EIS). *Agent i ’s utility function is Epstein-Zin with unit elasticity of intertemporal substitution (EIS):*

$$\log U_i(x) = (1 - \beta_i) \log x_0 + \frac{\beta_i}{1 - \gamma_i} \log \left(\sum_{s=1}^S \pi_{is} x_s^{1-\gamma_i} \right), \quad (2.3)$$

where x_0 is consumption at $t = 0$, x_s is consumption in state s at $t = 1$, $\beta_i \in (0, 1)$ is the discount factor, $\gamma_i > 0$ is RRA (the case $\gamma_i = 1$ corresponds to log utility as usual), and $\pi_{is} > 0$ is agent i ’s subjective probability of state s .

⁷See Kehoe (1998) and Geanakoplos and Walsh (2018) for further discussion of uniqueness in the presence of heterogeneous preferences.

In terms of the financial structure, suppose that there are only two financial assets, a stock (a claim to the aggregate endowment) and a risk-free bond. By interpreting $t = 0$ as state $s = 0$, there are effectively three assets, the other being a claim to $t = 0$ consumption. Therefore the asset structure is as follows.

Assumption 5. Let $e_s > 0$ ($s = 0, 1, \dots, S$) be the aggregate endowment of goods in state s . The asset payoff matrix and the aggregate endowment of capitalists' assets are given by

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e_1 & 1 \\ \vdots & \vdots & \vdots \\ 0 & e_S & 1 \end{bmatrix}, \quad n = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} \alpha_0 e_0 \\ \alpha_1 \\ 0 \end{bmatrix}.$$

Under these assumptions, we can show that a redistribution of wealth from a bondholder to a stockholder reduces the equity premium, while redistribution between laborers and capitalists does not affect the equity premium. To make the statement precise, we introduce some additional notation.

Taking $t = 0$ consumption as the numéraire, let P be the ex-dividend price of the stock (the value of the claim $(0, e_1, \dots, e_S)'$) and R_f the gross risk-free rate (reciprocal of the value of the claim $(0, 1, \dots, 1)'$). Since the risk-free asset is in zero net supply, the aggregate wealth of capitalists at $t = 0$ is $\alpha_0 e_0 + \alpha_1 P$. Since the wealth share of capitalist i is w_i , the budget constraint is

$$y_{i1} + P y_{i2} + \frac{1}{R_f} y_{i3} = w_i (\alpha_0 e_0 + \alpha_1 P). \quad (2.4)$$

Let

$$\phi_i = \begin{bmatrix} \phi_{i1} \\ \phi_{i2} \\ \phi_{i3} \end{bmatrix} = \frac{1}{w_i (\alpha_0 e_0 + \alpha_1 P)} \begin{bmatrix} y_{i1} \\ P y_{i2} \\ y_{i3} / R_f \end{bmatrix}$$

be the vector of capitalist i 's portfolio shares.

Now we can state our main theoretical result.

Theorem 2.2. Under Assumptions 3–5, the followings are true.

1. There exists a unique equilibrium. Letting $y_i = (y_{i1}, y_{i2}, y_{i3})'$ be capitalist i 's equilibrium asset holdings, we have

$$y_{i1} = \frac{w_i (1 - \beta_i)}{\sum_{i=1}^I w_i (1 - \beta_i)} \alpha_0 e_0, \quad (2.5)$$

and $(y_{i2}, y_{i3})_{i=1}^I$ solves

$$\begin{aligned} & \text{maximize} && \sum_{i=1}^I \frac{w_i \beta_i}{1 - \gamma_i} \log \left(\sum_{s=1}^S \pi_{is} (e_s y_{i2} + y_{i3})^{1 - \gamma_i} \right) \\ & \text{subject to} && \sum_{i=1}^I y_{i2} = \alpha_1, \quad \sum_{i=1}^I y_{i3} = 0. \end{aligned} \quad (2.6)$$

The equilibrium consumption allocation is given by $x_{i0} = y_{i1}$ and $x_{is} = e_s y_{i2} + y_{i3}$ for $s = 1, \dots, S$.

2. The equilibrium price-dividend ratio is given by

$$\frac{P}{e_0} = \frac{\alpha_0}{\alpha_1} \frac{\sum_{i=1}^I w_i \beta_i}{1 - \sum_{i=1}^I w_i \beta_i}. \quad (2.7)$$

Consequently, shifting wealth from an impatient agent (low β_i) to a patient agent (high β_i) increases the price-dividend ratio.

3. Let $R = (e_1/P, \dots, e_S/P)'$ be the vector of gross stock returns, $\pi = (\pi_1, \dots, \pi_S)'$ be any probability, and $\mu = \pi'(\log R - \log R_f)$ be the log equity premium. Then μ is independent of the capitalists' income shares α_0, α_1 . Shifting wealth from an agent who invests relatively more in the risk-free asset (high ϕ_{i3}) to an agent who invests relatively less (low ϕ_{i3}) reduces the log equity premium.

The intuition for Theorem 2.2 is as follows. In an economy with financial assets, the equilibrium prices and risk premia balance the agents' preferences and beliefs. Since the stock is the only saving vehicle in the aggregate (because the risk-free asset is in zero net supply), shifting wealth to a patient agent increases the demand for stocks, and hence its price rises. If wealth shifts into the hands of the natural stockholder (either the risk tolerant or optimistic agent), everything else fixed, the aggregate demand for the stock increases. Hence for markets to clear, the risk premium must fall to counterbalance the new demand of these agents.

A surprising aspect of Theorem 2.2 is that the equity premium is independent of the capitalist/laborer income shares α_0, α_1 and depends only on the wealth distribution among capitalists, $\{w_i\}_{i=1}^I$. The intuition is that while α_0, α_1 affect the overall level of asset prices and the risk-free rate by changing the relative income between $t = 0, 1$, they do not affect the equity premium because assets are held only by capitalists; the equity premium balances the relative demand of stocks and bonds, which depends only on the capitalists' wealth distribution.

Who is the natural bondholder in Theorem 2.2? We can answer this question by reducing the individual problem to a static optimal portfolio problem. Due to unit EIS, we have $y_{i1} = (1 - \beta_i)w_i(\alpha_0 e_0 + \alpha_1 P)$. Therefore the budget constraint (2.4) simplifies to

$$P y_{i2} + \frac{1}{R_f} y_{i3} = w_i \beta_i (\alpha_0 e_0 + \alpha_1 P).$$

Let $\theta_i = \frac{\phi_{i2}}{\phi_{i2} + \phi_{i3}} = \frac{P y_{i2}}{w_i \beta_i (\alpha_0 e_0 + \alpha_1 P)}$ be capitalist i 's portfolio share of stocks within savings. With a slight abuse of notation, let $e_1 = (e_1, \dots, e_S)' \in \mathbb{R}_{++}^S$ be the vector of aggregate endowments at $t = 1$. Then the vector of gross returns is $R = e_1/P$. Since agents have Epstein-Zin utility, the consumption-saving decision and the portfolio decision can be separated. By homotheticity, the optimal portfolio problem reduces to

$$\max_{\theta} E_i [u_i(R\theta + R_f(1 - \theta))], \quad (2.8)$$

where u_i is the Bernoulli utility with relative risk aversion γ_i (i.e., $u_i(x) = \frac{x^{1-\gamma_i}}{1-\gamma_i}$ if $\gamma_i \neq 1$ and $u_i(x) = \log x$ if $\gamma_i = 1$) and E_i is the expectation under agent i 's belief.

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

Proposition 2.3. *Suppose Assumptions 3–5 hold and agents have common beliefs. If $\gamma_1 > \dots > \gamma_I$, then $0 < \theta_1 < \dots < \theta_I$.*

Proposition 2.4. *Suppose Assumptions 3–5 hold and agents 1, 2 have common risk aversion. Assume that agent 1 is more pessimistic than agent 2 in the above sense. Then $\theta_1 < \theta_2$.*

Combining Theorem 2.2 together with either Proposition 2.3 or 2.4, provided that two agents have the same discount factor (hence the same $\phi_{i1} = 1 - \beta_i$), shifting wealth from a more risk averse or pessimistic agent to a more risk tolerant or optimistic agent reduces the equity premium. In particular, if the rich are relatively more risk tolerant, optimistic, or simply more likely to buy risky assets (for example due to fixed stock market participation costs), rising inequality should forecast declining excess returns.

Theorem 2.2 tells us that the wealth distribution qualitatively affects asset prices, but does it matter quantitatively? To address this issue we compute a numerical example calibrated at annual frequency. For simplicity we ignore the laborers, so $\alpha_0 = \alpha_1 = 1$. We specialize the above economy to one with two agents denoted by $i = A, B$. For preference parameters, let $\rho_i > 0$ be the discount rate of agent i and define the discount factor by $\beta_i = \frac{e^{-\rho_i}}{1+e^{-\rho_i}}$. We set $(\rho_A, \rho_B) = (0.015, 0.06)$ and $(\gamma_A, \gamma_B) = (1, 5)$, so we can interpret type A as the “rich” (patient, risk tolerant) and type B as the “poor” (impatient, risk averse). These preference parameters as well as the dividend growth distribution are taken from the infinite horizon calibration in Appendix B. Given these parameters, we can easily solve for the equilibrium by numerically solving the planner’s problem (2.6). Figure 1 shows the results.

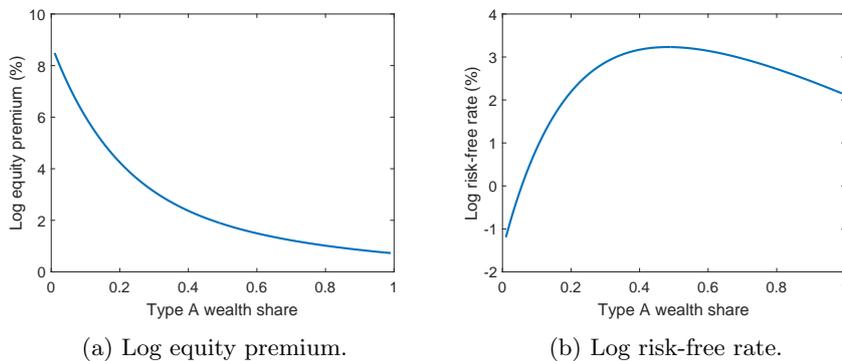


Figure 1: Wealth distribution and asset prices.

Figures 1a and 1b show the log equity premium and the log risk-free rate,

respectively. Consistent with Theorem 2.2, increasing the wealth share of type A agents monotonically decreases the equity premium. Since the equity premium ranges between about 8% and 1%, the wealth distribution has a quantitatively large effect.

3 Predictability of returns with inequality

In Theorem 2.2, we have theoretically shown that shifting wealth from an agent who holds comparatively fewer stocks to one who holds more reduces the subsequent equity premium. Many empirical papers show that the rich hold relatively more stocks than do the poor and argue that the rich are relatively more risk tolerant.⁸ Therefore, rising inequality should negatively predict subsequent excess stock market returns. In this section we construct a stationary measure of inequality and show that it predicts subsequent returns.⁹

3.1 Connecting theory to empirics

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

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

and test whether $\beta = 0$. There are several obstacles for implementing this type of regression. First, it is difficult to measure wealth, and hence of wealth inequality. The 1916–2000 top wealth share series (based on estate tax data) from Kopczuk and Saez (2004) are missing many years in the 50s, 60s, and 70s. The wealth share data of Saez and Zucman (2016) cover 1913–2012 but are *imputed* from capitalizing income. Second, our model considers redistribution amongst agents that participate in the stock market. Therefore we should look at the inequality amongst stockholders, but the above wealth inequality measures consider all agents. Third, these wealth measures are highly persistent, which introduces econometric problems. Fortunately, there is a way to construct a stationary proxy measure of capitalist inequality, which we describe below.¹⁰

We employ the Piketty and Saez (2003) income inequality measures for the U.S., which are available in updated form in a spreadsheet on Emmanuel Saez’s website.¹¹ In particular, we consider top income share measures based on tax return data, which are at the annual frequency and cover the period 1913–2015. These series reflect in a given year the percent of income earned by the top 1% of earners pretax. We also employ the top 0.1% share, the top 10% share, and the corresponding series that exclude realized capital gains income. Figure 2 shows these series, both including realized capital gains (Figure 2a) and excluding capital gains (Figure 2b). We can immediately see that all series seem to share a common U-shaped trend over the century, and the series including capital gains are more volatile than those without capital gains.

⁸See Haliassos and Bertaut (1995), Carroll (2002), Vissing-Jørgensen (2002), Campbell (2006), Wachter and Yogo (2010), Buccioli and Miniaci (2011), and Calvet and Sodini (2014).

⁹In this section we are only concerned with predictability, or correlation. We address causality in Section 4.

¹⁰Using our calibrated infinite horizon model, in Appendix B.5 we show that capital wealth inequality, income inequality, and our proxy are in theory all highly correlated.

¹¹<https://eml.berkeley.edu/~saez/>

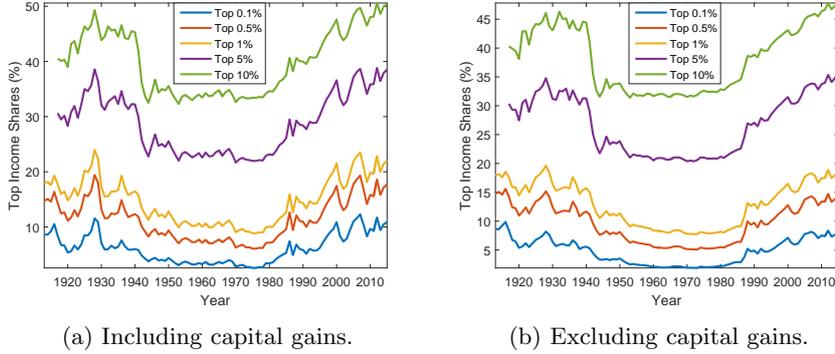


Figure 2: U.S. top income shares (1913–2015).

Let $\text{top}(x)$ ($\text{top}(x)^{\text{excg}}$) be the top $x\%$ income share including (excluding) capital gains. The fact that these two series seem to share a common trend motivates us to consider their difference

$$\text{top}(x) - \text{top}(x)^{\text{excg}}. \quad (3.1)$$

Using a Taylor approximation, we can connect this quantity to other measures of inequality, as described below.

Suppose that there are two agent types in the economy, say the top 1% and the rest (bottom 99%). Let us denote these two types by $i = A, B$. Let Y_i^k, Y_i^l be the total capital and labor income of type i and $Y_i = Y_i^k + Y_i^l$ be the total income of type i . Let $Y^k = Y_A^k + Y_B^k$ and $Y^l = Y_A^l + Y_B^l$ be the aggregate capital and labor income and $Y = Y^k + Y^l = Y_A + Y_B$ be the total aggregate income. Then the top income share (type A 's income share) is

$$\text{top}(1) = \frac{Y_A}{Y} = \frac{Y_A^k + Y_A^l}{Y_A^k + Y_A^l + Y_B^k + Y_B^l}.$$

Suppose that fraction ρ_i of type i 's capital income is comprised of realized capital gains. Then the top income share excluding realized capital gains is

$$\text{top}(1)^{\text{excg}} = \frac{(1 - \rho_A)Y_A^k + Y_A^l}{(1 - \rho_A)Y_A^k + Y_A^l + (1 - \rho_B)Y_B^k + Y_B^l} =: f(\rho_A, \rho_B).$$

Using Taylor's theorem, we can approximate the quantity in (3.1) as

$$\begin{aligned} \text{top}(1) - \text{top}(1)^{\text{excg}} &= f(0, 0) - f(\rho_A, \rho_B) \\ &\approx -\rho_A \frac{\partial f}{\partial \rho_A}(0, 0) - \rho_B \frac{\partial f}{\partial \rho_B}(0, 0) \\ &= \rho_A \frac{Y_A^k Y_B}{Y^2} - \rho_B \frac{Y_B^k Y_A}{Y^2}. \end{aligned} \quad (3.2)$$

Letting α be the capital income share in aggregate income, so $Y^k = \alpha Y$, it follows from (3.2) that

$$\text{top}(1) - \text{top}(1)^{\text{excg}} \approx \alpha \left(\rho_A \frac{Y_A^k}{Y^k} (1 - \text{top}(1)) - \rho_B \left(1 - \frac{Y_A^k}{Y^k} \right) \text{top}(1) \right). \quad (3.3)$$

Since Y_A^k/Y^k is the capital income share of the top 1%, who are more likely to be capital owners or entrepreneurs, it is reasonable to assume that the order of magnitude of Y_A^k/Y^k is at least that of $1 - Y_A^k/Y^k$. According to Figure 2a, the top 1% income share has evolved between 0.1 and 0.2, so $1 - \text{top}(1) \gg \text{top}(1)$. Therefore assuming that ρ_B is at most of the same order of magnitude as ρ_A , the second term inside the parenthesis of the right-hand side of (3.3) is much smaller than the first term. Ignoring the second term, we obtain

$$\begin{aligned} \text{top}(1) - \text{top}(1)^{\text{excg}} &\approx \alpha \rho_A \frac{Y_A^k}{Y^k} (1 - \text{top}(1)) \\ \iff \text{KGR}(1) &:= \frac{\text{top}(1) - \text{top}(1)^{\text{excg}}}{1 - \text{top}(1)} \approx \alpha \rho_A \frac{Y_A^k}{Y^k}. \end{aligned} \quad (3.4)$$

The left-hand side of (3.4), $\text{KGR}(1)$ (we will explain the acronym shortly), is a quantity that can be calculated from top 1% income shares including/excluding realized capital gains. (3.4) says that it has three components: $\alpha = Y^k/Y$ (the capital share of aggregate income), ρ_A (the fraction of realized capital gains income to total capital income for top earners), and Y_A^k/Y^k (the capital income share of top earners).

Before proceeding further, we need to make sure that the approximation (3.4) is empirically accurate. We address this issue in two ways. First, taking the logarithm of (3.4), we obtain

$$\log(\text{KGR}(1)) \approx \log \alpha + \log \rho_A + \log(Y_A^k/Y^k). \quad (3.5)$$

It is possible to approximate all quantities in the right-hand side of (3.5) from the data of Saez and Zucman (2016), at least for the period 1916–2012.¹² To evaluate the accuracy of the approximation (3.4), inspired by (3.5), in columns (1)–(3) of Table 1 we regress $\log(\text{KGR})$ on the logarithm of the three components for the top 0.1%, 1%, and 10% group. In each case R^2 is above 0.9, which suggests that the three components α , ρ_A , and Y_A^k/Y^k explain almost all of the variation in KGR. Furthermore, consistent with (3.5), with top 0.1% and 1% the constant term is insignificant and the other three coefficients are statistically not different from 1. This result suggests that the approximation (3.4) is indeed accurate. Second, we construct the two terms inside the parenthesis in (3.3). For the top 0.1% and 1%, the sample means for term 1 are 0.279 and 0.154, respectively, which are much larger than the term 2 means of 0.006 and 0.009. Additionally, the term 1 standard deviations are 0.165 and 0.091, respectively, while the term 2 standard deviations are 0.005 and 0.006. Therefore, for the top 0.1% and 1%, term 2 is indeed negligible relative to term 1, consistent with the accuracy of the approximation.¹³

Intuitively, what does $\text{KGR}(1)$ measure? The first component, $\alpha = Y^k/Y$, is the capital share of income, although we find that most variation in $\text{KGR}(1)$ is

¹²Our sources are the “AppendixTables(Aggregates)” and “AppendixTables(Distributions)” spreadsheets for Saez and Zucman (2016). We compute α as the sum of total positive business income, taxable interest, dividends, positive rents, estate and trust income, and net realized capital gains minus business and rental losses divided by total net taxable income. Y_A^k/Y^k is the series “Top taxable capital income shares, capital gains included in shares & rankings.” We compute ρ_A as the fraction of realized capital gains in the income of the top 1% divided by the fraction of capital income in the 1%’s income.

¹³For the top 10%, term 1 is larger and more volatile than is term 2, but the comparison is less stark: the term 1 and term 2 means are (0.053, 0.009), and the standard deviations are (0.028, 0.006).

Table 1: Decomposition of KGR

Regressors (t)	Dependent Variable: $\log(\text{KGR}(x))$					
	(1)	(2)	(3)	(4)	(5)	(6)
	0.1%	1%	10%	1%	1%	1%
Constant	-0.07 (0.34)	-0.05 (0.28)	1.17 (0.39)	-5.37 (1.33)	-2.68 (0.17)	-2.67 (0.44)
$\log \alpha$	1.15*** (0.20)	0.97*** (0.18)	1.65*** (0.26)	-0.82 (0.80)		
$\log \rho_x$	0.98*** (0.09)	1.12*** (0.10)	1.25*** (0.10)		1.00*** (0.11)	
$\log(Y_x^k/Y^k)$	0.97*** (0.08)	1.25*** (0.19)	3.47*** (0.45)			1.87*** (0.55)
Sample	1922- -2012	1916- -2012	1962- -2012	1916- -2012	1916- -2012	1916- -2012
R^2	0.94	0.91	0.95	0.04	0.78	0.14

Note: Newey-West standard errors in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). For $x = 0.1, 1, 10$, the table shows regressions of $\log(\text{KGR}(x))$ on its components according to (3.4): the logs of the capital income share (α), the realized capital gain share of capital income (ρ_x), and the top $x\%$'s share of capital income (Y_x^k/Y^k ; ranked by capital income including realized capital gains).

not due to α . The second component, ρ_A , could reflect two factors. First, since capital gains realization is more likely when prices are high and since the rich disproportionately hold capital, rising ρ_A should correlate with rising capital wealth inequality. Second, ρ_A could simply be capturing the timing of capital gains realization. We provide evidence *against* this second interpretation below (Section 4.1 and Table 10). The last component, Y_A^k/Y^k , is capital income inequality.

To see which of the three components mainly determines KGR, in columns (4)–(6) of Table 1, we regress $\log(\text{KGR}(1))$ on the components one at a time. The R^2 takes a large value 0.78 when we use only ρ_A , an intermediate value of 0.14 when we use Y_A^k/Y^k , and 0.00 when we use α (which is also insignificant). Therefore KGR is mostly reflecting variation in ρ_A , the fraction of realized capital gains income to total capital income for top earners. This is why we chose the name KGR: capital gains ratio. To some extent KGR moves with capital income inequality Y_A^k/Y^k . The capital share α does not appear to drive KGR. In summary, provided the timing component of ρ_A is not dominating (see Section 4.1), KGR measures capital wealth and income inequality.

An advantage of KGR is that it is stationary (the Phillips and Perron (1988) p-values are less than 0.01 for the top 0.1%, 1%, and 10%). In contrast, the raw top wealth and income series appear nonstationary, or at least highly persistent, and thus introduce econometric problems when used to predict stationary returns (Granger, 1981). KGR(1) actually looks very much like the detrended top 1% income share series. Figure 3 shows the KGR(1) series as well as the detrended versions of the raw top 1% series using the Kalman filter with an AR(1) cyclical component (see Appendix E for details) or subtracting the 10 year moving average.

In light of Figures 2a and 2b and the capitalist/laborer share irrelevance in

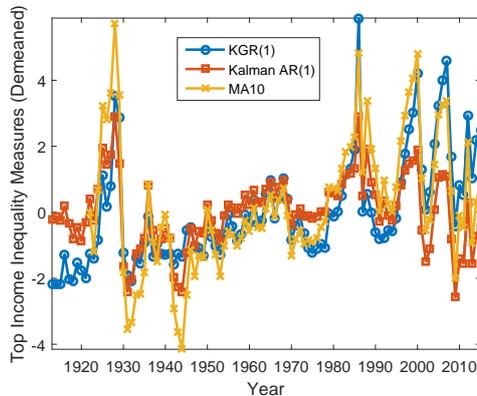


Figure 3: Time series plot of inequality measures.

Note: the graph shows the stationary component of the top 1% income share using KGR(1), the AR(1) Kalman filter (both demeaned), and subtracting the 10 year moving average. The units of all series are percentage points.

Theorem 2.2, there is arguably a non-econometric reason for detrending as well. Since in the figures both the top 0.1% and 10% (and the shares in between) appear to have a common U-shaped trend, it seems plausible that the slow-moving component of inequality is due to redistribution between the poor and rich uniformly rather than from intra-rich redistribution. Assuming the poor/non-rich are less likely to participate in financial markets, then the trends in inequality correspond to changes in the capitalist/laborer income shares (α_0, α_1) in Theorem 2.2, which are irrelevant for the equity premium according to our model. In this case, we would want to strip out the trends prior to predictive regressions. Intuitively, long-term trends in the capitalist/laborer income share should affect the overall level of asset markets but not the equity premium per se, which depends only on the intra-capitalist wealth distribution.

In the remainder of this section, we focus on the ability of KGR(1) defined by (3.4) to predict excess market returns in- and out-of-sample.

3.2 In-sample predictions

We obtain the U.S. stock market returns, risk-free rates, and other financial variables from the spreadsheet of Welch and Goyal (2008).¹⁴ Before 1926, stock returns are calculated from the S&P 500 index. After 1926, we use CRSP volume weighted average returns. We put returns into real terms using consumer price index (CPI) inflation, and when we say “returns”, it always means log returns. For example, if the gross return on stocks from year t to $t + 1$ is $R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$, the log return is $\log R_{t+1}$. Similarly, the excess return refers to the log excess return $\log R_{t+1} - \log R_{f,t}$. In some of the specifications below, we use five year annualized returns, which are compounded annually:

$$\log R_{t \rightarrow t+5} = \frac{1}{5} \sum_{k=1}^5 \log R_{t+k}.$$

¹⁴<http://www.hec.unil.ch/agoyal/>

The series P/D and P/E are the price-dividend and price-earnings ratios (in real terms) for the S&P 500 index. The spreadsheet also contains the Lettau and Ludvigson (2001) consumption-wealth ratio, commonly referred to as CAY, which spans the period 1945–2015. For presentation, we multiply CAY by 100. Our other controls are GDP growth and, inspired by Lettau et al. (2008) and Bansal et al. (2014), consumption growth variance. Annual data for GDP and consumption are from the website of the Federal Reserve Bank of St. Louis (FRED)¹⁵ and span 1930–2016. We estimate consumption growth variance using an AR(1)-GARCH(1,1) model for log consumption growth.

Table 2 shows the results of regressions of one year (t to $t + 1$) excess stock market returns on KGR(1) (time t), some classic return predictors (time t), and macro factors (time t). In column (1) we find that when KGR(1) rises by one percentage point in year t , subsequent one year excess market returns (January to December of year $t + 1$) decline on average by 2.7%. The coefficient is significant at the 1% level (using a Newey-West standard error), and the R^2 statistic is 0.051. It is clear, at least in sample, that KGR(1) forecasts the subsequent excess return on the stock market.

Table 2: Regressions of one year excess stock market returns on KGR(1) and other predictors

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	11.92 (2.74)	11.30 (4.06)	17.30 (8.07)	9.10 (16.82)	14.65 (10.84)	13.59 (3.63)
KGR(1)	-2.69*** (1.00)	-2.70** (1.25)	-3.38* (1.76)	-2.89* (1.54)	-2.56** (1.12)	-2.79** (1.37)
$\Delta \log(\text{GDP})$		0.36 (0.48)				
$\log(\text{CGV})$			-2.15 (2.97)			
$\log(\text{P/D})$				0.99 (5.66)		
$\log(\text{P/E})$					-1.12 (4.21)	
CAY						1.25* (0.76)
Sample	1913- -2015	1930- -2015	1930- -2015	1913- -2015	1913- -2015	1945- -2015
R^2	0.051	0.055	0.051	0.051	0.052	0.117

Note: Newey-West standard errors in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). KGR(1) is the proxy for top 1% capital inequality defined by (3.4). $\Delta \log(\text{GDP})$ is real GDP growth. CGV is consumption growth volatility, which is estimated 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.

Tables 20 and 21 (in Appendix D) show that the inverse relationship between inequality measures and subsequent excess returns also holds with the

¹⁵<http://research.stlouisfed.org/fred2/>

KGR(10) and KGR(0.1) series. Table 3 shows that all three versions of $KGR(x)$ also significantly predict five year excess returns. Figures 4a and 4b show the corresponding scatter and time series plots for five year returns. KGR(1) appears to forecast subsequent five year excess returns well except around 1986.¹⁶ Overall, a one percentage point increase in KGR(1) is associated with, roughly, a 2–4% decline in subsequent excess returns.

Table 3: Regressions of five year excess stock market returns on $KGR(x)$

Dependent Variable: t to $t + 5$ Excess Market Return			
KGR(x) version			
Regressors (t)	0.1%	1%	10%
Constant	10.20 (1.95)	10.61 (2.05)	12.25 (2.05)
KGR	-3.16** (1.25)	-2.19*** (0.84)	-2.06*** (0.65)
Sample	1913- -2015	1913- -2015	1917- -2015
R^2	0.173	0.181	0.219

Note: Newey-West standard errors in parentheses (8 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). Five year excess returns are annualized. $KGR(x)$ is the proxy for top $x\%$ capital inequality.

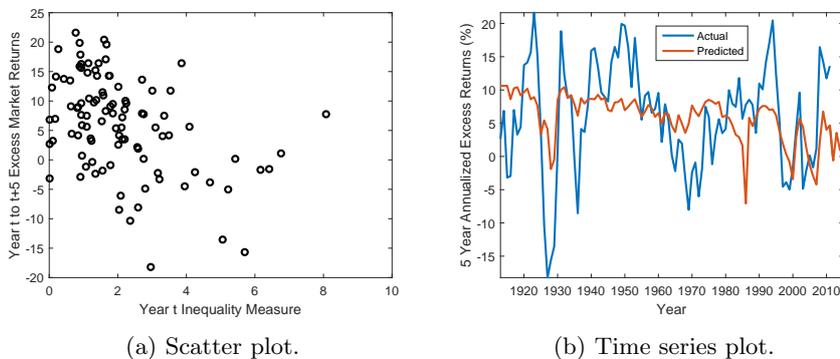


Figure 4: Year t to year $t + 5$ excess stock market return (annualized) vs. year t KGR(1) (proxy for top 1% capital inequality), 1913–2015.

Across specifications throughout the paper, KGR(10) performs relatively well in forecasting excess returns. This is surprising at first glance since the top 10% hold most financial wealth, and we argued in Section 3.1 that *intra*-capitalist inequality is what should matter for excess returns. There is, however, a straightforward explanation, which is that KGR(10) overlaps with KGR(1) and KGR(0.1) and contains the predictive power of higher income shares. In Table 4, we predict returns with *intra*-10% KGR analogs, for example, KGR(5–

¹⁶The capital gains tax rate increased from 20% in 1986 to 28% in 1987 (but announced in 1986), which gave investors an incentive to realize capital gains in 1986. In Section 4 we disentangle the inequality and timing components of KGR(1).

10) corresponding to the top 5–10% income share (see the table caption for a precise definition). These new variables form a decomposition of $KGR(10)$ since regressing it on $KGR(0.1), \dots, KGR(5-10)$ yields an R^2 of over 0.99 (with all regressors strongly significant). Consistent with our theory, it is the highest share series (top 0.1–0.5% and top 0.1%) with the largest coefficients (regressors are standardized in Table 4) and R^2 . The top 1–5% KGR is not significant, while $KGR(5-10)$ has a *positive* coefficient significant at the 10% level. So, while the $KGR(0.1)$ component of $KGR(10)$ inversely forecasts excess returns, the $KGR(5-10)$ component does the opposite, consistent with our story that redistribution to poorer capitalists *increases* excess returns.

Table 4: Regressions of one year excess stock market returns on KGR for finer income groups

Regressors (t) (Standardized)	Dependent Variable: t to $t + 1$ Excess Market Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	6.33 (1.88)	6.03 (1.82)	6.03 (1.86)	6.03 (1.85)	6.33 (1.83)	6.33 (1.82)
$KGR(10)$	-5.34*** (1.63)					
$KGR(0.1)$		-4.08** (1.70)				
$KGR(0.1-0.5)$			-4.41** (1.88)			
$KGR(0.5-1)$				-3.73** (1.81)		
$KGR(1-5)$					0.19 (2.14)	
$KGR(5-10)$						3.29* (1.73)
Sample	1917- -2015	1913- -2015	1913- -2015	1913- -2015	1917- -2015	1917- -2015
R^2	0.077	0.045	0.052	0.037	0.000	0.029

Note: see caption of Table 2. The other variables are the sub-income group components of KGR defined analogously. For example, $KGR(1-5) = (\text{top}(1-5) - \text{top}(1-5)^{\text{excg}}) \times \frac{1-\text{top}(1)}{1-\text{top}(5)}$, which is the correct analog since $KGR(5)$ is effectively $KGR(0-5)$, and $\text{top}(0) = 0$.

Given the strength of the relationship, a question immediately arises. Is there some mechanical, non-equilibrium explanation for the relationship between inequality and subsequent excess returns? For example, might stock returns somehow be *determining* the top share measures? For a few reasons, the answer is likely no. First, the relationship is between *initial* inequality and *subsequent* returns. Returns could affect contemporaneous top shares but not lagged top shares. One might still worry that our results are driven by our transformation of the top share series into $KGR(1)$. However, as we see in Appendix D, we get similar results with other methods of creating a stationary series.

But, one might say, we have known at least since Campbell and Shiller (1988)

and Fama and French (1988) that when prices are high relative to either earnings or dividends, subsequent excess market returns are low. The current price could indeed affect current inequality. Are the KGR series simply proxying for the price-dividend or price-earnings ratios, which are known to predict returns? Again, the answer seems to be no. As we see in columns (4) and (5) from Table 2, top shares predict excess returns even when controlling for the log price-dividend or price-earnings ratio. Including these controls barely affect the KGR(1) coefficient, which is large and significant. The P/D and P/E ratios, however, are not significant after controlling for top capital shares. Similar results hold for the KGR(10) series in Table 20. With respect to KGR(0.1) (Table 21), we lose significance when controlling for P/D due to large standard errors, but the top share coefficient remains large and significant when including P/E.

In columns (2), (3), and (6) from Table 2, we also control for real GDP growth, consumption growth variance (Lettau et al., 2008; Bansal et al., 2014), and CAY, which Lettau and Ludvigson (2001) show forecasts excess market returns. Including these controls (which also shortens the sample), we still see a strong relationship between the top income share and subsequent returns. Similar results hold for different percentiles of the top income share and detrending methods (Appendix D).

How do the components of $KGR(x)$ — α , ρ_x , and Y_x^k/Y^k —perform in forecasting excess returns? We see in Table 24 (Appendix D) that ρ_x , the primary driver of $KGR(x)$ according to Table 1, significantly predicts lower excess returns, while α and Y_x^k/Y^k are insignificant. Given that the realized capital gains component of KGR, ρ , is driving return prediction, a question arises. Are realized capital gains and not inequality per se predicting excess returns? We address this issue in Appendices B.5 and C. First, we calibrate and solve an infinite horizon version of our model and simulate model versions of both KGR and a purely mechanical measure of aggregate realized capital gains. Our Monte Carlo experiment shows that in our calibrated model (i) KGR, mechanical capital gains, and wealth inequality all inversely forecast excess returns, (ii) all three variables are substantially correlated, and (iii) KGR and mechanical capital gains have small sample properties better than those of wealth inequality. In short, both KGR and mechanical realized capital gains are quantitatively reasonable proxies for wealth inequality, which our model shows inversely forecasts excess returns.¹⁷ Second, we run a horse race between KGR in the data and empirical measures of mechanical realized capital gains. We find that they perform similarly but provide evidence that the distribution of realized capital gains across income groups matters beyond the overall level in forecasting returns.

In summary, there are three reasons we advocate using KGR as our proxy for capitalist wealth inequality in testing our theory. First, it is a reasonable proxy according to our calibrated model. Second, it is easily computed from frequently updated, publicly available data, and its computation requires no arbitrary parameters. Third, KGR is stationary and closely resembles detrended versions of income inequality, which has a very persistent component perhaps driven by forces we argue are less relevant for the equity premium. We elaborate

¹⁷Income inequality also inversely predicts excess returns, but it performs worse than the other variables.

on these points in Appendices B.5 and C.

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 (KGR in particular) to predict excess stock market returns out-of-sample.

Consider the predictive regression model for the equity premium,

$$y_{t+h} = \beta' x_t + \epsilon_{t+h}, \quad (3.6)$$

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

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

$$F = \frac{1}{\hat{\sigma}_\epsilon^2} \sum_{t=\lfloor \rho T \rfloor + 1}^T \left[(y_{t+h} - \hat{y}_{t+h|t}^N)^2 - (y_{t+h} - \hat{y}_{t+h|t}^A)^2 \right], \quad (3.7)$$

where $\hat{\sigma}_\epsilon^2$ is a consistent estimator of $\text{Var}[\epsilon_{t+h}]$ (which we estimate from the sample average of the squared OLS residuals of (3.6) using the whole sample), $\hat{y}_{t+h|t}^A = \hat{\beta}'_t x_t$ ($\hat{y}_{t+h|t}^N = \hat{\beta}'_{1t} x_t$) is the predicted value of y_{t+h} based on x_t using the ALT (NULL) model (here $\hat{\beta}_t, \hat{\beta}_{1t}$ are the OLS estimator of (3.6) using data only up to time t), T is the sample size, and $0 < \rho < 1$ is the proportion of observations set aside for initial estimation of β and β_1 . Theorems 3 and 4 of Hansen and Timmermann (2015) show that under the null ($H_0 : \beta_2 = 0$), the asymptotic distribution of F is a weighted sum of the difference of independent $\chi^2(1)$ variables.

For the regressors in the ALT model, following Welch and Goyal (2008), we consider the simplest possible case where $x_{1t} \equiv 1$ (constant) and x_{2t} consists of a single predictor. For the predictor x_{2t} , we consider $\text{KGR}(x)$ for $x = 0.1, 1, 10$ and valuation ratios ($\log(\text{P/D})$ and $\log(\text{P/E})$). 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 2015), so we cannot afford to use variables that are available only in shorter samples (e.g., CAY) for performing out-of-sample predictions, and (ii) since Welch and Goyal (2008) find that most predictor variables suggested in the literature are poor, there is no point in comparing many variables. The choice of the proportion of the training sample, ρ , is necessarily subjective. Small ρ leads to imprecise initial estimates of β , and large ρ leads to the loss of power. Hence we simply report results for $\rho = 0.2, 0.3, 0.4$. Table 5 shows the results.

Table 5: Out-of-sample performance in predicting 1-year excess returns

ρ	Predictor in the ALT Model				
	KGR(1)	KGR(10)	KGR(0.1)	log(P/D)	log(P/E)
0.2	3.67***	6.07***	2.67**	-0.12	0.77*
	(0.0040)	(0.0010)	(0.0131)	(0.1367)	(0.0515)
0.3	2.16**	3.19***	1.43**	0.23	1.34**
	(0.0153)	(0.0068)	(0.0436)	(0.1245)	(0.0360)
0.4	1.42**	2.94***	0.64*	-0.42	0.58*
	(0.0388)	(0.0081)	(0.0901)	(0.2781)	(0.0845)

Note: $\rho = 0.2, 0.3, 0.4$ is the proportion of observations set aside to compute an initial OLS estimate. columns correspond to the predictors included in the ALT model in addition to a constant. $KGR(x)$ is the proxy for top $x\%$ capital inequality defined by (3.4). The numbers in the table are the out-of-sample F statistic computed by (3.7). p-values (in parentheses) are computed by simulating 10,000 realizations from the asymptotic distribution based on Hansen and Timmermann (2015) (one sided). ***, **, and * indicate significance at 1%, 5%, and 10% levels.

According to Table 5, we can see that across specifications, the out-of-sample F statistic is positive and significant when we use $KGR(x)$, while it is insignificant for $\log(P/D)$ and weakly significant for $\log(P/E)$. (Note that since the asymptotic distribution of F depends on the NULL model, the relationship between 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 (3.7)) over the prediction period in Figure 5. The vertical axis is the cumulative sum for the NULL model minus the ALT, so a positive value favors the ALT. We can see that for all $KGR(x)$ specifications, the plots roughly monotonically increase up to 1980, decrease until 1990, and then increase again. This result is not surprising, since 1980s was a time when income inequality increased but the stock market did not suffer (Figure 2a). On the other hand, the $\log(P/D)$ and $\log(P/E)$ specifications deteriorate after 1970, especially so for $\log(P/D)$. This finding is consistent with Welch and Goyal (2008), who document that most of the prediction gains stem from the 1973–1975 Oil Shock.

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

4 Tax instruments and portfolios of the rich

The top 1% income share is an endogenous variable in the macro economy. 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. In this section, we adopt two approaches to address this issue. First, we use tax policy as an instrument for inequality. Second, we look at evidence on the portfolios of the rich, which are at the heart of our proposed mechanism linking inequality and returns.

Research on inequality suggests that increases (decreases) in top marginal

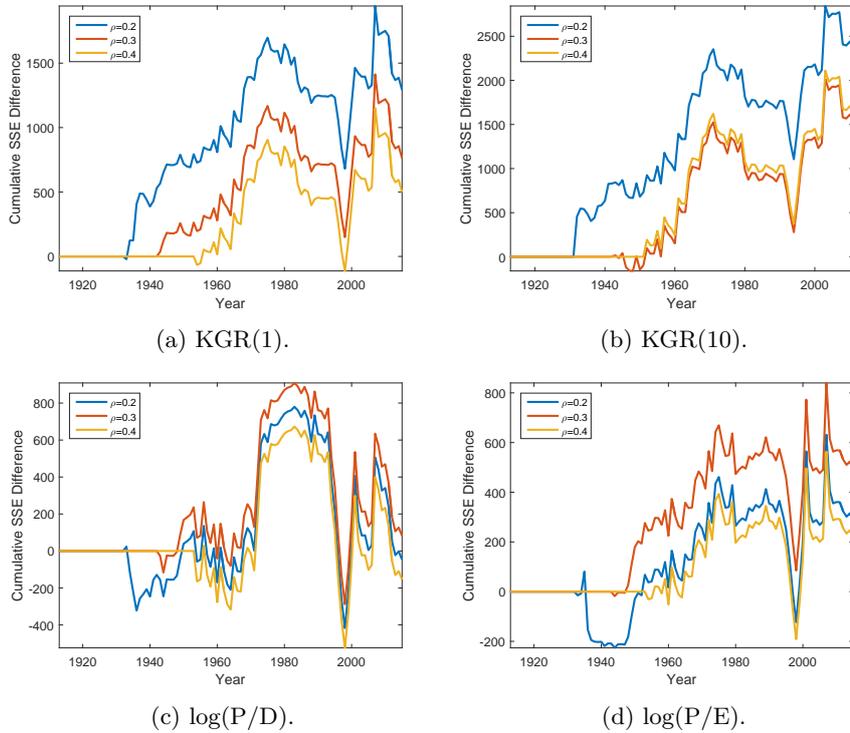


Figure 5: Annual performance in predicting subsequent excess returns.

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

tax rates reduce (exacerbate) inequality (Roine et al., 2009; Kaymak and Poschke, 2016). Indeed, the Piketty-Saez series appear to exhibit a U-shaped trend over the century, which might be due to the change in the marginal income tax rates. According to Figure 6, the marginal tax rate for the highest income earners increased from about 25% to 90% over the period 1930–1945 and started to decline in the 1960s, reaching about 40% in the 1980s. Thus the marginal tax rate exhibits an inverse U-shape that seems to coincide with the trend in the Piketty-Saez series.

Furthermore, top tax rate changes are the result of Congressional bills, which generally take years to pass and usually stem from wars or pro-long-term growth or anti-deficit ideologies (de Rugy, 2003a,b; Jacobson et al., 2007; Weinzierl and Werker, 2009; Romer and Romer, 2010). Therefore, while alterations in top tax rates impact inequality, their timing and justification are likely not the result of financial market fluctuations. Provided top tax rate changes have a muted effect on returns, except via inequality, they can serve as an instrument for top income shares. We address this “excludability” condition below.

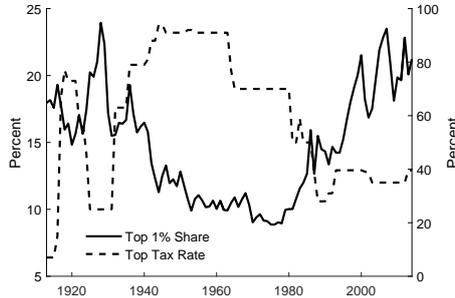


Figure 6: Top 1% income share including capital gains (left axis) and top marginal tax rate (right axis), 1913–2014. Source: IRS.

4.1 Instrumental variables regressions using changes in top estate tax rates

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

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

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

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

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

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

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

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

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

Both Piketty and Saez (2003) and Piketty (2003) argue that income inequality should decline in response to expansion of progressive estate taxation: capital gains comprise a substantial portion of the income of the rich, and high estate taxes decrease the ability and incentive to amass wealth in financial assets. Thus, increasing the top estate tax rate should disproportionately reduce

the wealth of the very rich and subsequently mitigate capital gains income inequality, which is driven by inequality in asset holdings. On the other hand, since estate taxes apply to both realized and unrealized capital gains, it is unlikely that estate taxes affect the timing of realizing capital gains beyond their incentive effects. Therefore current and lagged changes in the estate tax rates are a good candidate for an instrument. The first stage regressions in Table 6 confirm this hypothesis: contemporaneous and lagged changes in the top estate tax rate significantly explain a substantial portion of the variation in KGR(1) (and the 10% and 0.1% analogs).

Table 6: Regressions of KGR on contemporaneous and lagged changes in top estate tax rates

Regressors	Dependent Variable: $KGR(x)_t$		
	0.1%	1%	10%
Constant	1.52	2.37	3.11
ΔETR_t	-0.04***	-0.06***	-0.07***
ΔETR_{t-1}	-0.03**	-0.04*	-0.04*
ΔETR_{t-2}	-0.07***	-0.10***	-0.10***
ΔETR_{t-3}	-0.06***	-0.08***	-0.08***
R^2	0.26	0.24	0.19

Note: the table shows regressions of KGR on lagged changes in top estate tax rates (ETR). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants) according to Newey-West standard errors (4 lags). Sample: 1913–2015. Sources: Tax Foundation and IRS.

Table 6 suggests that changes in top estate tax rates can instrument for KGR in explaining excess returns. Whether one believes this instrument can test causation depends on if lagged changes in estate tax rates are excludable or not. One concern is that estate tax cuts stimulate the economy and thus stock market returns. Another concern is that even if estate tax rates only affect inequality, inequality may simply be proxying for the level of stock market, which we already know predicts returns. To control for these possibilities, we allow KGR, industrial production growth, and $\log(P/E)$ to be endogenous and instrument all three with contemporaneous and three lags of the change in the top estate tax rate (ΔETR for $t, t-1, t-2, t-3$) as well as the lagged price-earnings ratio ($\log(P/E)_{t-1}$).¹⁸ Table 7 shows the results of GMM estimation of the moment condition (4.2) (including industrial production growth and $\log(P/E)$ as controls). KGR is significant at the 5% level in predicting subsequent excess returns regardless of whether we use the top 0.1%, 1%, or 10% income share in constructing KGR.

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

¹⁸Industrial production growth (t) is significantly correlated with ΔETR for $t, t-1$; $\log(P/E)_t$ is significantly correlated with $\log(P/E)_{t-1}$. Hence the rank condition for identification holds.

Table 7: Instrumental variables GMM estimates of the effect of KGR, industrial production growth, and $\log(P/E)$ on one year excess stock market returns

Dependent Variable: t to $t + 1$ Excess Market Return			
Regressors (t)	KGR(x) version		
	0.1%	1%	10%
Constant	18.09 (24.05)	22.58 (23.85)	28.43 (24.78)
KGR(x)	-10.79** (4.54)	-7.52** (3.27)	-6.91** (3.08)
% Δ IP	-1.51*** (0.51)	-1.49*** (0.49)	-1.46*** (0.48)
$\log(P/E)$	3.71 (9.98)	2.61 (10.02)	1.90 (10.64)
J statistic	0.65 ($p = 0.72$)	0.69 ($p = 0.71$)	0.75 ($p = 0.69$)

Note: the table shows the results of two-step GMM estimation of the moment condition (4.2) (including industrial production growth and $\log(P/E)$ as controls). The initial weighting matrix is identity, and the second stage one is Newey-West (4 lags). Newey-West standard errors are in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). % Δ IP is the annual % change in the industrial production index. P/E is the S&P500 price-earnings ratios. The instruments are a constant, changes in the top estate tax rate (Δ ETR for $t, t - 1, t - 2, t - 3$), and the lagged price-earnings ratio ($\log(P/E)_{t-1}$). Sample: 1913–2015. Sources: Tax Foundation, IRS, and FRED.

we can show that the moment condition

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

holds, where $\alpha_2 = \beta_0 + \alpha\beta_2$. What would be a good candidate for z_{2t} ? Rational agents have an incentive to realize (delay) capital gains if they expect the capital gains tax rate to increase (decrease). Since tax rates in year $t + 1$ are announced in year t , we can use the change in the maximum capital gains tax rate from year t to $t + 1$, Δ CGTR $_{t+1}$, as an instrument z_{2t} for the timing component of KGR, x_{2t} . Table 8 adds Δ CGTR $_{t+1}$ to the first-stage regressions displayed in Table 6. As conjectured, the change in top capital gains tax rates from year t to $t + 1$ have positive and significant relationship with year t KGR. Current and lagged changes in estate tax rates, however, continue to have a strong inverse association with KGR. As rising capital gains and estate tax rates should, all else equal, discourage wealth accumulation amongst the rich, the positive coefficient on Δ CGTR $_{t+1}$ is likely reflecting the timing component of KGR (x_{2t}): when the rich expect capital gains taxes to rise, they move forward the realization of capital gains, which causes KGR to rise.

Thus, in Table 9 we jointly estimate the moment conditions (4.2) and (4.3) (including industrial production growth, $\log(P/E)$, and Δ CGTR $_{t+1}$ as controls) by multiple equation GMM using the instruments

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

$$z_{2t} = (1, \Delta$$
CGTR $_{t+1})'$,

Table 8: Regressions of KGR on contemporaneous and lagged changes in top estate tax rates and the one-period-ahead change in the capital gains tax rate

Regressors	Dependent Variable: $KGR(x)_t$		
	0.1%	1%	10%
Constant	1.54	2.39	3.14
ΔETR_t	-0.04***	-0.06***	-0.07***
ΔETR_{t-1}	-0.04***	-0.05**	-0.06**
ΔETR_{t-2}	-0.07***	-0.09***	-0.09**
ΔETR_{t-3}	-0.06***	-0.08***	-0.08***
$\Delta CGTR_{t+1}$	0.03***	0.04***	0.05***
R^2	0.29	0.27	0.22

Note: see caption of Table 6 for explanations. $\Delta CGTR_{t+1}$ is the one-period-ahead change in the maximum capital gains tax rate.

Table 9: Instrumental variables multiple equation GMM estimates of the effect of KGR, industrial production growth, and $\log(P/E)$ on one year excess stock market returns

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return		
	KGR(x) version		
	0.1%	1%	10%
Constant (α_1)	21.59 (25.14)	27.27 (25.05)	32.83 (25.81)
Constant (α_2)	-11.94 (64.50)	-6.09 (66.36)	-7.75 (76.83)
KGR(x) (inequality, β_1)	-10.93** (4.42)	-7.75** (3.19)	-7.17** (2.96)
KGR(x) (timing, β_2)	11.45 (34.02)	6.43 (22.77)	5.92 (20.59)
% ΔIP	-1.48* (0.88)	-1.47* (0.85)	-1.42* (0.83)
$\log(P/E)$	2.47 (10.13)	1.07 (10.22)	0.51 (10.74)
$\Delta CGTR_{t+1}$	-0.05 (1.17)	0.01 (1.12)	-0.03 (1.13)
J statistic	0.49 ($p = 0.49$)	0.50 ($p = 0.48$)	0.54 ($p = 0.46$)

Note: the table shows the results of two-step multiple equation GMM estimation of the moment conditions (4.2) and (4.3) (including industrial production growth, $\log(P/E)$, and $\Delta CGTR_{t+1}$ as controls). See caption of Table 7 for explanations.

respectively. The coefficients are positive but insignificant for the timing components (x_{2t}) identified by changes in future capital gains tax rates. The inequality components (x_{1t}), however, have negative and significant coefficients. This is true regardless of whether we use the top 0.1%, 1%, or 10% income share, and suggests that the causal effect of KGR on subsequent excess returns is driven by inequality rather than by the timing of capital gains realization.

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

4.2 Portfolios of the rich

While we showed in Section 3.2 that KGR predicts excess returns, two important concerns are whether KGR can be interpreted as a measure of inequality and, even if so, does rising inequality create excess demand for risky assets (from the rich), as in our theory, or simply correlate with other predictors. We argued in Section 4.1 by using tax instruments that at least part of KGR is capturing inequality and that the timing of realizing capital gains (the rich’s information) is not driving KGR. Although prediction appears robust to IV and including many controls, one might still wonder about omitted variables and the mechanism through which inequality affects returns.

To help address these concerns, in this section we provide evidence on the portfolios of the rich. A key implication of our theory is that redistribution from natural bondholders to stockholders should trigger further concentration of equities and a subsequent decline in the equity premium. We have already demonstrated that inequality predicts excess returns, and there are many papers showing the rich are the marginal buyers of stock (see Footnote 8). The remaining testable implication at first seems to be a link between equity concentration amongst the rich and subsequent excess returns.

But the relationship between equity concentration and future returns may be ambiguous. Recent studies have argued that investor returns rise with wealth (due to sophistication, skill, and/or information), even in the tail of the wealth distribution (Kacperczyk et al., 2014; Fagereng et al., 2016a,b). So, while rising equity concentration could reflect rising inequality and predict lower excess returns, increased concentration of equities in the hands of rich could also be due to the wealthy better exploiting positive market information and thus predict higher excess returns. In this section, we attempt to disentangle these two forces and provide evidence suggesting the inequality driven component of equity concentration does indeed forecast lower excess returns, as in our theory.

In Table 10 we regress the change (from year t to $t+1$) in the wealth portfolio composition of the rich, stock and bond wealth inequality in particular, on KGR in year t . The portfolio data is from Saez and Zucman (2016), who compute it from tax returns data but show that it is in line with values in the Survey of Consumer Finances.¹⁹ We compute asset inequality by dividing the asset’s level

¹⁹Our sources are the “AppendixTables(Aggregates)” and “AppendixTables(Distributions)”

contribution to the particular top wealth share by the total fraction of wealth in that asset class. So the 1% equity share (call it Eq(1)), is the fraction of equity wealth held by the richest 1%.

Table 10: Regressions of stock and bond wealth inequality on KGR

Regressors (t)	Dependent: t to $t + 1$ change in asset class wealth share					
	Equities share ($\Delta\text{Eq}(x)_{t+1}$)			Bonds share		
	0.1%	1%	10%	0.1%	1%	10%
Constant	-0.98 (0.52)	-1.35 (0.58)	-0.48 (0.21)	-0.03 (0.47)	-0.45 (0.62)	-0.36 (0.28)
KGR(x)	0.64*** (0.24)	0.52*** (0.19)	0.15*** (0.05)	0.07 (0.25)	0.21 (0.21)	0.09 (0.07)
Sample	1913- -2012	1913- -2012	1917- -2012	1913- -2012	1913- -2012	1917- -2012
R^2	0.06	0.06	0.05	0.00	0.01	0.01

Note: Newey-West standard errors in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). For each asset class (equities or bonds) and $x = 0.1, 1, 10$, the table shows regressions of the t to $t + 1$ change in the share of assets owned by the wealthiest $x\%$ on time t KGR(x) and a constant.

Table 10 shows that the equity coefficients from regressing ΔEq_{t+1} on KGR_t are positive and significant (equity wealth inequality rises), which suggests that the rich subsequently invest more in equities after KGR increases. For bonds the coefficients are insignificant. This is consistent with our story in which KGR measures inequality and the rich are more risk tolerant. When the rich get richer, there is subsequently a relative rise in stock wealth concentration (vs. bond wealth inequality). Despite the significant correlation between rising inequality and subsequent equity concentration in Table 10, equity inequality does not forecast lower excess returns. As we see in Table 11, there is no significant relationship between the change in the 0.1% or 10% equity share and subsequent excess returns. And the change in the 1% equity share significantly forecasts *higher* excess returns.

The reason for this correlation, which may be puzzling at first glance, is that top equity shares reflect not just inequality but also other factors that correlate with subsequent returns. Table 12 regresses ΔEq_t on KGR_{t-1} , $\log(\text{P/E})_t$, and ΔCGTR_{t+1} (the change in the maximum capital gains tax rate). For the 0.1%, 1%, and 10%, top equity shares are significantly positively correlated with lagged inequality. But equity concentration is also inversely related with the current level of the stock market and rising capital gains taxes (although the coefficients are only significant for the 1%, and the 10% ΔCGTR_{t+1} coefficient is estimated to be 0).

Since a low price-earnings ratio and expected capital gains tax cuts are associated with rising excess returns, it is now clear why equity concentration fails to forecast lower excess returns: top equity shares move for many reasons (including income inequality), but some drivers of equity inequality predict rising excess returns. In light of the research arguing the wealthy are more sophisticated investors that get better returns, there is a natural interpretation of spreadsheets for Saez and Zucman (2016).

Table 11: Regressions of one year excess stock market returns on high wealth equity shares

Dependent Variable:			
t to $t + 1$ Excess Market Return			
$\Delta\text{Eq}(x)$ version			
Regressors (t)	0.1%	1%	10%
Constant	6.21 (1.89)	6.37 (1.93)	6.18 (1.91)
$\Delta\text{Eq}(x)$	0.81 (0.68)	1.02** (0.50)	1.00 (1.43)
Sample	1913- -2012	1913- -2012	1917- -2012
R^2	0.01	0.03	0.00

Note: Newey-West standard errors in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). For $x = 0.1, 1, 10$, the table shows regressions of the t to $t + 1$ excess market return on the change in the share of equities owned by the wealthiest $x\%$ and a constant.

Table 12: Decomposition of high wealth equity shares

Dependent: change in high wealth equity share			
$\Delta\text{Eq}(x)_t$ version			
Regressors (t)	0.1%	1%	10%
Constant	0.35 (2.07)	1.92 (1.54)	0.54 (0.77)
$\text{KGR}(x)_{t-1}$	0.75*** (0.28)	0.66** (0.25)	0.20*** (0.07)
$\log(\text{P}/\text{E})_t$	-0.55 (0.78)	-1.33* (0.69)	-0.43 (0.31)
ΔCGTR_{t+1}	-0.04 (0.04)	-0.13** (0.05)	0.00 (0.01)
Sample	1913- -2012	1913- -2012	1917- -2012
R^2	0.07	0.18	0.06

Note: Newey-West standard errors in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). For $x = 0.1, 1, 10$, the table shows regressions of the change in the share of equities owned by the wealthiest $x\%$ on lagged $\text{KGR}(x)$, $\log(\text{P}/\text{E})$, the one-period-ahead change in the maximum capital gains tax rate, and a constant.

these findings. All else equal, rising equity concentration (from, say, increasing inequality), pushes down the equity premium. But equity concentration might also rise precisely *because* the stock market is going to do well from, for example, tax cuts or mean reversion and since the rich are better at timing the market.

To see that inequality driven equity concentration (vs. timing/sophistication driven concentration) inversely forecasts excess returns, we use KGR_{t-1} as an instrument for ΔEq_t . Table 13 shows that the instrumented top equity shares coefficients are negative and significant with respect to the 1% and 10%. (For

the 0.1%, the coefficient is negative but the p-value is 0.118.)

Table 13: Instrumental variables estimates of the effect of high wealth equity shares on one year excess stock market returns

Dependent Variable:			
t to $t + 1$ Excess Market Return			
$\Delta \text{Eq}(x)$ version			
Regressors (t)	0.1%	1%	10%
Constant	5.06 (2.57)	4.44 (2.42)	5.41 (3.15)
$\Delta \text{Eq}(x)$	-7.71 (4.93)	-6.41* (3.85)	-19.81** (9.19)
Sample	1913- -2012	1913- -2012	1917- -2012

Note: for $x = 0.1, 1, 10$, the table shows instrumental variables regressions of the t to $t + 1$ excess market return on the change in the share of equities owned by the wealthiest $x\%$ and a constant. Newey-West standard errors are in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). The instrument is $\text{KGR}(x)_{t-1}$.

5 International evidence

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

5.1 Data

We consider 29 countries, for the time period 1969–2015, spanning the continents: Americas (Argentina, Canada, Colombia, and 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, Japan, Malaysia, Singapore, South Korea, and Taiwan), and Oceania (Australia, Indonesia, and New Zealand). Due to missing data points for some countries, we have 815 observations when we include all countries and time periods.

The inequality panel data are from the World Wealth and Income Database.²⁰ To be consistent across countries, we use the “fiscal income” top 1% income share series. To calculate annual stock returns (end-of-period), we acquire from

²⁰<http://wid.world/>

Datastream the MSCI total return indexes in local currency. To convert returns into local real terms, we deflate the stock indexes by local CPI, which we obtain from the World Bank website,²¹ Haver Analytics,²² FRED, and the Taiwanese government.²³ See Appendix F for country-specific details.

5.2 International regression results

In Section 3, we showed that income concentration is inversely related to subsequent *excess* returns. However, quantitatively, this result was really about stock returns. Indeed, redoing column (1) of Table 2 with stock returns instead of excess returns, the KGR(1) coefficient is -2.61 with a Newey-West p-value of 0.027. With the 10% or 0.1% series, the coefficients are, respectively, -2.16 and -3.53 with p-values of 0.002 and 0.041. 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, while U.S. Treasury 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. Additionally, 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, we perform the international analysis using 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. Furthermore, for many countries we cannot calculate KGR(1) (because as of the time of writing, the WID database does not provide the top 1% series excluding *and* including capital gains income for many countries) and the samples are short or missing years. Therefore we use the raw top 1% income share data rather than attempt to estimate and remove common or heterogeneous time trends. Table 14 presents the panel regression results for both the whole sample and different groups.

First, we see in column (1) that when including all countries a one percentage point increase in the top income share is associated with a subsequent decline in stock market returns of 0.94% on average. The coefficient is significant at the 10% level with standard errors clustered by country (results are similar without clustering). column (2) does the same regression for only advanced economies (Australia, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, New Zealand, Portugal, Singapore, Spain, Sweden, Switzerland, Taiwan, U.K., and U.S.), and we obtain similar results.

Is the predictive power of the top income share uniform across countries? In either very large markets (such as 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, foreign investors own a substantial fraction of the domestic stock markets and should mitigate the role of local inequality. However, even if local inequality is less important in small and open financial markets, inequality amongst global investors should still impact returns in these markets.

²¹<http://data.worldbank.org/>

²²<http://www.haver.com/>

²³eng.stat.gov.tw

Table 14: Country fixed effects panel regressions of one year stock returns on local top income shares and U.S. KGR

Regressors (t)	Dependent Variable: t to $t + 1$ Stock Return			
	(1) All	(2) Advanced	(3) ex-U.S.	(4) ex-U.S.
Top 1%	-0.94* (0.52)	-1.01* (0.49)	-0.42 (0.70)	2.61 (1.55)
U.S. KGR(1)			-2.51*** (0.43)	-0.53 (0.75)
Top 1% ×homebias				-5.44** (2.42)
U.S. KGR(1) ×(1 – homebias)				-4.17** (1.60)
Country FE	✓	✓	✓	✓
Obs.	815	712	769	687
R^2 (w,b)	(0.00,0.05)	(0.01,0.03)	(0.02,0.13)	(0.03,0.27)

Note: Clustered standard errors in parentheses, ***1%, **5%, *10% (constants suppressed). ex-U.S.: All countries excluding U.S. R^2 (w,b): Within and between R-squared. Top 1% is the share of income going to the top 1% of earners (the “fiscal income” top 1% series from World Wealth and Income Database). The home bias measure is from Mishra (2015). See the main text and Appendix F for country details on series construction. Sample: 1969–2015.

In column (3) of Table 14, we also include the U.S. KGR(1) as a proxy for global investor inequality. Doing so makes the U.S. KGR(1) very significant but the local inequality insignificant. However, as we argued above, local inequality should matter only for relatively closed economies. Therefore in column (4), we also include the interaction terms between inequality and home bias measures. For this regression we use the ICAPM equity home bias measure of Mishra (2015), which takes value between 0 (no home bias) and 1 (complete home bias).²⁴ Specifically, we consider the model

$$\log R_{i,t+1} = \alpha_i + \beta_1 \text{top}(1)_{it} + \beta_2 \text{KGR}(1)_{\text{US},t} + \beta_3 \text{top}(1)_{it} \times \text{homebias}_i + \beta_4 \text{KGR}(1)_{\text{US},t} \times (1 - \text{homebias}_i) + \varepsilon_{i,t+1},$$

where α_i is the country fixed effect. According to column (4) of Table 14, the coefficients on the interaction terms (β_3, β_4) are negative and highly significant, while the linear terms (β_1, β_2) are insignificant. In particular, in a closed economy (homebias = 1), a one percentage point increase in the top 1% income share is associated with a subsequent decline in stock market returns of $-\beta_1 - \beta_3 = 2.83\%$ on average; in a small open economy (homebias = 0), a one percentage point increase in U.S. KGR(1) is associated with a subsequent decline in stock market returns of $-\beta_2 - \beta_4 = 4.70\%$. These findings are consistent with the conjecture that the local 1% share negatively predicts returns only for countries with higher home bias (relatively closed economies), and the global 1% share (proxied by U.S. KGR(1)) matters only for countries with lower home bias (small open economies).

²⁴Mishra (2015) does not calculate a home bias number for China, Ireland, Mauritius, or Taiwan.

6 Concluding remarks

In this paper we built a general equilibrium model with agents that are heterogeneous in wealth, risk aversion, and belief. We showed that the concentration of wealth/income drives down the subsequent equity premium. Our model is a mathematical formulation of Irving Fisher’s narrative that booms and busts are caused by changes in the relative wealth of the rich (the “enterpriser-borrower”) and the poor (the “creditor, the salaried man, or the laborer”). Consistent with our theory, we found that in the U.S. 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. The inverse relationship between returns and inequality is robust to controlling for standard return predictors and instrumenting with changes in estate taxes. It also holds outside of the U.S., although in relatively open economies with low home bias it is *U.S.* inequality that matters (as opposed to domestic inequality).

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

References

- Andrew Ang and Geert Bekaert. Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707, 2007. doi:10.1093/rfs/hhl021.
- Kenneth J. Arrow. *Aspects of the Theory of Risk-Bearing*. Yrjö Jahanssonin Säätiö, 1965.
- Pierluigi Balduzzi and Tong Yao. Testing heterogeneous-agent models: An alternative aggregation approach. *Journal of Monetary Economics*, 54(2): 369–412, March 2007. doi:10.1016/j.jmoneco.2005.08.021.
- Ravi Bansal, Dana Kiku, Ivan Shaliastovich, and Amir Yaron. Volatility, the macroeconomy, and asset prices. *Journal of Finance*, 69(6):2471–2511, December 2014. doi:10.1111/jofi.12110.
- Suleyman Basak and Domenico Cuoco. An equilibrium model with restricted stock market participation. *Review of Financial Studies*, 11(2):309–341, 1998. doi:10.1093/rfs/11.2.309.
- Claude Berge. *Espaces Topologiques: Fonctions Multivoques*. Dunod, Paris, 1959. English translation: Translated by E. M. Patterson. *Topological Spaces*, New York: MacMillan, 1963. Reprinted: Mineola, NY: Dover, 1997.
- Harjoat S. Bhamra and Raman Uppal. Asset prices with heterogeneity in preferences and beliefs. *Review of Financial Studies*, 27(2):519–580, 2014. doi:10.1093/rfs/hht051.

- Jacob Boudoukh, Matthew Richardson, and Robert F. Whitelaw. The myth of long-horizon predictability. *Review of Financial Studies*, 21(4):1577–1605, 2008. doi:10.1093/rfs/hhl042.
- Alon Brav, George M. Constantinides, and Christopher C. Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110(4):793–824, August 2002. doi:10.1086/340776.
- Alessandro Buccioli and Raffaele Miniaci. Household portfolio and implicit risk preference. *Review of Economics and Statistics*, 93(4):1235–1250, 2011. doi:10.1162/REST_a.00138.
- Laurent E. Calvet and Paolo Sodini. Twin picks: Disentangling the determinants of risk-taking in household portfolios. *Journal of Finance*, 69(2):867–906, April 2014. doi:10.1111/jofi.12125.
- John Y. Campbell. Household finance. *Journal of Finance*, 61(4):1553–1604, August 2006. doi:10.1111/j.1540-6261.2006.00883.x.
- John Y. Campbell and Robert J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228, 1988. doi:10.1093/rfs/1.3.195.
- John Y. Campbell and Samuel B. Thomson. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531, 2008. doi:10.1093/rfs/hhm055.
- Sean D. Campbell, Stefanos Delikouras, Danling Jiang, and George M. Korniotis. The human capital that matters: Expected returns and high-income households. *Review of Financial Studies*, 29(9):2523–2563, 2016. doi:10.1093/rfs/hhw048.
- Christopher D. Carroll. Portfolios of the rich. In Luigi Guiso, Michael Haliassos, and Tullio Jappelli, editors, *Household Portfolios*, chapter 10, pages 389–430. MIT Press, Cambridge, MA, 2002.
- Georgy Chabakauri. Dynamic equilibrium with two stocks, heterogeneous investors, and portfolio constraints. *Review of Financial Studies*, 26(12):3104–3141, December 2013. doi:10.1093/rfs/hht030.
- Georgy Chabakauri. Asset pricing with heterogeneous preferences, beliefs, and portfolio constraints. *Journal of Monetary Economics*, 75:21–34, October 2015. doi:10.1016/j.jmoneco.2014.11.012.
- Yeung Lewis Chan and Leonid Kogan. Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy*, 110(6):1255–1285, December 2002. doi:10.1086/342806.
- Yili Chien, Harold Cole, and Hanno Lustig. A multiplier approach to understanding the macro implications of household finance. *Review of Economic Studies*, 78(1):199–234, January 2011. doi:10.1093/restud/rdq008.

- YiLi Chien, Harold Cole, and Hanno Lustig. Is the volatility of the market price of risk due to intermittent portfolio rebalancing? *American Economic Review*, 102(6):2859–2896, October 2012. doi:10.1257/aer.102.6.2859.
- John S. Chipman. Homothetic preferences and aggregation. *Journal of Economic Theory*, 8(1):26–38, May 1974. doi:10.1016/0022-0531(74)90003-9.
- John H. Cochrane. The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575, 2008. doi:10.1093/rfs/hhm046.
- Timothy Cogley. Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics*, 49(2):309–334, March 2002. doi:10.1016/S0304-3932(01)00106-4.
- George M. Constantinides. Intertemporal asset pricing with heterogeneous agents without demand aggregation. *Journal of Business*, 55(2):253–267, April 1982.
- George M. Constantinides and Darrell Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104(2):219–240, April 1996. doi:10.1086/262023.
- George M. Constantinides and Anisha Ghosh. Asset pricing with countercyclical household consumption risk. *Journal of Finance*, 72(1):415–460, February 2017. doi:10.1111/jofi.12471.
- Jakša Cvitanić, Elyès Jouini, Semyon Malamud, and Clotilde Napp. Financial markets equilibrium with heterogeneous agents. *Review of Finance*, 16(1):285–321, 2012. doi:10.1093/rof/rfr018.
- Veronique de Rugy. Tax rates and tax revenue: The Mellon income tax cuts of the 1920s. *Cato Institute Tax and Budget Bulletin*, (13), February 2003a. URL <http://object.cato.org/sites/cato.org/files/pubs/pdf/tbb-0302-13.pdf>.
- Veronique de Rugy. High taxes and high budget deficits: The Hoover-Roosevelt tax increases of the 1930s. *Cato Institute Tax and Budget Bulletin*, (14), March 2003b. URL <http://object.cato.org/sites/cato.org/files/pubs/pdf/tbb-0303-14.pdf>.
- Bernard Dumas. Two-person dynamic equilibrium in the capital market. *Review of Financial Studies*, 2(2):157–188, 1989. doi:10.1093/rfs/2.2.157.
- Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity in returns to wealth and the measurement of wealth inequality. *American Economic Review: Papers and Proceedings*, 106(5):651–655, May 2016a. doi:10.1257/aer.p20161022.
- Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity and persistence in returns to wealth. NBER Working Paper 22822, 2016b.
- Eugene F. Fama and Kenneth R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25, October 1988. doi:10.1016/0304-405X(88)90020-7.

- Irving Fisher. *Introduction to Economic Science*. Macmillan, NY, 1910.
- Nicolae Gârleanu and Stavros Panageas. Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing. *Journal of Political Economy*, 123(3):670–685, June 2015. doi:10.1086/680996.
- John Geanakoplos. An introduction to general equilibrium with incomplete asset markets. *Journal of Mathematical Economics*, 19(1-2):1–38, 1990. doi:10.1016/0304-4068(90)90034-7.
- John Geanakoplos and Martin Shubik. The capital asset pricing model as a general equilibrium with incomplete markets. *Geneva Papers on Risk and Insurance Theory*, 15(1):55–71, 1990. doi:10.1007/BF01498460.
- John Geanakoplos and Kieran James Walsh. Uniqueness and stability of equilibrium in economies with two goods. *Journal of Economic Theory*, 174:261–272, March 2018. doi:10.1016/j.jet.2017.12.005.
- Christian Gollier. Wealth inequality and asset pricing. *Review of Economic Studies*, 68(1):181–203, January 2001. doi:10.1111/1467-937X.00165.
- William M. Gorman. Community preference fields. *Econometrica*, 21(1):63–80, January 1953. doi:10.2307/1906943.
- Clive W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121–130, May 1981. doi:10.1016/0304-4076(81)90079-8.
- Daniel L. Greenwald, Martin Lettau, and Sydney C. Ludvigson. *Origins of stock market fluctuations*. 2016.
- Fatih Guvenen. A parsimonious macroeconomic model for asset pricing. *Econometrica*, 77(6):1711–1750, November 2009. doi:10.3982/ECTA6658.
- Michael Haliassos and Carol C. Bertaut. Why do so few hold stocks? *Economic Journal*, 105(432):1110–1129, September 1995. doi:10.2307/2235407.
- James D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.
- James D. Hamilton. Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, 2018. doi:10.1162/REST_a_00706.
- Peter Reinhard Hansen and Allan Timmermann. Equivalence between out-of-sample forecast comparisons and Wald statistics. *Econometrica*, 83(6):2485–2505, November 2015. doi:10.3982/ECTA10581.
- Chiaki Hara, James Huang, and Christoph Kuzmics. Representative consumer’s risk aversion and efficient risk-sharing rules. *Journal of Economic Theory*, 137(1):652–672, November 2007. doi:10.1016/j.jet.2006.11.002.
- Juan Carlos Hatchondo. A quantitative study of the role of wealth inequality on asset prices. *Federal Reserve Bank of Richmond Economic Quarterly*, 94(1):73–96, 2008.

- John Heaton and Deborah J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy*, 104(3):443–487, June 1996. doi:10.1086/262030.
- Robert J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5(3):357–386, 1992. doi:10.1093/rfs/5.3.351.
- Darien B. Jacobson, Brian G. Raub, and Barry W. Johnson. The estate tax: Ninety years and counting. *SOI Bulletin*, 27(1):118–128, Summer 2007. URL <https://www.irs.gov/pub/irs-soi/ninetyestate.pdf>.
- Timothy C. Johnson. Inequality risk premia. *Journal of Monetary Economics*, 59(6):565–580, October 2012. doi:10.1016/j.jmoneco.2012.06.008.
- Kenneth L. Judd. Projection methods for solving aggregate growth models. *Journal of Economic Theory*, 58(2):410–452, December 1992. doi:10.1016/0022-0531(92)90061-L.
- Marcin Kacperczyk, Jaromir B. Nosal, and Luminita Stevens. Investor sophistication and capital income inequality. NBER Working Paper 20246, 2014.
- Barış Kaymak and Markus Poschke. The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics*, 77:1–25, February 2016. doi:10.1016/j.jmoneco.2015.10.004.
- Timothy J. Kehoe. Uniqueness and stability. In Alan Kirman, editor, *Elements of General Equilibrium Analysis*, chapter 3, pages 38–87. Wiley-Blackwell, 1998.
- Narayana R. Kocherlakota and Luigi Pistaferri. Asset pricing implications of Pareto optimality with private information. *Journal of Political Economy*, 117(3):555–590, June 2009. doi:10.1086/599761.
- Wojciech Kopczuk and Emmanuel Saez. Top wealth shares in the United States, 1916–2000: Evidence from estate tax returns. *National Tax Journal*, 57(2):445–487, June 2004. doi:10.17310/ntj.2004.2S.05.
- Martin Lettau and Sydney C. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849, June 2001. doi:10.1111/0022-1082.00347.
- Martin Lettau and Sydney C. Ludvigson. Measuring and modeling variation in the risk-return trade-off. In Yacine Aït-Sahalia and Lars Peter Hansen, editors, *Handbook of Financial Econometrics*, volume 1, chapter 11, pages 617–690. Elsevier, Amsterdam, 2010. doi:10.1016/B978-0-444-50897-3.50014-6.
- Martin Lettau, Sydney C. Ludvigson, and Jessica A. Wachter. The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies*, 21(4):1653–1687, 2008. doi:10.1093/rfs/hhm020.
- John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37, February 1965a. doi:10.2307/1924119.

- John Lintner. Security prices, risk, and maximal gains from diversification. *Journal of Finance*, 20(4):587–615, December 1965b. doi:10.1111/j.1540-6261.1965.tb02930.x.
- Francis A. Longstaff and Jiang Wang. Asset pricing and the credit market. *Review of Financial Studies*, 25(11):3169–3215, 2012. doi:10.1093/rfs/hhs086.
- Sydney C. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. In George M. Constantinides, Milton Harris, and René M. Stultz, editors, *Handbook of the Economics of Finance*, volume 2, chapter 12, pages 799–906. Elsevier, Amsterdam, 2013. doi:10.1016/B978-0-44-459406-8.00012-3.
- N. Gregory Mankiw. The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics*, 17(1):211–219, September 1986. doi:10.1016/0304-405X(86)90012-7.
- Rolf R. Mantel. Homothetic preferences and community excess demand functions. *Journal of Economic Theory*, 12(2):197–201, April 1976. doi:10.1016/0022-0531(76)90073-9.
- Michael W. McCracken. Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, 140(2):719–752, October 2007. doi:10.1016/j.jeconom.2006.07.020.
- Anil V. Mishra. Measures of equity home bias puzzle. *Journal of Empirical Finance*, 34:293–312, December 2015. doi:10.1016/j.jempfin.2015.08.001.
- Charles R. Nelson and Myung J. Kim. Predictable stock returns: The role of small sample bias. *Journal of Finance*, 48(2):641–661, June 1993. doi:10.1111/j.1540-6261.1993.tb04731.x.
- Peter C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988. doi:10.1093/biomet/75.2.335.
- Thomas Piketty. Income inequality in France, 1901–1998. *Journal of Political Economy*, 111(5):1004–1042, October 2003. doi:10.1086/376955.
- Thomas Piketty and Emmanuel Saez. Income inequality in the United States, 1913–1998. *Quarterly Journal of Economics*, 118(1):1–41, February 2003. doi:10.1162/00335530360535135.
- Walter Pohl, Karl Schmedders, and Ole Wilms. Higher-order effects in asset pricing models with long-run risks. *Journal of Finance*, 2018. doi:10.1111/jofi.12615.
- David E. Rapach and Guofu Zhou. Forecasting stock returns. In Graham Elliott and Allan Timmerman, editors, *Handbook of Economic Forecasting*, volume 2, chapter 6, pages 328–383. Elsevier, 2013. doi:10.1016/B978-0-444-53683-9.00006-2.
- David E. Rapach, Jack K. Strauss, and Guofu Zhou. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2), 821–862 2010. doi:10.1093/rfs/hhp063.

- Jesper Roine, Jonas Vlachos, and Daniel Waldenström. The long-run determinants of inequality: What can we learn from top income data? *Journal of Public Economics*, 93(7-8):974–988, August 2009. doi:10.1016/j.jpubeco.2009.04.003.
- Christina D. Romer and David H. Romer. The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, 100(3):763–801, June 2010. doi:10.1257/aer.100.3.763.
- Emmanuel Saez and Gabriel Zucman. Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, 131(2):519–578, May 2016. doi:10.1093/qje/qjw004.
- William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, September 1964. doi:10.1111/j.1540-6261.1964.tb02865.x.
- Robert F. Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, December 1999. doi:10.1016/S0304-405X(99)00041-0.
- Kjetil Storesletten, Chris I. Telmer, and Amir Yaron. Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics*, 10(4):519–548, October 2007. doi:10.1016/j.red.2007.02.004.
- Alexis Akira Toda and Kieran Walsh. The double power law in consumption and implications for testing Euler equations. *Journal of Political Economy*, 123(5):1177–1200, October 2015. doi:10.1086/682729.
- Alexis Akira Toda and Kieran James Walsh. Edgeworth box economies with multiple equilibria. *Economic Theory Bulletin*, 5(1):65–80, April 2017a. doi:10.1007/s40505-016-0102-3.
- Alexis Akira Toda and Kieran James Walsh. Fat tails and spurious estimation of consumption-based asset pricing models. *Journal of Applied Econometrics*, 32(6):1156–1177, September/October 2017b. doi:10.1002/jae.2564.
- Rossen Valkanov. Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics*, 68(2):201–232, May 2003. doi:10.1016/S0304-405X(03)00065-5.
- Annette Vissing-Jørgensen. Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. NBER Working Paper, 2002. URL <http://www.nber.org/papers/w8884>.
- Jessica A. Wachter and Motohiro Yogo. Why do household portfolio shares rise in wealth? *Review of Financial Studies*, 23(11):3929–3965, 2010. doi:10.1093/rfs/hhq092.
- Jiang Wang. The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics*, 41(1):75–110, May 1996. doi:10.1016/0304-405X(95)00854-8.
- Matthew C. Weinzierl and Eric D. Werker. Barack Obama and the Bush tax cuts (A). Case 709-037, Harvard Business School, January 2009.

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

A Proofs

A.1 Proof of Theorem 2.1

Since by Assumption 2 we have $e_i = 0$, the utility maximization problem becomes

$$\begin{aligned} & \text{maximize} && U_i(x) \\ & \text{subject to} && q'y \leq q'n_i, 0 \leq x \leq Ay. \end{aligned} \quad (\text{A.1})$$

Step 1. The planner's problem (2.2) has a unique solution.

Proof. Let

$$\Omega = \left\{ x = (x_i) \in \mathbb{R}_+^{SI} \mid (\exists y = (y_i)) (\forall i) x_i \leq Ay_i, \sum_{i=1}^I y_i = n \right\}$$

be the set of all feasible consumption allocations. Then the planner's problem (2.2) is equivalent to maximizing $f(x) = \sum_{i=1}^I w_i \log U_i(x_i)$ subject to $x \in \Omega$. By Assumption 1 and Berge (1959, p. 208, Theorem 3), each $U_i(x_i)$ is strictly concave. Since $\log(\cdot)$ is increasing and strictly concave, so is $\log U_i(x_i)$. Since f is continuous and strictly concave, to show the existence and uniqueness of a solution, it suffices to show that Ω is nonempty, compact, and convex. Clearly $\Omega \neq \emptyset$ because we can choose the initial endowment $y_i = n_i$ and $x_i = An_i = e_i$. Since Ω is defined by linear inequalities and equations, it is closed and convex. If $x \in \Omega$, by definition we can take $y = (y_i)$ such that $x_i \leq Ay_i$ for all i and $\sum_{i=1}^I y_i = n$. Then

$$\sum_{i=1}^I x_i \leq \sum_{i=1}^I Ay_i = A \sum_{i=1}^I y_i = An.$$

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

Let $x = (x_i)$ be the unique maximizer of f on Ω . Since f is strictly increasing, we have $x_i = Ay_i$ for some $y = (y_i)$ such that $\sum_{i=1}^I y_i = n$. If there is another such $y' = (y'_i)$, then $Ay_i = Ay'_i \iff A(y_i - y'_i) = 0$. Since by assumption A has full column rank, we have $y_i - y'_i = 0 \iff y_i = y'_i$. Therefore the planner's problem (2.2) has a unique solution. \square

Step 2. $x = (x_i)$ is a GEI equilibrium allocation and the Lagrange multiplier to the planner's problem gives the asset prices.

Proof. Let

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

be the Lagrangian of the planner's problem (2.2). By the previous step, a unique solution $y = (y_i)$ exists. Furthermore, since U_i satisfies the Inada condition, it must be $Ay_i \gg 0$. Hence by the first-order condition and the chain rule, we have

$$q' = w_i \frac{DU_i(Ay_i)A}{U_i(Ay_i)} \quad (\text{A.2})$$

for all i , where DU_i denotes the $(1 \times S)$ Jacobi matrix of the function U_i . Since U_i is homogeneous of degree 1, for all $x \gg 0$ and $\lambda > 0$ we have $U_i(\lambda x) = \lambda U_i(x)$. Differentiating both sides with respect to λ and setting $\lambda = 1$, we have $DU_i(x)x = U_i(x)$. Hence multiplying y_i from the right to (A.2), we get

$$q' y_i = w_i \frac{DU_i(Ay_i)Ay_i}{U_i(Ay_i)} = w_i.$$

Adding across i , since $\sum_{i=1}^I y_i = n$, we get

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

Therefore

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

so the budget constraint holds with equality. Furthermore, letting $\lambda_i = \frac{1}{w_i}$, by (A.2) we obtain $D[\log U_i(Ay_i)] = \lambda_i q'$, which is the first-order condition of the utility maximization problem (A.1) after taking the logarithm. Since $\log U_i$ is concave, y_i solves the utility maximization problem. Since $\sum_{i=1}^I y_i = n$, the asset markets clear, so $\{q, (x_i), (y_i)\}$ is a GEI. \square

Step 3. The GEI is uniquely given as the solution to the planner's problem (2.2).

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

$$\lambda_i q' = D[\log U_i(Ay_i)] = \frac{DU_i(Ay_i)A}{U_i(Ay_i)}. \quad (\text{A.3})$$

Multiplying n from the right and noting that $DU_i \gg 0$, $An \gg 0$, and $Ay_i \gg 0$ imply $U_i(Ay_i) > U_i(0) = 0$, we obtain $\lambda_i q' n > 0$. Since $\lambda_i \geq 0$, we must have $\lambda_i > 0$ and $q' n > 0$. By rescaling the price vector if necessary, we may normalize such that $q' n = 1$. Multiplying y_i to (A.3) from the right and using $DU_i(x)x = U_i(x)$ and the complementary slackness condition, we have

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

Substituting into (A.3), we obtain $q' = w_i D[\log U_i(Ay_i)]$, which is precisely (A.2), the first-order condition of the planner's problem (2.2) with Lagrange multiplier q . Since (y_i) is feasible and the objective function is strictly concave, (y_i) is the unique solution to the planner's problem. \square

A.2 Proof of Propositions 2.3 and 2.4

Let u be a general Bernoulli utility function with $u' > 0$ and $u'' < 0$. (In view of Theorem 2.2, we only need to assume $u(x) = \frac{1}{1-\gamma}x^{1-\gamma}$ or $u(x) = \log x$, but most of the following results do not depend on the particular functional form.) Suppose that there are two assets, one risky asset with gross return R and a risk-free asset with gross risk-free rate R_f . Let $R(\theta) := R\theta + R_f(1 - \theta)$ be the portfolio return, where θ is the fraction of wealth invested in the risky asset. Consider the optimal portfolio problem

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

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

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

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

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

2. Since $f'(\theta) = 0$ and $f'(0) = u'(R_f w)w(\mathbb{E}[R] - R_f)$, the result follows.
3. Dividing the first-order condition by w , we obtain $\mathbb{E}[u'(R(\theta)w)(R - R_f)] = 0$. 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 $F_\theta = \mathbb{E}[u''(R(\theta)w)(R - R_f)^2w] < 0$, it suffices to show $F_w \geq 0$. Let $\gamma(x) = -xu''(x)/u'(x) > 0$ be the relative risk aversion coefficient. Then

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

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

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

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

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

where the last equality uses the first-order condition. \square

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

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

so the less risk averse agent invests more aggressively.

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

$$\frac{d}{dx} \left(\frac{u_2'(w_2x)}{u_1'(w_1x)} \right) = \frac{w_2 u_2'' u_1' - u_2' w_1 u_1''}{(u_1')^2} = \frac{1}{x} \frac{u_2'}{u_1'} (\gamma_1(w_1x) - \gamma_2(w_2x)) > 0,$$

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

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

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

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

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

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

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

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

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

$$f_i'(\theta_i) = \sum_{s=1}^S \pi_{is} u'(R_f + X_s \theta_i) X_s = 0. \quad (\text{A.4})$$

Letting q be the stock price, since $R_s = e_s/q$ and $e_1 < \dots < e_S$, we have $X_1 < \dots < X_S$. Since $\pi_{is} > 0$ and $u' > 0$, by (A.4), it must be $X_1 < 0 < X_S$. Let $s^* = \max\{s \mid X_s < 0\}$ be the best state with negative excess returns. Clearly $1 \leq s^* < S$.

Using the definition of the likelihood ratio $\lambda_s = \pi_{1s}/\pi_{2s}$, by (A.4) we obtain

$$0 = f'_1(\theta_1) = \sum_{s=1}^S \pi_{1s} u'(R_f + X_s \theta_1) X_s = \lambda_{s^*} \sum_{s=1}^S \frac{\lambda_s}{\lambda_{s^*}} \pi_{2s} u'(R_f + X_s \theta_1) X_s.$$

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

$$\begin{aligned} 0 &= \lambda_{s^*} \sum_{s=1}^S \frac{\lambda_s}{\lambda_{s^*}} \pi_{2s} u'(R_f + X_s \theta_1) X_s \\ &< \lambda_{s^*} \sum_{s=1}^S \pi_{2s} u'(R_f + X_s \theta_1) X_s = \lambda_{s^*} f'_2(\theta_1), \end{aligned}$$

where the inequality is due to the fact that replacing $\lambda_s/\lambda_{s^*} \geq (\leq) 1$ by 1 for $s \leq (\geq) s^*$ makes the term less negative (more positive), and the inequality is strict for $s = 1$ or $s = S$. Therefore $f'_2(\theta_1) > 0$, and since f_2 is strictly concave and $f'_2(\theta_2) = 0$, we obtain $\theta_1 < \theta_2$. \square

A.3 Proof of Theorem 2.2

Step 1. The equilibrium is unique and is characterized by the planner's problem (2.6).

Proof. By the same argument as in the proof of Theorem 2.1, the Epstein-Zin utility function (2.3) with unit EIS is homogeneous of degree 1 and strictly concave. Hence by Theorem 2.1, the equilibrium is unique and is characterized by the planner's problem (2.2). Since we assumed that U_i has unit EIS, the objective function is additively separable with respect to x_{i0} . Therefore we can fix $(y_{i2}, y_{i3})_{i=1}^I$ and maximize over $(y_{i1})_{i=1}^I$. This problem is

$$\text{maximize } \sum_{i=1}^I w_i (1 - \beta_i) \log y_{i1} \quad \text{subject to } \sum_{i=1}^I y_{i1} = \alpha_0 e_0.$$

We can easily solve this problem analytically, and the solution is (2.5). Substituting this y_{i1} into (2.2), the remaining problem becomes (2.6). \square

Step 2. The price-dividend ratio is given by (2.7), and shifting wealth to a patient agent increases the price-dividend ratio.

Proof. Let P be the stock price, i.e., the value of the claim $(0, e_1, \dots, e_S)'$. Since the aggregate supply of traded shares is α_1 , the market capitalization of traded shares is $\alpha_1 P$. Since the risk-free asset is in zero net supply, the market capitalization of stocks must equal aggregate savings. Since aggregate wealth of capitalists (including $t = 0$ consumption) is $\alpha_0 e_0 + \alpha_1 P$, agent i has wealth

share w_i , and the saving rate out of wealth is β_i due to log utility, the aggregate savings is $S = \sum_{i=1}^I \beta_i w_i (\alpha_0 e_0 + \alpha_1 P)$. Setting $\alpha_1 P = S$ and solving for P , we get (2.7). From this expression it is clear that shifting wealth from a low β_i agent to a high β_i agent increases P/e_0 . \square

Step 3. The log equity premium is independent of capital income shares α_0, α_1 .

Proof. Note that the fraction of wealth agent i invests in the risk-free asset relative to *total* wealth is $\phi_{i3} = \beta_i(1 - \theta_i)$. Therefore the market clearing condition for the risk-free asset is

$$\sum_{i=1}^I w_i \beta_i (1 - \theta_i) (\alpha_0 e_0 + \alpha_1 P) = 0 \iff \sum_{i=1}^I w_i \beta_i (1 - \theta_i) = 0, \quad (\text{A.5})$$

which does not directly depend on α_0, α_1 . By homotheticity, the optimal portfolio problem (2.8) is equivalent to

$$\max_{\theta} \mathbb{E}_i [u_i(e_1 \theta + z(1 - \theta))], \quad (\text{A.6})$$

where $z = PR_f$. Since (A.6) does not directly depend on α_0, α_1 , the value of z that makes the market clearing condition (A.5) holds does not depend on α_0, α_1 . By the definition of the log equity premium, we have

$$\mu = \pi'(\log R - \log R_f) = \pi'(\log(e_1/P) - \log R_f) = \pi' \log e_1 - \log z, \quad (\text{A.7})$$

so the log equity premium is independent of α_0, α_1 . \square

Step 4. Shifting wealth from a high ϕ_{i3} agent to a low ϕ_{i3} agent reduces the log equity premium.

Proof. Suppose that in the initial equilibrium we have $\beta_1(1 - \theta_1) > \beta_2(1 - \theta_2)$, so agent 1 is the natural bondholder, and we transfer some wealth from agent 1 to 2. Since $z = PR_f$ is the only relevant parameter in the optimal portfolio problem (A.6), if z is unchanged after the wealth transfer, then all agents choose their original portfolios. Letting P' be the stock price and w'_i the wealth share of agent i after the transfer, by assumption we have $w'_1 < w_1$, $w'_2 > w_2$, and $w'_i = w_i$ for $i > 2$. Then the new aggregate demand of the risk-free asset is

$$\sum_{i=1}^I w'_i \beta_i (1 - \theta_i) (\alpha_0 e_0 + \alpha_1 P') = (\alpha_0 e_0 + \alpha_1 P') \sum_{i=1}^I w'_i \beta_i (1 - \theta_i),$$

which has the same sign as $\sum_{i=1}^I w'_i \beta_i (1 - \theta_i)$. But by (A.5), we have

$$\begin{aligned} \sum_{i=1}^I w'_i \beta_i (1 - \theta_i) &= \sum_{i=1}^I w'_i \beta_i (1 - \theta_i) - \sum_{i=1}^I w_i \beta_i (1 - \theta_i) \\ &= \beta_1(1 - \theta_1)(w'_1 - w_1) + \beta_2(1 - \theta_2)(w'_2 - w_2) \\ &= \epsilon(\beta_2(1 - \theta_2) - \beta_1(1 - \theta_1)), \end{aligned}$$

where $\epsilon := w_1 - w'_1 = w'_2 - w_2 > 0$ because $w'_1 + w'_2 = w_1 + w_2$. Therefore if we shift wealth from an agent who invests more in the risk-free asset (high

$\phi_{i3} = \beta_i(1 - \theta_i)$) to an agent who invests less (low $\phi_{i3} = \beta_i(1 - \theta_i)$), it will result in an excess supply of risk-free assets.

Since θ_i solves (A.6), applying Lemma A.1 for $R = e_1$ and $R_f = z$, for large enough z we have $\theta_i < 0$ for all i . Therefore there is an excess demand of risk-free assets. Hence by the intermediate value theorem (continuity trivially holds by the maximum theorem), in the new equilibrium (which is unique) $z = PR_f$ must increase. Therefore the log equity premium (A.7) must decrease. \square

B Infinite horizon model with estate tax

In this appendix we present an infinite horizon version of the model in Section 2 with overlapping generations and an estate tax.

B.1 Model

Agents There are a continuum of agents. Agents have Epstein-Zin preferences and can be either of the two types indexed by $i = A, B$. The discount factor, relative risk aversion, and elasticity of intertemporal substitution of each type are denoted by $\beta_i \in (0, 1)$, $\gamma_i > 0$, and $\varepsilon_i > 0$, respectively.

To ensure stationarity, we assume that agents die with probability $\delta > 0$ each period. The wealth of a deceased agent is passed on to its child after paying the estate tax. Fraction ν_i of newborn agents are of type i , where $\nu_A + \nu_B = 1$, independent of the type of their parents. Furthermore, there is an exogenous influx of agents (e.g., new entrepreneurs, immigrants) to the economy, among which fraction ν_i is of type i . The revenue from the estate tax is distributed equally among these new agents without parents (“orphans”).²⁵

Estate tax The estate tax rate follows an exogenous finite-state Markov chain. Let these states be indexed by $s = 1, \dots, S$ and $P = (p_{ss'})$ be the transition probability matrix, which we assume to be irreducible. Let τ_{is} be the estate tax rate in state s applied to type i . The estate tax rate announced at time t , which is τ_{is_t} , is applied to the wealth of type i agents who pass away between time t and $t + 1$.

Endowment growth Aggregate dividend growth takes finitely many values and is independent and identically distributed (i.i.d.) over time. Let g_j be the log aggregate dividend growth rate in state j and p_j be its probability, where $j = 1, \dots, J$.

Financial structure There are two financial assets, a stock (claim to aggregate dividend) and a bond (risk-free asset in zero net supply). Furthermore, there are perfectly competitive insurance companies that offer life insurance.

²⁵There is no need to specify the exogenous population growth rate since it does not play any role. The reader may wonder why we need exogenous population growth in addition to overlapping generations. The main reason is we need to do something with the estate tax revenue. Because redistributing the estate tax revenue across existing agents would kill the homogeneity (and hence tractability) of the model, we have decided to introduce exogenous population growth.

We assume that agents care about their children's utility as their own, so they buy life insurance to just cover the estate tax payment. Therefore, if R_f is the gross risk-free rate at a particular point in time and τ_{is} is the estate tax rate, the effective risk-free rate that type i agents face is $\tilde{R}_f = \frac{R_f}{1+\delta\tau_{is}}$. (The same applies to the stock returns.)²⁶

B.2 Equilibrium

Because there are only two financial assets but JS contingencies (J states for aggregate dividend growth and S states for estate tax), markets are incomplete. The state variables are the wealth share of type A agents, which we denote by $x \in [0, 1]$, and the exogenous estate tax state $s \in \{1, \dots, S\}$. (The dividend growth state $j \in \{1, \dots, J\}$ is not a state variable due to the i.i.d. assumption.)

Let w be the wealth of a typical type i agent, $R_f(x, s)$ be the gross risk-free rate given the current state (x, s) , and $R(x, s, s', j')$ be the gross stock return when the current state is (x, s) and the next period's estate tax and dividend growth states are (s', j') . Then the budget constraint is

$$w' = \frac{1}{1 + \delta\tau_{is}} (R(x, s, s', j')\theta + R_f(x, s)(1 - \theta))(w - c), \quad (\text{B.1})$$

where θ is the fraction of wealth invested in the stock and c is consumption. Note that the next period's wealth is the same regardless of whether the agent passes away or not because the estate tax payment is covered by life insurance. Letting $V_i(w, x, s)$ be the value function of a type i agent, it satisfies the Bellman equation

$$V_i(w, x, s) = \max_{c, \theta} \left((1 - \beta_i)c^{1-1/\varepsilon_i} + \beta_i \mathbf{E} [V_i(w', x', s')^{1-\gamma_i} | x, s]^{\frac{1-1/\varepsilon_i}{1-\gamma_i}} \right)^{\frac{1}{1-1/\varepsilon_i}}, \quad (\text{B.2})$$

where the next period's wealth satisfies the budget constraint (B.1).

A recursive equilibrium is defined by a price-dividend ratio $q(x, s)$, gross stock returns $R(x, s, s', j')$, gross risk-free rate $R_f(x, s)$, and agents' value functions and optimal consumption-portfolio rule such that (i) the value functions satisfy the Bellman equation (B.2) and the consumption-portfolio rules are the argmax, (ii) markets for consumption, stock, and risk-free asset clear, (iii) the law of motion for the state variables is consistent with individual choice, and (iv) the gross stock returns and price-dividend ratio are consistent.

B.3 How to solve for equilibrium

Since agents have homothetic preferences and the budget constraint (B.1) is homogeneous of degree 1 in (w, c) , we can guess that the value function takes the form

$$V_i(w, x, s) = a_i(x, s)w \quad (\text{B.3})$$

²⁶To see why this is the correct insurance premium, suppose that an agent has savings w and portfolio return R . Then the next period's wealth is Rw with probability $1 - \delta$ (if he survives) and $(1 - \tau)Rw$ with probability δ (if he passes away, where τ is the estate tax rate). The agent wants to cover the loss τRw by purchasing life insurance. If the insurance company charges premium a , it gets Raw at the beginning of the next period, which is used to finance the insurance payment $\delta\tau Rw$. Therefore $a = \delta\tau$.

for some coefficient $a_i(x, s) > 0$. Using the Bellman equation (B.2), we can solve for the optimal consumption rule for each agent analytically. Consider an agent with parameters $(\beta, \gamma, \varepsilon)$. Let a, a' be the coefficients of the value function (B.3) for the current and subsequent period and R be the gross return on wealth, fixing the portfolio. Then (B.2) becomes

$$aw = \max_c \left((1 - \beta)c^{1-1/\varepsilon} + \beta \mathbb{E}[(a'R(w - c))^{1-\gamma}]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}.$$

Letting $m = c/w$ be the propensity to consume out of wealth and

$$\rho = \mathbb{E}[(a'R)^{1-\gamma}]^{\frac{1}{1-\gamma}}, \quad (\text{B.4})$$

the above equation further simplifies to

$$a = \max_m \left((1 - \beta)m^{1-1/\varepsilon} + \beta\rho^{1-1/\varepsilon}(1 - m)^{1-1/\varepsilon} \right)^{\frac{1}{1-1/\varepsilon}}.$$

The maximization over m is equivalent to

$$\max_{m \in (0,1)} \frac{1}{1-1/\varepsilon} \left((1 - \beta)m^{1-1/\varepsilon} + \beta\rho^{1-1/\varepsilon}(1 - m)^{1-1/\varepsilon} \right).$$

The first-order condition is

$$(1 - \beta)m^{-1/\varepsilon} = \beta\rho^{1-1/\varepsilon}(1 - m)^{-1/\varepsilon} \iff m = \frac{(1 - \beta)^\varepsilon}{(1 - \beta)^\varepsilon + \beta^\varepsilon\rho^{\varepsilon-1}}. \quad (\text{B.5})$$

Using (B.5) and the Bellman equation, we obtain

$$a^{1-1/\varepsilon} = (1 - \beta)m^{1-1/\varepsilon}(m + (1 - m)) \iff m = (1 - \beta)^\varepsilon a^{1-\varepsilon}. \quad (\text{B.6})$$

Therefore the optimal consumption rate of type i in state (x, s) is

$$m_i(x, s) = (1 - \beta_i)^{\varepsilon_i} a_i(x, s)^{1-\varepsilon_i}. \quad (\text{B.7})$$

Note that comparing (B.5) and (B.6), we obtain

$$a = \left((1 - \beta)^\varepsilon + \beta^\varepsilon\rho^{\varepsilon-1} \right)^{\frac{1}{\varepsilon-1}}.^{27} \quad (\text{B.8})$$

We can solve for the equilibrium using the projection method (Judd, 1992; Pohl et al., 2018). The policy functions are (i) coefficients of the value function $a_i(x, s)$ ($i = A, B$), (ii) portfolios $\theta_i(x, s)$ ($i = A, B$, where θ is the fraction of savings invested in stocks), (iii) gross risk-free rate $R_f(x, s)$, and (iv) gross stock return $R(x, s, s', j')$. If we use $(N - 1)$ -degree Chebyshev polynomials to approximate each, then we have

$$N(2S + 2S + S + S^2J) = NS(5 + SJ)$$

unknown coefficients. To determine these coefficients, we need the same number of equations.

The number of equations are: (i) $2NS$ consistency conditions for the coefficient of value function (2 agents, S estate tax states, and N Chebyshev nodes), (ii) $2NS$ first-order conditions for portfolio choice (2 agents, S estate tax states, and N Chebyshev nodes), (iii) NS market clearing conditions for bonds, and (iv) NS^2J consistency conditions for stock returns, which we describe below in details.

²⁷If $\varepsilon = 1$, we can take the limit of (B.8) as $\varepsilon \rightarrow 1$ to obtain $a = (1 - \beta)^{(1-\beta)}(\beta\rho)^\beta$.

1. The consistency condition for the coefficient of value function is (B.8), where ρ is calculated using (B.4) with the return on wealth

$$R_i(\theta; x, s) := \frac{1}{1 + \delta\tau_{is}} (R(x, s, s', j')\theta + R_f(x, s)(1 - \theta)) \quad (\text{B.9})$$

with $\theta = \theta_i(x, s)$.

2. Suppressing the i subscript, since the optimal portfolio problem reduces to

$$\max_{\theta} \frac{1}{1 - \gamma} \text{E} [(a')^{1-\gamma} R(\theta; x, s)^{1-\gamma} | x, s],$$

the first-order condition is

$$\text{E} [(a')^{1-\gamma} R(\theta(x, s); x, s)^{-\gamma} (R(x, s, s', j') - R_f(x, s)) | x, s] = 0. \quad (\text{B.10})$$

3. Since type i 's saving rate and bond portfolio are $1 - m_i$ and $1 - \theta_i$, and the wealth share of type A agents is x , the bond market clearing condition is

$$(1 - m_A)(1 - \theta_A)x + (1 - m_B)(1 - \theta_B)(1 - x) = 0, \quad (\text{B.11})$$

where m_i is given by (B.6) and $\theta_i = \theta_i(x, s)$.

4. Finally we derive the consistency condition for the stock returns. Letting W_i be the aggregate wealth held by type i agents at the beginning of the period, by commodity market clearing we obtain

$$D = \sum_{i=A,B} m_i W_i,$$

where D is aggregate dividend. Since the risk-free asset is in zero net supply, all aggregate savings must be in stocks. Hence its price must be

$$P = \sum_{i=A,B} (1 - m_i) W_i.$$

Therefore the price-dividend ratio is

$$q(x, s) = \frac{P}{D} = \frac{(1 - m_A)x + (1 - m_B)(1 - x)}{m_A x + m_B(1 - x)}. \quad (\text{B.12})$$

Assuming next period's type A wealth share x' is known, we can compute the gross stock return as

$$R(x, s, s', j') = \frac{P' + D'}{P} = \frac{P'/D' + 1}{P/D} \frac{D'}{D} = \frac{q(x', s') + 1}{q(x, s)} e^{g_{j'}}, \quad (\text{B.13})$$

where the price-dividend ratio $q(x, s)$ is given by (B.12). (B.13) is the consistency condition for the stock return.

Since the next period's type A wealth share x' appears in (B.13) and also implicitly in $a' = a(x', s')$ in (B.4), to close the solution algorithm we need to derive the equation of motion for the type A wealth share x . By the budget

constraint (B.1), the optimal consumption rule (B.6), and the return on wealth (B.9), the wealth of a typical type i agent evolves according to

$$w'_i = R_i(\theta_i; x, s)(1 - m_i)w_i.$$

Note that this w'_i is the same regardless of whether the agent passes away or not because the estate tax payment is covered by life insurance. Since fraction δ of agents die, children's type is independent of parents' type, and the estate tax revenue is distributed across orphans, it follows that

$$\begin{aligned} W'_i &= \underbrace{(1 - \delta)R_i(1 - m_i)W_i}_{\text{Aggregate wealth of type } i \text{ agents who survived}} \\ &+ \underbrace{\nu_i \delta \sum_{i=A,B} R_i(1 - m_i)W_i}_{\text{Aggregate wealth of type } i \text{ newborn agents with parents}} \\ &+ \underbrace{\nu_i \delta \sum_{i=A,B} \tau_{is} R_i(1 - m_i)W_i}_{\text{Aggregate wealth of type } i \text{ newborn agents without parents}} \\ &= (1 - \delta)R_i(1 - m_i)W_i + \nu_i \delta \sum_{i=A,B} (1 + \tau_{is})R_i(1 - m_i)W_i, \end{aligned}$$

where we have abbreviated as $R_i = R_i(\theta_i; x, s)$. Putting all the pieces together, we obtain

$$x' = \frac{(1 - \delta)R_A(1 - m_A)x + \nu_A \delta \sum_{i=A,B} (1 + \tau_{is})R_i(1 - m_i)x_i}{\sum_{i=A,B} (1 + \delta \tau_{is})R_i(1 - m_i)x_i}, \quad (\text{B.14})$$

where $x_A = x$ and $x_B = 1 - x$.

We can now apply the projection method to numerically solve for the equilibrium, as follows.

1. Choose a degree N of Chebyshev polynomial approximation.
2. For any vector of coefficients

$$\mathbf{c} = \left((c_{isn}^a, c_{isn}^\theta)_{i=A,B}, (c_{sn}^{R_f}), (c_{ss'j'n}^R)_{s'=1}^S \right)_{s=1}^S \in \mathbb{R}^{NS(5+SJ)},$$

approximate the policy functions as

$$\log a_i(x, s) = \sum_{n=0}^{N-1} c_{isn}^a T_n(2x - 1),$$

where $T_n(\cdot)$ is the n -degree Chebyshev polynomial and c_{isn}^a is its coefficient in estate tax state s . (Here we write $T_n(2x - 1)$ instead of $T_n(x)$ because the state space for x is $[0, 1]$, whereas Chebyshev polynomials are defined on $[-1, 1]$. The mapping $x \mapsto 2x - 1$ maps $[0, 1]$ into $[-1, 1]$.) The same applies for $\log \theta_i$, $\log R_f$, and $\log R$.

3. Define the residual of the equilibrium conditions $F : \mathbb{R}^{NS(5+SJ)} \rightarrow \mathbb{R}^{NS(5+SJ)}$ by stacking the left-hand side minus the right-hand side of the consistency condition for the Bellman equation (B.8), the portfolio first-order condition (B.10), the bond market clearing condition (B.11), and the consistency condition for the stock returns (B.13), where the type A wealth share is evaluated at the points corresponding to the roots of the N -degree Chebyshev polynomial, so $x_n = \frac{1}{2} \left(1 + \cos \frac{2n-1}{2N} \pi \right)$.
4. Find coefficients \mathbf{c} such that $F(\mathbf{c}) = 0$.

B.4 Calibration

We calibrate the model at annual frequency. We obtain the real stock prices, dividend, and risk-free rates for the period for the period 1871–2012 from Robert Shiller’s spreadsheet.²⁸ We assume that log dividend growth is independent and identically distributed as a Gaussian mixture distribution with two components. Using maximum likelihood, the estimated mean, standard deviation, and proportion of each component are $\mu = (-0.0264, 0.0263)$, $\sigma = (0.2071, 0.0616)$, and $p = (0.245, 0.755)$. Figure 7 shows the histogram and the fitted density.

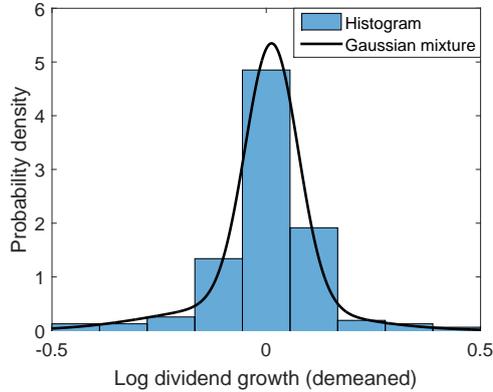


Figure 7: Histogram and fitted Gaussian mixture density of log dividend growth.

To make the solution algorithm simple, we discretize the Gaussian mixture density by a three-point distribution using the Gaussian quadrature for Gaussian mixtures. Thus we have $J = 3$, $g = (g_j) = (-0.3584, 0.0094, 0.2779)$, and $p = (p_j) = (0.0549, 0.8552, 0.0899)$. For the preference parameters, we assume that agents have unit EIS ($\varepsilon_A = \varepsilon_B = 1$) so that the infinite horizon model is as close as possible to the two period model in Section 2. Note that in this case (B.7) implies $m_i = 1 - \beta_i$, so the price-dividend ratio in (B.12) can be computed explicitly. Since this formula is essentially identical to (2.7), as in Theorem 2.2, shifting wealth from an impatient to a patient agent increases the

²⁸<http://www.econ.yale.edu/~shiller/data.htm>

price-dividend ratio. Finally, for the discount factor and relative risk aversion, we set $(\beta_A, \beta_B) = (0.985, 0.94)$ and $(\gamma_A, \gamma_B) = (1, 5)$ so that the unconditional log equity premium, stock volatility, and the price-dividend ratio are roughly the same as in the data.

We set the death probability to $\delta = 1/40 = 0.025$ so that inheritance occurs on average every 40 years. Since we model type A as the rich stock holder, we set $\nu_A = 1/10 = 0.1$ so that ν_A roughly corresponds to the fraction of top 1% agents among the top 10%. Since only very rich households pay estate taxes, we assume that the estate tax rate for type B is zero, and that the estate tax for type A take two values, low ($\tau_{AL} = 0.2$) and high ($\tau_{AH} = 0.8$). The estate tax state switches between the two state with probability 0.05, so the transition probability matrix is $P = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$, which implies that the same tax rate applies for an average of $1/0.05 = 20$ years.

Having chosen all parameter values, we solve for the equilibrium using the projection method as described above. Figure 8 shows the equity portfolios, equity premium, risk-free rate, and price-dividend ratio when we change the type A wealth share x as well as the stationary distribution of x . Table 15 shows some asset pricing moments in the data and the model.

Table 15: Asset pricing moments.

Moment	Data	Model
Log equity premium (%)	3.84	3.83
Volatility of returns (%)	17.4	16.6
Average real interest rate (%)	2.49	1.32
Average price-dividend ratio	26.4	26.9

B.5 Monte Carlo regression analysis

In this section, we conduct a Monte Carlo experiment to see how different variables in the model predict excess returns at varying sample sizes. Given the global numerical solution, simulating time series from the model is straightforward. First, we create an evenly-spaced 101 point grid on $[0, 1]$ for the type A wealth share x . Then we simulate an estate tax Markov chain and an i.i.d. path for dividend growth and calculate the evolution of x using (B.14). All other series, such as optimal portfolios, are calculated from the policy functions, which depend only on x and the estate tax state. Off-grid values are interpolated with piecewise cubic Hermite polynomials. We use 10,000 simulations and consider four different sample sizes (50, 100, 500, and 1,000 years), with a burn-in period of 500 years.

To compare our simulations with our empirical analysis, we construct a model version of KGR. Let n_{it}^s be the number of shares of stock held by type i after trading in time t . Similarly, let n_{it}^b be the number of bonds held. Note that in equilibrium $\sum_{i=A,B} n_{it}^s = 1$ and $\sum_{i=A,B} n_{it}^b = 0$. Using the fact that total wealth is $P_t + D_t$, these quantities can be computed from the simulated series as $n_{it}^s = \theta_{it}(1 - m_{it})x_{it}(P_t + D_t)/P_t$ and $n_{it}^b = (1 - \theta_{it})(1 - m_{it})x_{it}(P_t + D_t)R_{f,t}$. Time t per share stock income excluding capital gains is D_t , while per bond

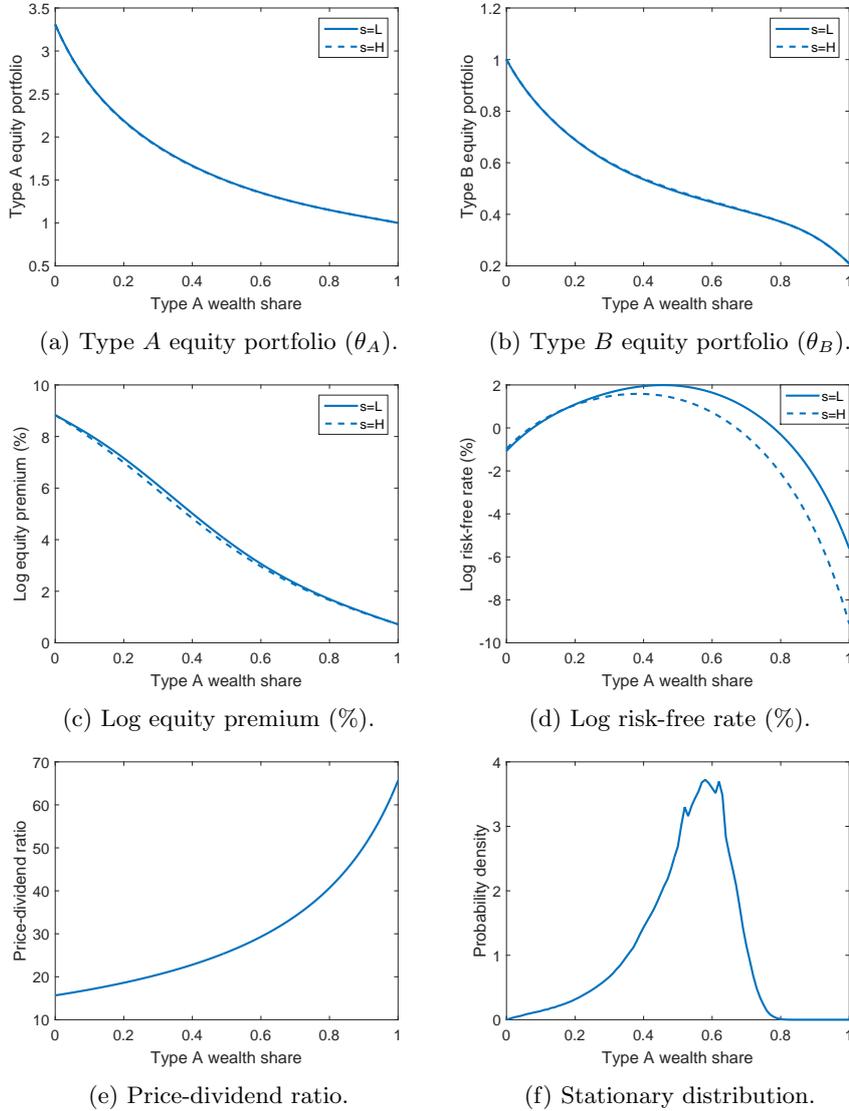


Figure 8: Solution to the infinite horizon model.

interest income is $(R_{f,t-1} - 1)/R_{f,t-1}$.²⁹ Therefore, type i 's income at t is, excluding capital gains,

$$Y_{it} = n_{i,t-1}^s D_t + n_{i,t-1}^b (R_{f,t-1} - 1)/R_{f,t-1}.$$

To construct KGR in the model, we need measures of *realized* capital gains and some corresponding cost basis, which are irrelevant for our agents since they care about their total wealth and not its composition or history. Additionally, annual frequency realized capital gains from the data are affected by intraperiod trades that are beyond the scope of our model. Indeed, in the data investors

²⁹We are implicitly assuming that $t - 1$ bond shares trade at price $1/R_{f,t-1}$ and pay 1 at t : $1/R_{f,t-1}$ in principal and $1 - 1/R_{f,t-1}$ in interest.

may begin and end a year with similar stock portfolios while over the course of the year realizing substantial capital gains/losses for liquidity, tax, or other purposes. Therefore, since the model cost basis is arbitrary and since gross trading in the data is much richer than is the effectively net trading that comes from the model, we adopt a parsimonious method of accounting for realized capital gains in our simulations.

Specifically, we suppose that trading activity entails a fraction λ of the stock market being sold each year so that the aggregate cost basis and realized capital gains evolve according to

$$\text{CB}_t = (1 - \lambda)\text{CB}_{t-1} + \lambda P_{t-1}, \quad (\text{B.15a})$$

$$\text{RCG}_t = \lambda(P_t - \text{CB}_t), \quad (\text{B.15b})$$

respectively. These realized capital gains are attributed to agent i according to his previous stock holdings so that $\text{RCG}_{it} = n_{i,t-1}^s \text{RCG}_t$. Beyond simplicity, an additional advantage of this method is that RCG_t has an easily computed empirical counterpart, which we explore in Appendix C.

As we argued in Section 3.1, top income shares in the data (10% and above) appear to follow a common U-shaped trend, perhaps driven by highly persistent redistribution between financial market participant capitalists and workers. Since our model so far is one of capitalists, to make simulated income inequality measures and KGR comparable with the data, we introduce non-capitalist (“laborer”) income $Y_{L,t}$ and assume that the capitalists share of income,

$$\alpha_t = \frac{Y_{A,t} + Y_{B,t}}{Y_{A,t} + Y_{B,t} + Y_{L,t}},$$

follows an exogenous, slowly moving process. As in Theorem 2.2 from the two period model, α_t is irrelevant for asset prices, which are determined by the capitalists. We calibrate α_t to behave like $\text{top}(10)^{\text{excg}}$, the empirical Piketty-Saez top 10% income share excluding realized capital gains. In particular, we fit an AR(1) process to

$$z_t = \log \frac{\text{top}(10)_t^{\text{excg}}}{1 - \text{top}(10)_t^{\text{excg}}},$$

simulate paths for z_t , and set $\alpha_t = 1/(1 + e^{-z_t})$. This procedure ensures that the α_t simulations stay on $(0, 1)$ and have persistence and variance similar to those of $\text{top}(10)^{\text{excg}}$.

Combining these elements, we can define the type A income share excluding realized capital gains, the income share including realized capital gains, and KGR as

$$\text{top}(A)_t^{\text{excg}} = \frac{Y_{A,t}}{Y_{A,t} + Y_{B,t} + Y_{L,t}} = \alpha_t \frac{Y_{A,t}}{Y_{A,t} + Y_{B,t}} \quad (\text{B.16a})$$

$$\begin{aligned} \text{top}(A)_t &= \frac{Y_{A,t} + \text{RCG}_{it}}{Y_{A,t} + Y_{B,t} + Y_{L,t} + \text{RCG}_t} \\ &= \alpha_t \frac{Y_{A,t} + \text{RCG}_{it}}{Y_{A,t} + Y_{B,t} + \alpha_t \text{RCG}_t} \end{aligned} \quad (\text{B.16b})$$

$$\text{KGR}(A)_t = \frac{\text{top}(A)_t - \text{top}(A)_t^{\text{excg}}}{1 - \text{top}(A)_t}. \quad (\text{B.16c})$$

As we see in (B.16a)–(B.16c), due to our definitions of cost basis and realized capital gains, $\text{top}(A)$ is not restricted to be between 0 and 1, which can generate spurious KGR outliers because the denominator of (B.16c) is $1 - \text{top}(A)_t$. In our Monte Carlo analysis, we throw out simulations in which $\text{top}(A)$ is ever less than 0 or greater than 1 or the denominator of (B.16b) is less than 0, although for sufficiently small λ this happens only very rarely within 1,500 years (the burn-in period plus the maximum sample size). However, if λ is too small, realized capital gains become unrealistically low relative to overall capital income. Below we consider two different values for λ . With $\lambda = 0.005$, only 2% of simulations are discarded, while average RCG_t/D_t is 9%, compared with 26% in the Saez and Zucman (2016) spreadsheet (measured as total net realized capital gains divided by capital and business income excluding capital gains). With $\lambda = 0.01$, 14% of simulations are discarded, while average RCG_t/D_t rises to 14%. However, both 9% and 14% are within a standard deviation (20%) of the empirical mean. Furthermore, the two λ 's yield similar results otherwise, so we conclude our below findings are driven neither by unrealistic levels for RCG_t nor by our discarding procedure.

Table 16 shows model correlations calculated from averaging across simulations. $\text{KGR}(A)$, type A 's income share including realized capital gains, and total realized capital gains scaled by dividends are all substantially correlated with the type A wealth share x . Since x inversely forecasts excess returns (Figure 8), one would expect these other measures to potentially proxy for wealth inequality in forecasting the equity premium.³⁰

In Table 17, with $T = 1000$ we see that x , KGR, $\text{top}(A)$, and RCG/D all inversely forecast excess returns, although income inequality underperforms in terms of R^2 and power (at the 10% level), likely due to highly persistent movements in the capitalist income share α_t . KGR is slightly underpowered (the null of no effect is rejected at the 10% level in 83–86% of simulations, depending on λ). As the sample size falls, the performances of KGR and RCG/D improve relative to x . With $T = 50, 100$, KGR and RCG/D actually slightly outperform x in terms of R^2 and power to capture the relationship between inequality and excess returns.

In summary, according to our calibrated model, KGR and RCG/D can serve as proxies for x in testing the relationship between inequality and the equity premium. And at smaller sample sizes, these proxies are even more likely than is wealth inequality to pick up the relationship (although all forecasters are underpowered for low T in our simulations). Surprisingly, comparing Table 17 with the empirical counterpart (Table 2), at $T = 100$ the regression coefficient and R^2 for KGR are quite similar in the data and model.

C The role of mechanical realized capital gains

We saw in Section B.5 that our calibrated model predicts (i) aggregate realized capital gains inversely forecast excess returns and (ii) realized capital gains are highly correlated with KGR (Table 16). We show in this section that these

³⁰The estate tax is also a state variable and in theory also forecasts returns, but, as we see in Figure 8, the direct relationship between the estate tax and the equity premium is quantitatively minuscule (the price-dividend ratio is theoretically independent of estate tax due to the unit EIS assumption).

Table 16: Model correlations

	$\lambda = 0.005$		$\lambda = 0.01$	
	correlation with x	correlation with $\text{KGR}(A)$	correlation with x	correlation with $\text{KGR}(A)$
x	1	0.86	1	0.89
$\text{KGR}(A)$	0.86	1	0.89	1
$\text{top}(A)$	0.71	0.81	0.75	0.83
$\frac{\text{RCG}}{D}$	0.88	0.93	0.91	0.95

Note: The table shows 1000 year model correlations averaged over 10,000 simulations. x is type A 's wealth share, KGR and income inequality $\text{top}(A)$ are defined by (B.16c) and (B.16b), and RCG is defined by (B.15b). λ is the fraction of the stock market annually sold to realize capital gains.

Table 17: Model regressions of one year excess stock returns on various predictors

	Dependent variable: t to $t + 1$ excess market return						
	$\lambda = 0.005$				$\lambda = 0.01$		
	x	Regressors (t)		$\frac{\text{RCG}}{D}$	Regressors (t)		$\frac{\text{RCG}}{D}$
<u>$T = 50$</u>							
ave. coeff.	-0.41	-3.81	-0.65	-1.45	-1.78	-0.62	-0.60
power	0.31	0.37	0.26	0.39	0.36	0.29	0.38
ave. R^2	0.04	0.05	0.04	0.05	0.05	0.04	0.05
<u>$T = 100$</u>							
ave. coeff.	-0.25	-2.26	-0.39	-0.87	-1.06	-0.36	-0.36
power	0.33	0.39	0.28	0.44	0.40	0.30	0.43
ave. R^2	0.02	0.03	0.02	0.03	0.03	0.02	0.03
<u>$T = 500$</u>							
ave. coeff.	-0.12	-0.95	-0.16	-0.36	-0.46	-0.15	-0.15
power	0.66	0.63	0.43	0.72	0.64	0.49	0.73
ave. R^2	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<u>$T = 1000$</u>							
ave. coeff.	-0.11	-0.77	-0.13	-0.28	-0.39	-0.13	-0.13
power	0.90	0.83	0.63	0.90	0.86	0.70	0.92
ave. R^2	0.01	0.01	0.00	0.01	.01	.01	.01

Note: The table shows results from regressing model simulated log excess returns on a constant and either lagged x (type A wealth share), $\text{KGR}(A)$, $\text{top}(A)$ (type A income share including realized capital gains), or RCG/D (realized capital gains divided by dividends) at different sample sizes ($T = 50, 100, 500, 1000$). KGR and $\text{top}(A)$ are defined by (B.16c) and (B.16b), and RCG is defined by (B.15b). λ is the fraction of the stock market annually sold to realize capital gains. "ave. coeff." and "ave. R^2 " are the OLS coefficient and R^2 averaged across 10,000 simulations. "power" is the fraction of simulations in which the coefficient is significant at the 10% level.

implications are borne out in the data. In the model, realized capital gains (as well as KGR) forecast excess returns *only* because they proxy for wealth

inequality. However, one might be concerned that in the data realized capital gains forecast returns for reasons unrelated to inequality and that KGR simply proxies for aggregate realized capital gains. While this hypothesis is difficult to test in the data since KGR is substantially correlated with realized capital gains (as in our theory), in this section we also perform a horse race between KGR and purely mechanical measures of aggregate realized capital gains. In particular, we show that realized capital gains do not clearly outperform KGR and that the distribution of capital gains across income groups appears to matter beyond its aggregate level (Table 18). We also find that while both KGR and mechanical realized capital gains are significantly associated with rising wealth inequality, conditional on KGR, realized capital gains are not significantly related to changing wealth inequality (Table 19).

We define realized capital gains according to (B.15a) and (B.15b) but consider two different versions. First, we define RCG_t^I/E_t by letting P_t and E_t be the annual stock index and earnings from the Welch and Goyal (2008) spreadsheet. This version is the one closest to RCG/D in the model simulations from Section B.5. Second, we define RCG_t^W/GDP_t by letting P_t and GDP_t be the household wealth measure HNOCEA and nominal GDP from FRED. While RCG_t^I/E_t spans our sample, RCG_t^W/GDP_t is only available post-1947. For presentation, we scale each series to have the standard deviation of KGR. We show results for $\lambda = 0.1$. The results for $\lambda = 0.005$ (our simulation value in Section B.5) are similar, but RCG_t^W/GDP_t becomes very persistent with low λ . In all cases, we set $CB_0 = P_0$. Additionally, we let $RCG(x)_t$ denote the realized capital gains of the top $x\%$ by income, divided by total income including capital gains (computed from the Saez and Zucman (2016) spreadsheets). Both RCG_t^I/E_t and RCG_t^W/GDP_t are highly correlated with KGR, as predicted by our theory. For example, the correlations between $KGR(1)$ and RCG_t^I/E_t and RCG_t^W/GDP_t are, respectively, 0.40 and 0.50. Also predicted by our theory, RCG_t^I/E_t and RCG_t^W/GDP_t both inversely forecast excess returns (the coefficients are significant at the 1% level with Newey-West standard errors).

In columns (1) and (2) of Table 18, we see that controlling for KGR, RCG_t^I/E_t does not significantly forecast excess returns.³¹ When we restrict the sample to post-1947, RCG_t^W/GDP_t is significant while KGR is not (columns (3) and (4)). However, the distribution of realized capital gains across income groups still matters: in columns (5) and (6) the realized capital gains of the top 1% and 10% inversely forecast returns, even when controlling for overall mechanical realized capital gains.

In summary, while KGR is correlated with mechanical realized capital gains (as predicted by theory), there is not strong evidence that KGR is irrelevant once controlling for mechanical capital gains. KGR performs worse post-1947 with the household wealth definition of mechanical capital gains, but even in that case the distribution of realized capital gains across income groups still forecasts excess returns. While mechanical realized capital gains are, like KGR, a potential proxy for wealth inequality, KGR is our preferred forecaster for several reasons. First, KGR is much more strongly associated with changes in wealth inequality (according to the Saez and Zucman (2016) 1% measure), as we see in Table 19. Second, KGR is readily computed from the Piketty and Saez (2003) data, which are frequently updated, while computing mechanical

³¹To save space, we omit the results for $KGR(0.1)$, which are similar to those for $KGR(1)$.

Table 18: Regressions of one year excess stock market returns on KGR and mechanical realized capital gains

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	12.12 (2.94)	14.62 (3.28)	17.83 (4.23)	18.56 (4.19)	19.20 (4.24)	20.63 (4.08)
KGR(1)	-1.56 (1.21)		-1.64 (1.35)			
KGR(10)		-1.98** (1.00)		-1.70 (1.07)		
$\frac{RCG^I}{E}$	-1.90 (1.23)	-1.59 (1.20)				
$\frac{RCG^W}{GDP}$			-2.72** (1.19)	-2.42* (1.23)	-2.81** (1.28)	-2.56* (1.29)
RCG(1)					-1.85* (1.05)	
RCG(10)						-1.90** (0.82)
Sample	1913- -2015	1917- -2015	1947- -2015	1947- -2015	1947- -2012	1947- -2012
R^2	0.07	0.09	0.14	0.15	0.16	0.17

Note: Newey-West standard errors in parentheses (4 lags). ***, **, and * indicate significance at 1%, 5%, and 10% levels (suppressed for constants). $KGR(x)$ is the proxy for top $x\%$ capital inequality defined by (3.4). $RCG(x)$ is the realized capital gains of the top $x\%$ by income, divided by total income (including realized capital gains). RCG^I/E is mechanical realized capital gains (stock index definition), divided by earnings. RCG^W/GDP is mechanical realized capital gains (household wealth definition), divided by GDP.

Table 19: Regressions of the change in top 1% wealth share on KGR and mechanical realized capital gains

Regressors (t)	Dependent Variable: $t - 1$ to t Change in Top 1% Wealth Share			
	(1)	(2)	(3)	(4)
Constant	-0.84 (0.21)	-0.49 (0.18)	-0.86 (0.24)	-0.46 (0.20)
KGR(1)	0.36*** (0.11)	0.25** (0.10)		
KGR(10)			0.26*** (0.10)	0.18** (0.09)
$\frac{RCG^I}{E}$	0.03 (0.09)		0.05 (0.10)	
$\frac{RCG^W}{GDP}$		-0.00 (0.08)		-0.00 (0.09)
Sample	1913- -2012	1947- -2012	1917- -2012	1947- -2012
R^2	0.14	0.19	0.11	0.15

Note: see caption of Table 18 for explanation. The top 1% wealth share is from Saez and Zucman (2016).

realized capital gains entails choosing the parameter λ and an initial cost basis.

D Robustness of predictability

As described in Section 3.2, Tables 20 and 21 show that the inverse relationship between the capital wealth and income inequality and subsequent excess returns also holds with the KGR(10) and KGR(0.1) series, although the result is slightly stronger for the 10% series and weaker for the 0.1% series (due to large standard errors in multiple regressions).

Table 20: Regressions of one year excess stock market returns on KGR(10) and other predictors

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	14.83 (3.11)	13.15 (4.29)	18.68 (7.38)	10.69 (16.15)	21.03 (9.38)	15.77 (3.75)
KGR(10)	-2.83*** (0.86)	-2.67*** (1.01)	-3.09** (1.31)	-3.08** (1.28)	-2.58*** (0.98)	-2.78*** (1.07)
$\Delta \log(\text{GDP})$		0.40 (0.48)				
$\log(\text{CGV})$			-1.96 (2.66)			
$\log(\text{P/D})$				1.48 (5.46)		
$\log(\text{P/E})$					-2.58 (3.82)	
CAY						1.33* (0.77)
Sample	1917- -2015	1930- -2015	1930- -2015	1917- -2015	1917- -2015	1945- -2015
R^2	0.077	0.075	0.069	0.078	0.080	0.147

Note: see caption of Table 2 for explanations.

Table 22 detrends the top income share by subtracting the 10 year moving average, and the results are similar to Table 2. Table 23 repeats the analysis of Table 2 but with the Kalman filter with an AR(1) cyclical component as discussed in Appendix E. (The likelihood ratio test rejects the i.i.d. cyclical component assumption ($p = 0.00$), and fails to reject AR(1) against AR(2) ($p = 0.27$.) This filter is one-sided (the cycle estimate in year t is based only on data up to year t).³² The results are roughly the same as in Table 2.

³²Following the advice of Hamilton (2018), we do not consider the Hodrick-Prescott (HP) filter because it is two sided: in contrast to the Kalman filter, the HP filter uses past, current, and future data to obtain a smooth trend, thereby potentially introducing a look-ahead bias. For example, since the rich are likely to be more exposed to the stock market, when the stock market goes up at year $t + 1$, the rich will be richer than usual. But then the trend in the top income share will shift upwards, and the year t deviation of the top income share will be lower. Therefore the low income share at year t may spuriously predict a high stock return at $t + 1$.

Table 21: Regressions of one year excess stock market returns on KGR(0.1) and other predictors

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	11.12 (2.59)	10.15 (3.90)	15.19 (8.17)	9.67 (18.14)	14.32 (11.05)	12.00 (3.65)
KGR(0.1)	-3.65** (1.52)	-3.37* (1.90)	-4.28 (2.66)	-3.80 (2.50)	-3.41** (1.74)	-3.38 (2.19)
$\Delta \log(\text{GDP})$		0.33 (0.49)				
$\log(\text{CGV})$			-1.74 (3.12)			
$\log(\text{P/D})$				0.51 (6.15)		
$\log(\text{P/E})$					-1.32 (4.33)	
CAY						1.22 (0.75)
Sample	1913- -2015	1930- -2015	1930- -2015	1913- -2015	1913- -2015	1945- -2015
R^2	0.045	0.043	0.038	0.045	0.045	0.096

Note: see caption of Table 2 for explanations.

Table 22: Regressions of one year excess stock market returns on top income shares (detrended using a 10 year moving average) and other predictors

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	6.72 (1.91)	4.86 (2.52)	9.32 (5.41)	18.78 (17.06)	25.59 (10.13)	7.30 (1.68)
Top 1%	-2.01** (0.93)	-2.27** (1.07)	-2.75* (1.65)	-1.61* (0.96)	-1.68* (0.95)	-2.54** (1.33)
$\Delta \log(\text{GDP})$		0.42 (0.46)				
$\log(\text{CGV})$			-2.04 (3.16)			
$\log(\text{P/D})$				-3.64 (4.93)		
$\log(\text{P/E})$					-6.98* (3.69)	
CAY						1.71** (0.79)
Sample	1922- -2015	1930- -2015	1930- -2015	1922- -2015	1922- -2015	1945- -2015
R^2	0.042	0.054	0.047	0.048	0.062	0.109

Note: see caption of Table 2 for explanations.

Table 23: Regressions of one year excess stock market returns on top income shares (detrended using the Kalman filter) and other predictors

Regressors (t)	Dependent Variable: t to $t + 1$ Excess Market Return							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	6.49 (1.83)	6.37 (1.83)	7.02 (1.76)	4.90 (2.53)	5.34 (4.03)	20.51 (13.03)	20.45 (10.03)	7.68 (1.68)
Top 1%	-3.06* (1.63)			-3.71** (1.76)	-3.17 (2.10)	-2.58* (1.53)	-2.97* (1.64)	-4.69*** (1.71)
Top 1% (AR(2))		-2.72* (1.53)						
Top 0.1%			-4.78** (2.06)					
$\Delta \log(\text{GDP})$				0.48 (0.44)				
$\log(\text{CGV})$					0.69 (2.34)			
$\log(\text{P/D})$						-4.28 (3.78)		
$\log(\text{P/E})$							-5.21 (3.61)	
CAY								1.71*** (0.65)
Sample	1913- -2015	1913- -2015	1913- -2015	1930- -2015	1930- -2015	1913- -2015	1913- -2015	1945- -2015
R^2	0.029	0.026	0.036	0.049	0.035	0.040	0.042	0.115

Note: see caption of Table 2 for explanations.

Table 24: Regressions of one year excess stock market returns on KGR components

(t)	Dependent Variable: t to $t + 1$ Excess Market Return						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	19.22 (10.04)	11.64 (5.30)	17.22 (10.93)	-10.20 (46.16)	11.97 (3.89)	11.85 (3.74)	13.59 (4.07)
α	-0.61 (0.49)						
$Y_{0.1}^k/Y^k$		-0.27 (0.25)					
Y_1^k/Y^k			-0.24 (0.24)				
Y_{10}^k/Y^k				0.17 (0.53)			
$\rho_{0.1}$					-0.16** (0.08)		
ρ_1						-0.20* (0.10)	
ρ_{10}							-0.30** (0.14)
Sample	1913- -2015	1913- -2015	1913- -2015	1962- -2015	1913- -2015	1913- -2015	1913- -2017
R^2	0.02	0.01	0.01	0.00	0.03	0.03	0.04

Note: see caption of Table 1.

E 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

$$y_t = g_t + u_t, \quad (\text{E.1})$$

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

$$(1 - L)^2 g_t = \sigma_1 \epsilon_{1t}, \quad \epsilon_{1t} \sim i.i.d. N(0, 1), \quad (\text{E.2a})$$

$$\phi(L)u_t = \sigma_2 \epsilon_{2t}, \quad \epsilon_{2t} \sim i.i.d. N(0, 1), \quad (\text{E.2b})$$

where L is the lag operator and

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

is the lag polynomial for the autoregressive process. For concreteness, assume

$p = 1$ so $\phi(z) = 1 - \phi_1 z$. Then (E.1) and (E.2) can be written as

$$\begin{aligned} x_t := \begin{bmatrix} g_t \\ g_{t-1} \\ u_t \end{bmatrix} &= \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \phi_1 \end{bmatrix} \begin{bmatrix} g_{t-1} \\ g_{t-2} \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\ &=: Ax_{t-1} + B\epsilon_t. \end{aligned} \tag{E.3}$$

Hence (E.1) and (E.3) reduce to

$$x_t = Ax_{t-1} + B\epsilon_t, \tag{E.4a}$$

$$y_t = Cx_t, \tag{E.4b}$$

where $C = [1 \ 0 \ 1]$. (E.4a) is the state equation and (E.4b) is the observation equation of the state space model. We can then estimate the model parameters $\phi_1, \sigma_1, \sigma_2$ as well as the trend $\{g_t\}$ by maximum likelihood:³³ see Chapter 13 of Hamilton (1994) for details. The extension to a general AR(p) model is straightforward.

F International data

For each of the 29 countries, the top income share series is the “fiscal income” top 1% share from the World Wealth and Income Database (<http://wid.world/>). These series are constructed from tax data and represent the fraction of income accruing to the top 1% of income earners. However, the treatment of capital gains and the definition of a tax unit varies across countries. See the documentation and citations at <http://wid.world/> for details on the construction of the top share in each country.

We calculate annual stock market returns for each country as the percentage change in the end-of-December local currency MSCI total return index (MSRI dividend convention) converted to real terms by dividing by the local consumer price index (CPI). The daily indexes are from Datastream.

For most countries, the annual CPI is from the World Bank Development Indicators (<http://data.worldbank.org/>). For China and Argentina the CPI is from Haver (<http://www.haver.com/>), the U.K. CPI is from FRED, and the Taiwanese CPI is from their government statistics website (eng.stat.gov.tw). For Germany, the pre-1992 CPI is from West Germany, and to get a single consistent German series we combine data from Haver and the World Bank.

Combining these data, we get 815 country/return/lagged-inequality observations (covered return years in parentheses): Argentina (1998–2005), Australia (1970–2014), Canada (1970–2011), China (1993–2015), Colombia (1994–2011), Denmark (1971–1973, 1975–2011), Finland (1988–1990, 1994–2010), France (1970–2015), Germany (1972, 1975, 1978, 1981, 1984, 1987, 1990, 1993, 1996, 1999, 2002–2012), India (1993–2000), Indonesia (1988, 1991, 1994, 1997, 1999–2005), Ireland (1988–2010), Italy (1975–1996, 1999–2010), Japan (1970–2011), Korea (1996–2013), Malaysia (1989, 1994–1996, 2001–2004, 2006, 2010–2013), Mauritius (2003–2009, 2011–2012), Netherlands (1971, 1974, 1976, 1978, 1982, 1986,

³³In our empirical analysis, we use the Matlab `ssm` command to specify the state space model, `estimate` command to estimate parameters by maximum likelihood, and `filter` command to estimate the states.

1990–2013), New Zealand (1988–2014), Norway (1970–2012), Portugal (1990–2006), Singapore (1970–1992, 1994–2013), South Africa (1993–1994, 2003–2013), Spain (1982–2013), Sweden (1970–2014), Switzerland (1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996–2011), Taiwan (1988–2014), U.K. (1970–1980, 1982–2008, 2010–2013), and U.S. (1970–2015).