Asset Pricing with Endogenously Uninsurable Tail Risks

Hengjie Ai and Anmol Bhandari*

February 26, 2017

This paper studies asset pricing in a setting where idiosyncratic risks in labor productivities are uninsurable due to limited commitment. Firms provide insurance to workers using long-term contracts but neither side can commit to these relationships. Under the optimal contract, sufficiently adverse shocks to worker productivity are uninsured. In general equilibrium, exposure to down-side tail risks results in higher risk premia, more volatile returns and variation of returns across firms. The risk sharing patterns are also consistent with the observed cross-sectional heterogeneity in earnings and wealth sensitivities to aggregate shocks.

Key words: Equity premium puzzle, dynamic contracting, tail risk, limited commitment

*Hengjie Ai (hengjie@umn.edu) is affiliated with the Carlson School of Management, University of Minnesota, Anmol Bhandari (bhandari@umn.ed) is at the Department of Economics, University of Minnesota. The authors would like to thank Urban Jermann, Hanno Lustig, Lukas Schmid, Stijn Van Nieuwerburgh and seminar participants at the UBC winter finance conference, Wharton international finance workshop, Minnesota Workshop in Macroeconomic Theory, and the 2016 NBER AP meetings and Rochester University for their helpful comments. The authors would also like to thank Jincheng Tong and Chao Ying for their assistance and comments on the paper.

1 Introduction

A key challenge for asset pricing theories is to simultaneously account for the large magnitude of equity premia and its substantial variation over time and across firms. In this paper we provide an incomplete-markets based asset pricing model that uses limited commitment as a micro-foundation to address these patterns in risk premia. Uninsured tail risks arise as an outcome of optimal risk sharing arrangements and the essential mechanisms that drive risk prices in our model are disciplined by data on individual earning dynamics and how they vary over business cycles. Our setting is consistent with the observed cyclicality of factor shares in the aggregate and the cross-sectional heterogeneity in individual earnings and wealth exposures to aggregate shocks.

The setup consists of two types of agents: "capital owners" and "workers". Capital owners are well-diversified and provide insurance to workers against idiosyncratic fluctuations in labor productivities using long-term compensation contracts.[1] The key feature that distinguishes our paper from standard representative agent asset pricing models is that neither firm owners and workers can commit to contracts that yield a value lower than their outside options. We embed these contracting frictions in a general equilibrium setting with aggregate shocks and then study the resulting asset pricing implications.

While both of the limited commitment constraints are required to match earning dynamics, the down-side risks in labor earnings, a key feature in the data, are driven mainly by the firm-side limited commitment. Perfect risk sharing implies that compensation contracts insure workers against idiosyncratic labor productivity shocks. When firms cannot commit to negative net present value projects, large drops in labor productivities must be accompanied by reductions in workers earnings. This

[1] Guiso et al. (2005), and more recently Juhn et al. (2017) use matched employer-employee data to provide empirical support for the role of labor compensation contracts as an insurance mechanism that firms provide to workers.

feature of the optimal contract is consistent with the empirical evidence presented in Gosh (2016), who use the Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX) data to demonstrate that significant drops in labor earnings are also associated with substantial reductions in individual consumption.

In general equilibrium, exposure to down-side risks under the optimal contract drives several of our asset pricing results. First, it generates a more volatile stochastic discount factor than that in an otherwise identical economy without limited commitment. With recursive utility and persistent countercyclical idiosyncratic risks, the prospect of future lack of risk sharing raises workers' current marginal utilities. As a result, the optimal risk sharing scheme compensates by allocating a higher share of aggregate output to the workers. Therefore, labor share moves negatively with the aggregate output. The counter-cyclicality of labor share translates into a pro-cyclical consumption share of all unconstrained investors, including the capital owners, and amplifies risk prices. In our quantitative analysis we find that Sharpe ratios are doubled due to agency frictions.

Second, our model produces substantial predictable variations in the risk premium, especially over long horizons. The dynamics of the pricing kernel depend on the fraction of firms that are likely to hit their limited commitment constraint. This introduces persistent variations in the volatility of the equilibrium stochastic discount factor and makes returns predictable. Regressing returns on the claim to the aggregate consumption claim on price-dividend ratios, we find that R-squares are similar to those in the data and also increase as we consider longer horizon returns. This time variation in discount rates also amplifies the response of asset prices to aggregate shocks and further elevates the market equity premium. In addition, because unisured downside risks amplifies the volatility of SDF in our model, tail risks measured from the cross section of equity returns have strong predictive power for aggregate stock market returns, consistent with the evidence provided in Kelly and Jiang (2014) and Herskovic et al. (2015).

Third, the above economic mechanism also results in a significant heterogeneity in the cross-section of equity returns. Under the optimal contract, labor compensation insures workers against aggregate productivity shocks and is counter-cyclical, making the residual capital income pro-cyclical and more exposed to aggregate shocks. This delivers a form of operating leverage at the firm level. In particular, firms that have experienced adverse idiosyncratic shocks have a higher fraction of their value promised to workers and are therefore more sensitive to aggregate shocks. As a result, they have lower valuation ratios and higher expected returns. This implication of our model is therefore consistent with the well-documented value premium, the fact that market-to-book ratio is negatively related to expected returns in the cross section.

Lastly, the risk sharing arrangement in our model is consistent with the cross-sectional variation in earnings and wealth sensitivities to aggregate shocks. In our model, individual earnings are history dependent and workers with histories of adverse shocks experience a larger earnings drop on average when the economy goes into a recession. Moreover, by analyzing the consumption replicating portfolio, we find that wealthy agents endogenously hold a higher fraction of wealth in the stock market, whereas low income workers invest more in the riskless asset. This is because workers who realize favorable productivity shocks are typically unconstrained, and therefore their marginal rate of substitutions are equalized with that of the well-diversified capital owners whose consumption is more exposed to shocks that affect the aggregate endowment. These outcomes are in line with recent studies such as Fagereng et al. (2016) who use administrative data on wealth and income from Norway and document that individuals with more uninsured labor income risk hold less risky portfolios.

This paper builds on the literature on incomplete market models with limited commitment. Kehoe and Levine (1993) and Alvarez and Jermann (2000) develop a theory of incomplete market based on one-sided limited commitment. On the asset pricing side, Alvarez and Jermann (2001) and Chien and Lustig (2009) study the asset pricing implications in such an environment. Most of the above theory builds on the Kehoe and Levine (1993) framework and implies that agents who experience large positive income shocks have an incentive to default because they have better outside options. As a result, positive income shocks cannot be insured while tails risks in labor income are perfectly insured. Our paper develops a model of two-sided lack of commitment and we focus on the general equilibrium effects of firm-side limited commitment that have not been studied before.[2]

Our paper is related to asset pricing models with exogenously incomplete markets that builds on Mankiw (1986) and Constantinides and Duffie (1996). Constantinides and Duffie demonstrate how countercyclical volatility in individual income amplify aggregate risk premia in general equilibrium. Schmidt (2015) and Constantinides and Ghosh (2014) calibrate incomplete market models to recent administrative data on earnings and show that higher moments of labor income shocks requires a significant risk compensation. For tractability, the Constantinides and Duffie (1996) framework requires the assumption of independent income shocks to rule out trading of financial assets in equilibrium. Heaton and Lucas (1996) and Storesletten et al. (2007) are amongst the few papers that depart from the no-trade equilibria and study risk premia in quantitative incomplete-market models.[3]

Different from the above papers, we take an optimal contracting approach to micro-found incomplete markets and use empirical evidence on labor earnings dynamics to discipline the choice of the parameters governing agency frictions in the quantitative exercise. Our model allows trading of financial assets and we explicitly characterize the history dependence of labor earnings under the optimal contract. We show that the implications of our model is consistent with the empirical evidence on the cross-sectional variation of the exposures of earnings and wealth to aggregate shocks as documented in Guvenen et al. (2014).

[2]The principal-side limited commitment problem in our model has a similar structure to those studied in Bolton et al. (2014) and Ai and Li (2015). However, none of these papers allow for aggregate risks and study asset pricing and the equity premium.

[3]From the theoretical perspective, Krueger and Lustig (2010) provide conditions under which idiosyncratic risks are irrelevant for risk prices.

The theoretical predictions of our model are also consistent with the recent literature that emphasize the importance of labor share dynamics in understanding the equity market. Our operating leverage results connect to insights in Danthine and Donaldson (2002), Berk and Walden (2013). More recently, Favilukis and Lin (2015) and Favilukis et al. (2016) develop models with sticky wages to demonstrate the importance of counter-cyclical labor share in explaining equity premium and credit risk premium in production economies. The implication of our model that variations in labor share can account for a large fraction of aggregate stock market variations is consistent with the evidence documented in Greenwald et al. (2014) and Ludvigson et al. (2014).

Our computational method builds on the work of Krusell and Smith (1998). Using techniques standard in the dynamic contracting literature such as Thomas and Worrall (1988) and Atkeson and Lucas (1992), we represent our equilibrium allocations recursively with the help of distribution of promised values as a state variable. However, in contrast to those papers, our environment has aggregate shocks and the distribution of promised values responds to such shocks even in the ergodic steady state. As in Krusell and Smith (1998), we approximate the forecasting problem of long lived agents by assuming that agents use few relevant moments of the distribution of promised values to guess future state prices.

The paper is organized as follows. We layout the environment - preferences, technology and the key contracting frictions in Section 2. In Section 3 we discuss the key features of the optimal contract. In Section 4 we derive our key results about endogenous tail risks, procylical consumption share of the capital owners and implications of operating leverage analytically. Finally in Section 5 we present the quantitative implications of our model after calibrating it to several U.S. aggregate and cross-sectional facts. Section 6 concludes.

2 The Model

Preferences and technology

We consider a discrete time infinite horizon economy with $t = 0, 1, 2 \ldots$. There are two groups of agents, a unit measure of capital owners and a unit measure of workers. Preferences are homogeneous across both groups of agents and represented by the Epstein-Zin form with risk aversion γ and intertemporal elasticity of substitution (IES) ψ. Production takes place in a continuum of firms that each hire one worker. Output of firm j at time t, $y_{j,t}$ is

$$y_{j,t} = Y_t s_{j,t}, \tag{1}$$

where $s_{j,t}$ is a worker-specific productivity and Y_t is the aggregate productivity common across all workers. The aggregate technological possibilities evolve stochastically with

$$\ln Y_{t+1} - \ln Y_t = g_{t+1}, \tag{2}$$

where g_t is a finite state Markov process with transition matrix Π and a typical element in Π is $\pi\left(g'|g\right)$. The worker-specific productivity follows

$$\ln s_{j,t+1} - \ln s_{j,t} = \varepsilon_{j,t+1}, \tag{3}$$

where $\varepsilon_{j,t}$ is i.i.d. across workers distributed with density $f(\varepsilon_{j,t}|g_t)$ and we normalize $\mathbb{E}\left[e^{\varepsilon_{j,t}}|g_t\right] = 1$. We use $(g_t, \varepsilon_{j,t})$ to denote time t exogenous shocks for a worker and $\left(g^t, \varepsilon_j^t\right) = \{g_s, \varepsilon_{j,s}\}_{s=0}^t$ to denote the history of the shocks up to time t.

The multiplicative specification in (1) is useful for aggregation.[4] The law of large numbers together with the assumption that $e^{\varepsilon_{j,t}}$ has a unit mean implies that

[4] Constantinides and Duffie (1996) and Heathcote et al. (2014) use a similar specification to model idiosyncratic and aggregate risk in labor productivities.

$Y_t(g^t) = \mathbb{E}\left[y_t|g^t\right]$. Allowing the distribution of $\varepsilon_{j,t}$ to depend on the aggregate state of the economy g_t is important, and as we show later, helps capture several features of the dynamics of individual labor income and consumption over business cycles.

Contracts and markets

Workers enter into long-term contracts with firms that promise them labor compensation as a function of the histories of idiosyncratic and aggregate shocks. We denote the compensation contract for worker j using $\mathbf{C}_j \equiv \left\{C_t\left(g^t, \varepsilon_j^t\right)\right\}_{t=0}^{\infty}$. Capital owners are endowed with ownership claims to these firms and have no labor income. There is a competitive market where capital owners can trade a full set of Arrow-Debreu securities. These assumptions imply that capital owners are fully diversified with respect to firm-specific shocks.

Let $\Lambda_{t,t+k}$ denote the price of a claim to one unit of consumption in state g^{t+k} denominated in state g^t consumption numeraire, where its history dependence is suppressed to save notation. The valuation of a firm with contract $\mathbf{C}$ after history g^t is given by

$$V_t\left[\mathbf{C}|g^t, \varepsilon^t\right] = \mathbb{E}\left[\left\{\sum_{j=0}^{\infty} \Lambda_{t,t+k}(y_{t+k} - C_{t+k})\right\}\Big|g^t, \varepsilon^t\right]$$

and a workers utility from $\mathbf{C}$ solves the Epstein-Zin preference recursion

$$U_t\left[\mathbf{C}|g^t, \varepsilon^t\right] = \left[(1-\beta)\, C_t^{1-\frac{1}{\psi}} + \beta\left(\mathbb{M}_t\left\{U_{t+1}\left[\mathbf{C}|g^{t+1}, \varepsilon^{t+1}\right]\right\}\right)^{1-\frac{1}{\psi}}\right]^{\frac{1}{1-\frac{1}{\psi}}},$$

where the certainty equivalent operator $\mathbb{M}_t$ that maps $(g^{t+1}, \varepsilon^{t+1})$ measurable random variable to a (g^t, ε^t) measurable random variable and is defined as

$$\mathbb{M}_t\left\{U_{t+1}\left[\mathbf{C}|g^{t+1}, \varepsilon^{t+1}\right]\right\} \equiv \left(\mathbb{E}\left\{U_{t+1}\left[\mathbf{C}|g^{t+1}, \varepsilon^{t+1}\right]\right\}^{1-\gamma}\Big|g^t, \varepsilon^t\right)^{\frac{1}{1-\gamma}}.$$

Compensation contracts are optimally determined by maximizing firm value imposing two-sided limited commitment: neither firm owners nor workers can commit to contracts that yield payoffs lower than their outside options. For a worker-firm pair, if the contract is reneged, workers retain a fraction of their skills and the firm is dissolved. As a result, firms' outside options are zero, and workers' outside options depend on their productivity. We use $\underline{U}_t(g^t, \varepsilon^t)$ to denote the outside option of a worker after the history $\{g^t, \varepsilon^t\}$.

A worker with initial productivity y_0 is offered a compensation contract that achieves a life-time utility U_0. As in Atkeson and Lucas (1992), we supres worker's name and index contracts by their initial conditions (U_0, y_0). Given a stochastic process for Arrow prices$\{\Lambda_t\}_t$, the contract $\mathbf{C}(U_0, y_0) = \{C_t(g^t, \varepsilon^t | U_0, y_0)\}_{t=0}^{\infty}$ solves :

$$\max_{\mathbf{C}} V_0[\mathbf{C}|g^0, \varepsilon^0] \tag{4}$$

$$U_0[\mathbf{C}|g^0, \varepsilon^0] \geq U_0,$$

$$U_t[\mathbf{C}|g^t, \varepsilon^t] \geq \underline{U}_t(g^t, \varepsilon^t) \quad \forall (g^t, \varepsilon^t), \tag{5}$$

$$V_t[\mathbf{C}|g^t, \varepsilon^t] \geq 0 \quad \forall (g^t, \varepsilon^t). \tag{6}$$

Equations (5) and (6) represent limited commitment on worker's side and on firm's side, respectively. Workers in our economy have access to state-contingent insurance in the form of compensation contracts but limited commitment frictions restrict the set of feasible compensation contracts. In absence of these constraints, workers would be completely insured with respect to their idiosyncratic shocks.[5]

Finally, the consumption of workers implied by the optimal contracts and total

[5]We do not explicitly impose (5) and (6) on capital owners' consumption. This is without loss of generality as capital owners have diversified claims to firms and no labor endowment; this implies that their consumption is only a function of aggregate shocks and hence the limited liability constraints are slack.

consumption of the capital owners, $X_t(g^t)$ satisfies the resource constraint

$$\int \mathbb{E}\left[\mathbf{C}\left(g^t, \varepsilon^t \middle| U_0, y_0\right) \mid g^t\right] \Phi\left(U_0, y_0\right) dU_0 dy_0 + X_t(g^t) = Y_t\left(g^t\right) \qquad \forall g^t,$$

where $\Phi\left(U_0, y_0\right)$ denotes density of the initial distribution of (U_0, y_0).

2.1 Recursive formulation

In this section we will characterize the firm's optimal contracting problem recursively and define a recursive competitive equilibrium for our economy. The homogeneity assumptions on preferences and technology specification help us to construct an equilibrium that has two state variables, (ϕ, g), where ϕ is a one dimensional density that summarizes the agent types and g is the Markov state of aggregate productivity.

We use the notation $\Lambda\left(g' \middle| \phi, g\right)$ for the one-step-ahead stochastic discount factor (SDF) in state (ϕ, g). That is, $\Lambda\left(g' \middle| \phi, g\right)$ is the price measured in state (ϕ, g) consumption numeraire for one unit of consumption good delivered in the next period contingent on the realization of aggregate shock g'. To make the notation compact, we use $z' = (g', \varepsilon')$ for the vector of realization of next period shocks and $\Omega(dz'|g)$ as the measure over z' given the current aggregate state g. Given our stochastic structure we can factor the joint density $\Omega(dz'|g) \equiv \pi(g'|\, g) f(d\varepsilon'|\, g')$, where $f(d\varepsilon'|\, g')$ is the conditional distribution of ε' in aggregate state g'.

Let $V\left(U, y \middle| \phi, g\right)$ be the value function that attains the optimum of the sequence problem (4). In Appendix 1, we show that $V\left(U, y \middle| \phi, g\right)$ satisfies

$$V\left(U, y \middle| \phi, g\right) = v\left(\frac{U}{y}\middle| \phi, g\right) y,$$

for some function $v\left(\cdot \middle| \phi, g\right)$ that represents "normalized" firm values. We then introduce the notation for normalized utility and normalized consumption:

$$u = \frac{U}{y}; \quad c = \frac{C}{y},$$

The normalized continuation utility u can be interpreted as a measure of a worker's share in the firm's valuation, because U is the total utility delivered by the compensation contract and y is firm size. We express the optimal contracting problem recursively as:

$$v\left(u|\phi,g\right) = \max_{c,\{u'(z')\}} \left\{ (1-c) + \int \Lambda\left(g'|\phi,g\right) e^{g'+\varepsilon'} v\left(u'\left(z'\right)|\phi',g'\right) \Omega(dz'|g) \right\} \tag{7}$$

$$s.t: \quad u = \left[(1-\beta)\, c^{1-\frac{1}{\psi}} + \beta \left(\mathbb{M}_g \left\{ e^{g'+\varepsilon'} u'(z') \right\} \right)^{1-\frac{1}{\psi}} \right]^{\frac{1}{1-\frac{1}{\psi}}} \tag{8}$$

$$\mathbb{M}_g \left\{ e^{g'+\varepsilon'} u'(z') \right\} = \left\{ \int \left[e^{g'+\varepsilon'} u'\left(z'\right) \right]^{1-\gamma} \Omega(dz'|g) \right\}^{\frac{1}{1-\gamma}} \tag{9}$$

$$v\left(u'\left(z'\right)|\,\phi',g'\right) \geq 0, \ \ for \ \ all \ \ z'. \tag{10}$$

$$u'(z') \geq \underline{u}(g) \tag{11}$$

Constraint (7) is the promise keeping constraint that ensure the contract delivers the initial promised value. Constraints (10) and (11) are the recursive counterpart of equations (5) and (6). Here we have assumed that the outside values for the agent, $U_t^*(g^t, \varepsilon^t)$ can be expressed as $\underline{u}(g_t)y_t$.[6]

Finally, we describe the construction of our aggregate state variable ϕ, which we will refer to as the "summary measure". Let $c\,(u)$ denote the compensation policy for the optimal contracting problem (7) and let $\Phi\,(U,y)$ denote the joint distribution of (U,y). In general, $\Phi\,(U,y)$ is needed as a state variable in the construction of a

[6]This restriction is without loss of generality if workers retain a fraction of their labor productivity when they leave the firm.

recursive equilibrium, because the resource constraint,

$$\int\int c\left(\frac{U}{y}\right)y\Phi\left(U,y\right)dydU+X=Y,$$

depends on $\Phi\left(U,y\right)$. We can write the integral in the above equation as

$$\begin{aligned}\int\int c\left(\frac{U}{y}\right)y\Phi\left(U,y\right)dydU &= Y\int\int c\left(u\right)\frac{y}{Y}\Phi\left(yu,y\right)ydydu \\ &= Y\int c\left(u\right)\left[\int\frac{y}{Y}\Phi\left(yu,y\right)ydy\right]du.\end{aligned}$$

We define the summary measure $\phi\left(u\right)=\int\frac{y}{Y}\Phi\left(yu,y\right)ydy$. Note that for a given y, the expression $\Phi(yu,y)y$ is the joint density of (u,y) and thus $\phi(u)$ is the share of endowment owned by agents of type u. The resource constraint then simplifies to

$$\int c\left(u\right)\phi\left(u\right)du=1-x\left(\phi,g\right), \tag{12}$$

where $x\left(\phi,g\right)=\frac{X}{Y}$ denotes the capital owners' share in aggregate consumption in state (ϕ,g). The above procedure reduces the two-dimensional distribution Φ into a one-dimensional measure ϕ and simplifies our analysis.

2.2 Equilibrium

Let $u'\left(u,g',\varepsilon'|\,\phi,g\right)$ be the optimal policy for continuation utilities, the law of motion of ϕ , $\phi'\equiv\Gamma\left(g'|\,\phi,g\right)$ is given by

$$\forall\tilde{u},\qquad \phi'\left(\tilde{u}\right)=\int\phi\left(u\right)\int e^{\varepsilon'}f\left(\varepsilon'|\,g'\right)I_{\{u'(u,g',\varepsilon'|\phi,g)=\tilde{u}\}}d\varepsilon'du. \tag{13}$$

We now define the recursive competitive equilibrium:

Definition 1. *A recursive competitive equilibrium consists of a law of motion for* ϕ, $\Gamma\left(g'|\,\phi,g\right)$, *SDF* $\left\{\Lambda\left(g'|\,\phi,g\right)\right\}_{g'}$, *capital owners' consumption share* $x\left(\phi,g\right)$, *value* *functions $v(u|\,\phi,g)$ and associated policy functions $c(u|\,\phi,g)$, $u'(u,g',\varepsilon'|\,\phi,g)$ such that i) the SDF is consistent with the capital owners' consumption:*[7]

$$\Lambda(g'|\,\phi,g)=\beta\left[\frac{x(\phi',g')\,e^{g'}}{x(\phi,g)}\right]^{-\frac{1}{\psi}}\left[\frac{w(\phi',g')\,e^{g'}}{\mathbb{M}_g e^{g'}w(\phi',g')}\right]^{\frac{1}{\psi}-\gamma},\tag{14}$$

where the capital owners' utility is defined using the recursion

$$w(\phi,g)=\left[(1-\beta)\,x(\phi,g)^{1-\frac{1}{\psi}}+\beta\left(\mathbb{M}_g e^{g'}w(\phi',g')\right)^{1-\frac{1}{\psi}}\right]^{\frac{1}{1-\frac{1}{\psi}}}.\tag{15}$$

ii) given the SDF, and the law of motion of ϕ, the value function and the policy functions solve the optimal contracting problem in (7) and given the policy functions, the law of motion of the summary measure ϕ satisfies (13); and iii) the policy functions and the summary measure ϕ satisfy the resource constraint (12).

3 Optimal Contract

As a point of departure, we derive the policy functions for an economy where constraints (10) and (11) are absent and then use them to contrast with the limited commitment case.

With perfect risk sharing and homothetic preferences, consumption of all agents in the economy must be proportional to each other. Therefore, the compensation of a worker with normalized promised value u, is proportional to u and the valuation of the firm, the policy functions for normalized continuation utility $u'(u,g',\varepsilon'|\,\phi,g)$ are

$$v^{FB}(u)=\left(\frac{u^{FB}(g)^{1-\frac{1}{\psi}}}{1-\beta}\right)\left[1-\frac{u}{u^{FB}(g)}\right],\quad u'(u,g',\varepsilon'|\,\phi,g)=\frac{u^{FB}(g')}{u^{FB}(g)}u\cdot e^{-\varepsilon'},\tag{16}$$

[7] Here we specify the SDF as a function of the capital owners' consumption directly without explicitly specifying the capital owners' consumption and portfolio problem for brevity. Because the the capital owners are well-diversified, their consumption and investment choices are standard.

where $u^{FB}(g)$ is the utility to consumption ratio of a representative agent who consumes the aggregate endowment.[8]

With perfect risk sharing, continuation utility does not respond to idiosyncratic shocks and thus normalized continuation utility $u'(u, g', \varepsilon' | \phi, g)$ must be inversely proportional to the idiosyncratic shock $e^{\varepsilon'}$ as in equation (16) . In particular, an extremely negative shock in ε' pushes $u'(u, g', \varepsilon' | \phi, g)$ toward infinity and can result in a firm value, $v^{FB}(u)$ in equation (16) to be negative. Since workers consume the same fraction of aggregate consumption, a firm with a worker who is extremely unproductive sustains a negative profits that drives the ex-post valuations below zero. Conversely, an extremely large realization of ε' revises $u'(u, g', \varepsilon' | \phi, g)$ downwards and will violate the limited commitment constraint of the worker. Clearly, the perfect risk sharing contract violates constraints (10) and (11) and is therefore not feasible under limited commitment.

Let $\phi' = \Gamma(g'|\phi, g)$ and $\overline{u}(\phi, g)$ be the level of normalized utility at which the limited commitment constraint binds, that is, $v(\overline{u}(\phi, g)|\phi, g) = 0$. The next proposition characterizes the properties of $u'(u, \varepsilon', g'|\phi, g)$ under the optimal contract with limited commitment.

Proposition 1. *There exists* $\underline{\varepsilon}(u, g'|\phi, g)$ *and* $\overline{\varepsilon}(u, g'|\phi, g)$ *with* $\frac{\partial \underline{\varepsilon}(u,g'|\phi,g)}{\partial u} > 0$ *and* $\frac{\partial \overline{\varepsilon}(u,g'|\phi,g)}{\partial u} > 0$ *such that*

$$u'(u, \varepsilon', g'|\phi, g) = \begin{cases} \overline{u}(\phi', g') & \varepsilon' \leq \underline{\varepsilon}(u, g'|\phi, g), \\ \underline{u}(g') & \varepsilon' \geq \overline{\varepsilon}(u, g'|\phi, g), \end{cases}$$

and $u'(u, \varepsilon', g'|\phi, g)$ *is strictly decreasing in* ε' *for all* $\varepsilon' \in (\underline{\varepsilon}(u, g'|\phi, g), \overline{\varepsilon}(u, g'|\phi, g))$.

By the above proposition, sufficiently low realizations or high realizations of ε'

[8] $u^{FB}(g)$ solves $u^{FB}(g) = \left[(1-\beta) + \beta \left(\sum_{g'} \pi(g'|g) \left(u^{FB}(g') e^{g'} \right)^{1-\gamma} \right)^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \right]^{\frac{1}{1-\frac{1}{\psi}}} \quad \forall g$

are not associated with changes in u'. As a result, the unnormalized continuation utility, $U' \propto e^{-\varepsilon'}u'$ moves one to one with ε' and so do the levels of future labor compensation. Thus earnings fall proportionally with negative shocks for all $\varepsilon' \in (-\infty, \underline{\varepsilon}(u|g', \phi, g))$ and the worker is compensated with higher levels of future compensation for $\varepsilon' \in (\overline{\varepsilon}(u, g'|\phi, g), \infty)$.[9]

While both of the limited commitment constraints are important to match earning dynamics, the down-side risks in labor earnings, a key feature in the data, are driven mainly by the firm-side limited commitment. Furthermore, exposure to such tail risk plays a key role in driving our quantitative asset pricing results. In next section we focus on the asset pricing implications of firm-side limited commitment.

4 Role of Firm-side limited commitment

We analyze a special case of our model that is analytically tractable and present three results that demonstrate how limited commitment on the firm side enhances the equity premium. First, under the optimal contract, idiosyncratic tail risks, i.e., a sufficiently adverse realization of ε draws, are not hedged. Second, combined with recursive preferences, persistent countercyclical idiosyncratic risks translate into a countercyclical aggregate labor share and a procyclical consumption share of the capital owners, making the equilibrium stochastic discount factor more volatile. Third, the optimal contract generates a form of operating leverage, making claims to unproductive firms more risky than aggregate consumption.

We initialize the economy at time 0 where all workers have the same promised utility u_0^* and make the following assumptions on preferences and technologies.

[9] A natural question is whether the fall in earnings upon a binding firm-side limited commitment constraint reflects a drop in hours worked or a drop in wages per hour. Our model is consistent with both interpretations and the allocation after an adverse ε' shock could reflect arrangements such as temporary layoffs, workers shifting from full time to part time jobs, wage cuts, or a mix thereof. Given our agency frictions, separation is not strictly necessary for the implementation of the optimal allocation.

Assumption 1. *Preferences satisfy* $\gamma \geq \psi = 1$.

The crucial assumption here is preference for early resolution of uncertainty. The assumption of unit elasticity of intertemporal substitution is merely for tractability and will be relaxed in the quantitative exercise.

Assumption 2. *Aggregate shocks* $g_t \in \{g_L, g_H\}$ *with* $g_L < g_H$. *From period one on, the transition probability from state* g *to state* g' *satisfies* $\pi(g'|\, g) = 1$ *if* $g' = g$. *The distribution* $f(\varepsilon|g = g_H)$ *is degenerate, and* $f(\varepsilon|g = g_L)$ *is a negative exponential with parameter* λ.[10]

Here we assume that booms ($g_t = g_H$) and recessions ($g_t = g_L$) are permanent and that there are no idiosyncratic shocks in booms as a parsimonious way of modeling persistent aggregate shocks and countercyclical idiosyncratic shocks.

Assumption 3. *For* $t = 2, 3, \cdots$, *workers consume* $C_t = \alpha y_t$.

For tractability, in this simple example, we focus on the optimal contracting problem in period one by assuming the absence of any risk sharing from period two and on. In particular, for $t = 2, 3, \cdots$, workers consume an α fraction of their output. Assumption 4 captures the key feature of the optimal contract in our full model: workers' consumption is more exposed to idiosyncratic shocks in recessions due to limited commitment.

Assumption 4. *The outside option for workers* $\underline{u}(g) = 0$.

To focus on the channels through which firm-side limited commitment operates we assume that workers can fully commit.

As mentioned before, Assumption 1-4 help in obtaining analytical results. In the quantitative section we do not impose any of these assumptions and solve the equilibrium numerically.

[10] See Appendix 3 for the properties of the negative exponential distribution.

Uninsurable tail risks

Our first result is a special case of Proposition 1 and explains how the optimal contract generates endogenously uninsurable tails risks. We use node H to denote the state in period one with a high aggregate growth rate, node L for the state with low aggregate growth, and $C_L(\varepsilon)$ to denote a worker's consumption at node L conditioning on the realization of idiosyncratic shock ε.

Proposition 2. *Under Assumptions 2-4, there exists a* $\overline{\varepsilon}$ *such that* $\forall\varepsilon < \underline{\varepsilon}$, *the elasticity of compensation with respect to idiosyncratic shocks,* $\frac{\partial \ln C_L(\varepsilon)}{\partial \varepsilon} = 1$, *and* $\lim_{\varepsilon \to -\infty} C_L(\varepsilon) = 0$.

Under full risk sharing, workers' compensation does not respond to idiosyncratic shocks, implying a zero elasticity of compensation with respect to ε. On the other hand, in the absence of any risk sharing, this elasticity is one for all ε. Proposition 2 implies that tail risks are not insured under the optimal contract with limited commitment on firm side.

To understand how uninsured tail risks translate into a higher volatility of the equilibrium stochastic discount factor, we need to examine the implications of market clearing and general equilibrium.

Procyclical consumption share of capital owners

Our next result is that fixing IES, a sufficiently high risk aversion is both necessary and sufficient for the consumption share of capital owners to be procyclical. Because capital owners are unconstrained, their marginal rate of substitution must be the relevant stochastic discount factor for all assets in the economy. As a result, the procyclicality of dividend share makes the stochastic discount factor more volatile in our framework relative to an otherwise identical model with full risk sharing. We denote capital owners' consumption at node H and L as x_H and x_L respectively and formalize our result in the following proposition.

Proposition 3. *Under Assumptions 1-4, there exists a* $\hat{\gamma} \in [1, 1+\lambda)$ *such that* $\gamma > \hat{\gamma}$ *implies* $x_H > x_L$*. Moreover as* $\gamma \to 1$*,* $x_H < x_L$*.*

The above proposition has two implications. First, when preference are separable, even though the optimal contract generates uninsurable tail risks (Proposition 1), the pricing kernel implied by the capital owner's consumption is less volatile than the pricing kernel in an otherwise identical but frictionless economy.

This outcome comes from imposing market clearing. Countercyclical idiosyncratic risk implies that relative to booms, a larger fraction of agents get constrained in recessions. Since constrained firms cut compensation, in the aggregate there are more resources available. These resources are allocated between the capital owners and the unconstrained workers by equating their intertemporal marginal rates of substitution. With expected utility this amounts to equalizing the growth rates of consumption of the capital owners and the unconstrained agents. Therefore, for $\gamma = 1 = \frac{1}{\psi}$, the consumption share of both capital owners and unconstrained agents must increases and $x_L > x_H$.

The second implication of the above proposition is that keeping IES fixed, a large enough risk aversion results in a procyclical consumption share of capital owners. As risk aversion exceeds the inverse of intertemporal elasticity of substitution, contemporaneous marginal utilities are decreasing in both consumption and continuation values. Recessions that are persistent and associated with lack of risk sharing in the future imply lower continuation values and higher marginal utilities in the current period for unconstrained agents. These conditions imply that optimal risk sharing is now achieved by transferring resources *away* from the capital owners to these unconstrained agents. Proposition 3 says that for sufficiently high risk aversion, this incentive is strong enough to dominate the effect of market clearing and deliver procylical consumption shares for capital owners.

Our result about what drives capital owners' consumption and eventually risk prices is in contrast with the exogenously incomplete market models, for example, Constantinides and Duffie (1996) and several other papers that build on it. In those papers all agents are marginal investors of the stock market and hence countercyclical uninsurable risks in consumption automatically translate into a more volatile pricing kernel. In our model, idiosyncratic risks are uninsured due to binding incentive compatibility constraints. However, agents with binding limited liability constraints will not be the marginal investors. The asset pricing implications depend on the dynamics of the marginal utility of unconstrained agents: the capital owners and some workers who receive high productivity shocks. We show that in order to obtain a more volatile pricing kernel, we need recursive preferences with the property that marginal utility is decreasing in continuation utilities.

Operating leverage

For simplicity, we have assumed that the distribution of promised utility is a point mass on u_0^* in period 0. Keeping u_0^* and equilibrium prices fixed, we now use the notation $\left\{c_0(u_0), \{u'(u_0, \varepsilon, g)\}_{\varepsilon,g}\right\}$ for the optimal contract of a firm with initial promised utility u_0, with the understanding that such agents are of measure 0 for $u_0 \neq u_0^*$.

Our final result is that the optimal labor compensation contract creates a form of operating leverage and elevates the risk exposure of dividends relative to that in an otherwise identical representative firm model. At the aggregate level, it means that the corporate sector's equity is more risky than the claim to aggregate consumption. In the cross section, this mechanism generates heterogeneity in expected returns with respect to u_0: more constrained firms have lower valuation ratios (or market-to-book ratios) and higher risk exposure. We summarize our result in the following proposition.

Proposition 4. *(Operating leverage)*

Under Assumptions 1-4, there exists a $\hat{\gamma} \in [1, 1+\lambda)$ *such that* $\gamma > \hat{\gamma}$ *implies:*

$$\frac{v\left(u\left(u_0^*, g_H\right) \mid g_H\right)}{E\left[e^{\varepsilon} v\left(u'\left(u_0^*, \varepsilon, g_L\right) \mid g_L\right)\right]} > 1. \tag{17}$$

In addition, for all $u_0 < \hat{u}$, *where* $\hat{u}$ *is defined as* $\underline{\varepsilon}(\hat{u}, g_L) = \ln \frac{1+\lambda}{\lambda}$,

$$\frac{\partial}{\partial u_0}\left(\frac{v\left(u\left(u_0, g_H\right) \mid g_H\right)}{E\left[e^{\varepsilon} v\left(u'\left(u_0, \varepsilon, g_L\right) \mid g_L\right)\right]}\right) > 0. \tag{18}$$

Because we assume unit IES, the price-to-dividend ratio of the aggregate consumption claim is constant in a representative agent economy. Therefore, the first part of Proposition 4 implies that firms' cash flow in our model has a higher exposure to aggregate shocks than in an otherwise identical representative agent economy. This is because compensation to workers must provide insurance against aggregate shocks under the optimal contract. Equation (18) provides a comparative statics result with respect to u_0: the higher the fraction of firms' cash flow provided to workers, the more riskyis the residual dividend.

High u_0 firms can be interpreted as low market-to-book ratio firms — these firms promised a large fraction of their cash flow to workers and therefore have low valuation ratios. They are more likely to be constrained and have higher expected returns reflecting the higher endogenous operating leverage. Proposition 4 therefore is also consistent with the value premium, the empirical fact that stocks with lower market-to-book ratios have higher historical average returns.

In the quantitative section of the paper, we show that the operating leverage effect is quantitatively significant and creates a high market equity premium as well as a substantial variation in expected returns in the cross section with a only modest level of risk aversion.

5 Quantitative Analysis

In order to confront our model with data we make two extensions to our Section 2 baseline setting. First, we modify the the aggregate productivity process to have unpredictable shocks to its growth rates with stochastic volatility that better matches the time series properties of aggregate consumption. We change equation (2) to

$$\ln Y_{t+1} - \ln Y_t = g_{t+1} + \sigma(g_{t+1})\eta_{t+1} \tag{19}$$

where g_t is a finite state Markov process as before and η_t is i.i.d standard Gaussian. We allow the volatility of the Gaussian component $\sigma(g)$ to depend on the persistent aggregate state g.[11]

The second extension is that we allow entry and exit of workers to maintain a stationary long-run distribution of idiosyncratic productivities in the model. Existing workers exit the work force at rate $\kappa > 0$ per period. A measure $\frac{1}{\kappa}$ of new workers start every period with the initial condition $s_0 = 1$ and $u_0(g) = u^{new}(g)$, where $u^{new}(g)$ is to be calibrated. This assumption ensures that the total measure of workers in the economy is always one, and the distribution of idiosyncratic productivity is stationary.

Neither of the above extensions are crucial for our asset pricing implications. However, they allow our model to capture some quantitative features of the data, and therefore make it possible for us to use relevant moments in the data to discipline parameter choices in calibration.

We now discuss the numerical methods that we use to solve the general model and then apply them to analyze its quantitative implications.

[11] Equilibrium prices and the optimal contract satisfy homogeneity property, and the presence of i.i.d η shocks does not increase the state space of the equilibrium value and policy functions.

5.1 Numerical algorithm

The computational challenge in our setup is that the distribution ϕ is an infinite dimensional state variable. We use a numerical procedure similar to that of Krusell and Smith (1998) and replace ϕ with suitable summary statistics. The distribution ϕ enters the problem through its effect on the stochastic discount factor and market clearing. To approximate the law of motion Γ, we conjecture that agents project down the stochastic discount factor on the space spanned by $\{g_t, x_t\}$ and use $x' = \Gamma_x(g'|g, x)$ as the law of motion for the transition of x_t.

Using x as the state variable is both numerically efficient and computationally convenient. Note that the stochastic discount factor depends on ϕ for two reasons. First, ϕ affects capital owners consumption, and this effect is completely summarized by x. Second, ϕ is a sufficient statistic to forecast future prices. Our method is numerically efficient because given the law of motion of x, the equilibrium stochastic discount factor is completely determined without solving for the optimal contract. Through market clearing conditions in equation (12) we observe that the x appropriately summarizes higher moments of ϕ as the function $c(\cdot)$ is strictly increasing and convex. This choice contrasts our algorithm from that in Krusell and Smith (1998) who use the first moment of the distribution of wealth as a sufficient state statistic.

Replacing ϕ with x and $\Gamma(\cdot)$ with the forecasting function $\Gamma_x(\cdot)$, one can compute the SDF, $\Lambda(g', \eta'|g, x)$.[12] Given $\Lambda(g', \eta'|g, x)$, we can solve the optimal contract to obtain optimal compensation $c(u|g, x)$ and continuation values $u'(\varepsilon', g', u|g, x)$. Then we simulate a panel of agents using these policy functions and update the law of motion $\Gamma_x(\cdot)$ using simulated data. We repeat this procedure until the the function $\Gamma_x(\cdot)$ converges.

---

[12] With η shocks: $\Lambda\left(g', \eta' \middle| \phi, g\right) = \beta \left[\frac{x\left(\phi', g'\right) e^{g' + \sigma(\eta')\eta'}}{x(\phi, g)}\right]^{-\frac{1}{\psi}} \left[\frac{w\left(\phi', g'\right) e^{g' + \sigma(\eta')\eta'}}{n(\phi, g)}\right]^{\frac{1}{\psi} - \gamma} \Bigg|_{\phi' = \Gamma(\phi | g, g', \eta')}$.

Appendix 6 describes the detailed steps and diagnostics necessary to implement this fixed point numerically.

5.2 Calibration

Using the algorithm outlined in the previous section, we now assess the model's ability to generate sizable equity premia and other relevant asset pricing implications in a calibrated economy. We discipline the stochastic process for aggregate shocks and the conditional distributions for idiosyncratic productivity shocks so that the endogenous labor earnings implied by our optimal contracting framework are consistent with several aggregate and cross-sectional moments of U.S. labor earnings.

A period is a quarter and we time aggregate outcomes in our model to report annual moments. There are two sets of parameters, those governing the stochastic process for aggregate shocks, Y_t and worker specific parameters. We discuss them separately.

We assume that aggregate productivity process g is a two-state Markov chain with state space $\{g_H, g_L\}$ and refer to "booms" as states with $g = g_H$ and "recessions" as states with $g = g_L$. The aggregate shock process $\{g_t, \eta_t\}$ is calibrated as in Ai and Kiku (2013) who jointly estimate the levels of g_H, g_L, the Markov transition matrix, and the volatility parameters, $\sigma(g_H)$ and $\sigma(g_L)$ from post-war aggregate consumption data. Our calibration implies an average duration of 12 years for booms and 4 years for recessions. The parameters for the aggregate shocks are summarized in the top panel of Table 2.

For the worker specific parameters, we assume that $f(\varepsilon|g)$ is Gaussian in booms and follows a mixture of a Gaussian and a fat-tailed distribution with negative exponential form in recessions with one extra parameter. This assumption leaves us with two parameters $\{\mu_i, \sigma_i\}_{i \in H,L}$ for the Gaussian distribution per aggregate state, two parameters $\{\lambda, \varepsilon^{max}\}$ for the negative exponential, and $\rho \in (0,1)$ as the mixing probability that represents the probability from a draw of the negative exponential distribution. Since shocks ε are modeled as shares we have $\int e^{\varepsilon} f(\varepsilon|g) d\varepsilon = 1$. We further assume a conditional mean of unity for each of the individual distribution in the mixture too. These restrictions imply $\mu_i = -\frac{\sigma_i^2}{2}$ and $\varepsilon^{max} = \log \frac{1+\lambda}{\lambda}$.

For the rest of the parameters, we set $\underline{u}(g) = \underline{\alpha} u^{FB}(g)$ and $u^{new}(g) = \alpha^{new} u^{FB}(g)$ where $u^{FB}(g)$ is the utility consumption ratio in absence of limited commitment and is defined in footnote 8. We set $\kappa = 1\%$ per quarter to reflect an average duration of staying in the work force for 25 years. This leaves us with 6 parameters $\{\sigma_L, \sigma_H, \lambda, \rho, \underline{\alpha}, \overline{\alpha}\}$ and we next discuss the moments in the data that we use to calibrate them.

As a baseline we will use moments from Guvenen et al. (2014) who use detailed administrative data from Social Security Administration (SSA) to document the cyclical properties of earnings over business cycles for the sample 1979-2010 as our targets. The advantage of the administrative data is that the tail moments, which are crucial for disciplining the model's workings, are measured very accurately. However, a major drawback is that we are unable to control employment status, transfers from the government and demographics that we abstract from in our setup. Ignoring these aspects could lead us to overestimate the effects of agency frictions.

As an alternative, we also use the Panel Study of Income Dynamics (PSID) and construct cross-sectional moments for labor earnings after tax and transfers and after controlling for several observable characteristics.[13] A trade-off to using PSID is that the sample is small and the estimates of tail moments could be potentially noisy. However comparing the results across the two data sets gives a reasonable range to evaluate the key implications of the model.

[13] The PSID is a longitudinal household survey of U.S. households with a nationally representative sample of over 18,000 individuals. Information on these individuals and their descendants has been collected continuously, including data covering employment, income, wealth, expenditures, health, education, and numerous other topics. The PSID data were collected annually 1968-97 and biennially after 1997.

To obtain our target moments, we compute several cross-sectional moments of one-year log earnings level and earnings growth which are then averaged separately for "boom-years" and "recession-years" using the classification as in Guvenen et al. (2014).[14] For the PSID, we restrict the sample to households where "head of household" is male, working age between 15 and 64, and reported at least 500 hours of work in two consecutive years. We use the residual of post-tax labor earnings after controlling for age of head, the age square, family size and education level of head as the observable characteristics as our measure of earnings.[15] The first two columns in Table 1 reports the summary statistics for both Guvenen et al. (2014) and PSID.

We see that the two data sets give qualitatively similar patterns. The standard deviation of one year earnings growth is approximately acylical and recessions are more left-skewed than booms. The main difference is that PSID measures a lower standard deviation of earnings growth, a lower spread in log earnings, and the left tail has a smaller skewness as compared to the SSA data.

For a given set of values for $\{\sigma_L, \sigma_H, \lambda, \rho, \underline{\alpha}, \alpha^{new}\}$, we simulate the model to compute the model's counterparts of the moments reported in Table 1. In addition, we also target two aggregate moments: the average share labor compensation in aggregate consumption of 70% and the standard deviation of labor share of 5%.[16] The values for the parameters are obtained by minimizing an unweighted Euclidean distance between the simulated moments and those we target in the data. Table 1, reports fit for the two calibrations, one aimed at the SSA data and the other targeting the PSID data. Although, all six parameters are jointly calibrated and in general there is no exact one-to-one mapping between the moments and individual parameters, below we comment on how the moments we include in our calibration

[14]This classification yields the following dates for the last four recessions: 1980–82, 1991–92, 2000–01, and 2007–09.

[15]Tax payments are computed using TAXSIM and as observables we use age of head, the age square, family size, and education level of head.

[16]These are obtained from the BEA using "Wages and Salaries" (Line 3) from Table 2.1 in NIPA as a measure of labor compensation

are informative of the parameters given the structure of the problem.

The average labor share and the spread of labor compensation in the cross-section discipline $\{\underline{\alpha}, \alpha^{new}\}$. A higher outside option for existing workers and higher initial share of surplus for new workers increase the average labor share in the economy. In absence of the worker-side limited commitment, labor compensation is adjusted downwards whenever there is a bad enough idiosyncratic shock. Over long simulations this introduces a negative drift in the average labor compensation and also delivers an earnings distribution that is concentrated.[17] On the other hand, a too high value for α^{new} that makes the worker-side limited commitment more prominent. This sharply reduces negative skewness of earnings growth as labor compensation is adjusted upwards for workers who are up against their outside option.

The excess skewness in recessions relative to booms helps identify the parameters of the mixture distribution in recessions. We estimate a $\lambda = 4.2$ and $\rho \approx 1.5\%$ which implies a very small probability of a draw from the left-skewed distribution. This underscores that the possibility of tail risk is key for operating leverage to be at work; the realized skewness is driven by implications of the optimal contract, and not by recurrent draws of the underlying productivity distributions itself. In addition, for both calibrations, $\{\sigma_i\}$ are helpful in targeting the cross-sectional standard deviation of earnings growth in booms and recessions.

The bottom panel of Table 2 reports the values for all the parameters governing idiosyncratic risks using both the SSA and PSID calibrations.

[17]As is well known in several settings with incomplete markets, over extremely low frequencies, the optimal contract drifts towards region of the state space where agency frictions are minimized. For a more general result see Albanesi and Armenter (2012) about tail properties of second best allocations.

5.3 Results

Asset Pricing

The equity premium We report the mean and the standard deviation of the excess returns and the risk-free interest rate in Table 3. With a risk aversion, $\gamma = 5$ and an IES, $\psi = 2$, our model generates a high equity premium and a low risk-free interest rate for both calibrations: SSA and PSID. Without assuming any financial leverage, the annualized excess return on the aggregate consumption claim is about 3.4% in our model compared to 0.6% in an otherwise identical model but without limited commitment.

Our model generates a high magnitude of equity premia because agency frictions amplify the volatility of the stochastic discount factor. As explained in Proposition 3, countercyclical tail risk together with limited commitment on the firm-side implies a procyclical consumption share of the marginal investors. In Table 3, we see that the maximum Sharpe ratio is approximately doubled due to agency frictions.

The second reason for the high equity premium in our model is the large volatility of stock returns. We discuss this implication next.

Return volatility and predictability Our model produces substantial variations in the volatility of the stochastic discount factor over time. This property accounts for both, a high volatility of the equity market return and a significant predictability of future excess returns by price-to-dividend ratios; this is an empirical fact documented by several papers such as Campbell and Shiller (1988), Fama and French (1988), and Hodrick (1992).

The general equilibrium implications of the agency problem introduce a new channel that raises the volatility of the stochastic discount factor in recessions relative to booms. The intuition for this endogenous variation comes from the persistent changes in the capital owners' share of aggregate consumption, x_t. In our model, prolonged recessions are associated with low levels of the capital owner's consumption share. This implies that small shocks that affect x_t translate into large variations in $\frac{x_{t+1}}{x_t}e^{g_{t+1}}$, which is the consumption growth rate of the capital owners. In equilibrium, the amplified volatility of the capital owner's consumption is compensated by a higher risk premium. On the other hand, in booms, the level of x_t is high, and shocks to x_t have limited impact of the volatility of its growth rate. This asymmetry results in countercyclical risk prices and predictability of market returns by valuation ratios.

Countercyclical volatility of SDF amplifies return volatility. In Table 3, we see that the standard deviation of the consumption claim in our model is 9.27% for the SSA calibration and 10.39% for the PSID calibration. This return volatility is about four times higher than its counterpart in the economy with full commitment. Most of the increases in the volatility of market return comes from time-varying discount rates as the standard deviation of consumption growth is modest, about 2.5% per annum.

In Table 4, we report the results of predictability regressions in our model and in the data. We regress excess stock market returns measured for one to five year horizons on the log price-to-dividend ratio at the start of the measuring period. The left panels report coefficients and R^2 of these regressions using the SP500 returns over the period of 1947-2007, and or data construction follows Beeler and Campbell (2012). We report the same regression results from our simulated data in the right panel. Overall, our model produces regressions coefficients and R^2 fairly close to those in the data and we match the pattern that predictability is higher for longer horizon returns.

The cross-section of expected returns Our last asset pricing result is on the cross-section of expected returns. As explained in Proposition 4, labor contracts generate a form of operating leverage and produces cross-sectional heterogeneity in expected returns. In our model, firms that experience a history of negative shocks have a higher fraction of their cash flow promised to workers and therefore lower valuation ratios. That is, negative productivity shocks raise workers' share u_t and lower the ratio of the market value of the firm relative to output, $v(u_t)$. Because labor compensation insures workers against aggregate shocks, higher workers' share in these firm implies that the equity claims are more risky. Our model therefore provides an optimal contracting based explanation of the value premium (Fama and French (1992), Fama and French (1995)), the fact that valuation ratios negatively predict stock returns in the cross-section.

In Figure 1, we plot the expected firms returns as a function of the key state variable, u for booms (dashed) and recessions (solid). High u firms are more risky and earn an extra premium of about about 2.5% from their higher cash-flow exposure.

Household earnings and wealth exposures to aggregate risks

Our model implies history dependence in individual earning dynamics. In particular, agents who differ in previous histories of their idiosyncratic shocks also differ in their sensitivity of earnings and wealth to aggregate shocks. This feature is absent in several asset pricing models with exogenously incomplete markets, for example, Constantinides and Duffie (1996), Schmidt (2015), and Constantinides and Ghosh (2014). In order to ensure a tractable equilibrium with zero trade in financial assets, all the above papers assume stylized processes for workers' earning growth dynamics with no history dependence.

We start with implications on earnings and then turn to the analysis of household portfolios.

Endogenous earnings exposure Guvenen et al. (2014) document that low income workers lose a significantly higher percentage of their earnings in recessions than high income individuals. Endogenous earnings dynamics generated by our model are in line with the patterns in data.[18]

Using moments reported in Guvenen et al. (2014), in Figure 2 we plot the cumulative earnings losses as a function of their five-year cumulative earnings prior to the 2008-2010 recession. Figure 3 is the counterpart of Figure 2 using data simulated from our model, where we plot the median change in earnings in a three-year recession for workers ranked by their pre-recession earnings level. The vertical axis is the average change in wages for agents in the corresponding percentile. As before the dashed line describes the implication of a model with full commitment, and the dotted line is the implication of our model with limited commitment. We see that the model replicates these patterns quite well. We highlight that this particular pattern of earnings losses is a unique feature of optimal contracting with firm side limited commitment: agents who received low productivity in the past are more likely to be constrained in the future and face larger average drops in labor compensation.

Endogenous wealth exposure and market segmentation In our model, the capital owners and the unconstrained workers share the same stochastic discount factor while the constrained agents' Euler equation holds with an inequality. This implication of our model leads to heterogeneous exposure of wealth to aggregate risk and can be interpreted as endogenous stock market participation.

In general, all agents' equilibrium consumption under the optimal contract can be replicated by an appropriate portfolio strategy. The sensitivity of this portfolio's value with respect to aggregate shocks is a measure of the equity exposure held by an agent. To formalize the above interpretation of our calibrated model, we define $p\left(u \middle| \phi, g\right)$ as

[18] Appendix B in Guvenen et al. (2014) suggests that cyclical unemployment risk and wage cuts are both important factors that drive tail risks in earnings. As mentioned in footnote 9, our model is consistent with either interpretation.

the present value of workers' compensation normalized by y.[19] We calculate Δ_C as the sensitivity of worker's consumption portfolio with respect to the aggregate shock g:

$$\Delta_C(u,g,x) = \frac{\mathbb{E}\left[p\left(u'\left(u,g_H,\varepsilon'\right|g,x)\right|g_H,x'\left(g_H\right)\right)\left|u,g,x\right]}{\mathbb{E}\left[p\left(u'\left(u,g_L,\varepsilon'\right|g,x)\right|g_L,x'\left(g_L\right)\right)\left|u,g,x\right]}. \tag{20}$$

Intuitively, $\Delta_C(u,g,x)$ calculates the ratio of the value of a worker's compensation portfolio in booms relative to that in recessions and is a measure of its exposure to aggregate risks. A replicating portfolio with only risk free bonds will have $\Delta_c = 1$ and higher values indicate higher exposure to stocks. We plot $\Delta_C(u,g,x)$ for booms and recessions (bottom panel) in Figure 4. We see that as a function of u, Δ_C is downward sloping, confirming the interpretation that workers who have previously faced low productivity shocks and consequently have a currently high u also have a lower exposure to equity markets.

6 Conclusion

We present an asset pricing model where risk premia are amplified by agency frictions. Under the optimal contract, sufficiently adverse shocks to worker productivity are uninsured as firms cannot commit to negative net present value projects. In general equilibrium, exposure to down-side tail risks results in a more volatile stochastic discount factor and time variations in discount rates. These features of the pricing kernel yield quantitatively large and volatile risk premia and generate a substantial cross-sectional variation in returns across firms. Our model is also consistent with observations on how individual earnings and wealth vary in their exposure to aggregate shocks.

[19] The expression for $p(u|\phi,g)$ is given by

$$p\left(u|\,\phi,g\right) = c\left(u|\,\phi,g\right) + (1-\kappa)\int \Lambda\left(g',\eta'|\,\phi,g\right)p\left(u\left(z'\right)|\phi',g'\right)e^{g'+\varepsilon'}\Omega(dz'|g).$$

References

Ai, H. and D. Kiku (2013): "Growth to Value: Option Exercise and the Cross Section of Equity Returns," *Journal of Financial Economics*, 107, 325–349.

Ai, H. and R. Li (2015): "Investment and CEO Compensation under Limited Commitment," *Journal of Financial Economics.*, 116, 452–472.

Albanesi, S. and R. Armenter (2012): "Intertemporal Distortions in the Second Best," *The Review of Economic Studies*, 79, 1271–1307.

Alvarez, F. and U. J. Jermann (2000): "Efficiency, equilibrium, and asset pricing with risk of default b," 68, 775–797.

——— (2001): "Quantitative Asset Pricing Implications of Endogenous Solvency Constraints," *Review of Financial Studies*, 14, 1117–1151.

Atkeson, A. and R. Lucas (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59, 427–453.

Beeler, J. and J. Y. Campbell (2012): "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *Critical Finance Review*, 1, 141–182.

Berk, J. B. and J. Walden (2013): "Limited Capital Market Participation and Human Capital Risk," *Review of Asset Pricing Studies*, 3, 1–37.

Bolton, P., N. Wang, and J. Yang (2014): "Liquidity and Risk Management: Coordinating Investment and Compensation Policies," *Working Paper, Columbia University.*

Campbell, J. Y. and R. J. Shiller (1988): "The dividend-price ratio and expectations of future dividends and discount factors," *Review of financial studies*, 1, 195–228.

Chien, Y. and H. Lustig (2009): "The Market Price of Aggregate Risk and the Wealth Distribution," *Review of Financial Studies*, 23, 1596–1650.

Constantinides, G. and A. Ghosh (2014): "Asset Pricing with Countercyclical Household Consumption Risk," *Working Paper*, 1–60.

Constantinides, G. M. and D. Duffie (1996): "Asset pricing with heterogenous consumers," 104, 219–240.

Danthine, J.-P. and J. B. Donaldson (2002): "Labor Relations and Asset Returns," *Review of Economic Studies*, 69, 41–64.

Fagereng, A., L. Guiso, and L. Pistaferri (2016): "Portfolio choices, firm shocks and uninsurable wage risk," Tech. rep., National Bureau of Economic Research.

Fama, E. and K. French (1992): "The cross section of expected stock returns," *The Journal of Finance*, 2, 427–465.

——— (1995): "Size and book-to-market factors in earnings and returns," *The Journal of Finance*, 50, 135–155.

Fama, E. F. and K. R. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–24.

Favilukis, J. and X. Lin (2015): "Wage Rigidity: A Quantitative Solution to Several Asset Pricing Puzzles," *Review of Financial Studies.*

Favilukis, J., X. Lin, and X. Zhao (2016): "The Elephant in the Room: The Impact of Labor Obligations on Credit Risk," *Working Paper, University of British Columbia.*

Gosh, A. (2016): "Income Versus Consumption Inequality: The Role of Time-Varying Higher Moments," .

Greenwald, D. L., M. Lettau, and S. C. Ludvigson (2014): "The Origins of Stock Market Fluctuations," *National Bureau of Economic Research Working Paper Series*.

Guiso, L., L. Pistaferri, and F. Schivardi (2005): "Insurance within the firm," *Journal of Political Economy*, 113, 1054–1087.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122, 621–660.

Heathcote, J., K. Storesletten, and G. L. Violante (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104, 2075–2126.

Heaton, J. and D. J. Lucas (1996): "Evaluating the effects of incomplete markets on risk sharing and asset pricing," *Journal of political Economy*, 104, 443–487.

Herskovic, B., B. Kelly, H. Lustig, and S. V. Nieuwerburgh (2015): "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*.

Hodrick, R. J. (1992): "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357–386.

Juhn, C., K. McCue, H. Monti, and B. Pierce (2017): "Firm Performance and the Volatility of Worker Earnings," Tech. rep., National Bureau of Economic Research.

Kehoe, T. and D. Levine (1993): "Debt-Constrained Asset Markets," *Review of Economic Studies*, 60, 865–888.

Kelly, B. and H. Jiang (2014): "Tail Risk and Asset Prices," *Review of Financial Studies*, 27, 2841–2871.

Krueger, D. and H. Lustig (2010): "When is Market Incompleteness Irrelevant for the Price of Aggregate Risk (and when is it not)?" *Journal of Economic Theory*, 145, 1–41.

Krusell, P. and A. A. Smith (1998): "Macroeconomy Per Krusell," *Journal of Political Economy*, 106, 867–896.

Ludvigson, S. C., M. Lettau, and M. Sai (2014): "Capital Share Risk and Shareholder Heterogeneity in U.S. Stock Pricing," *NBER Working paper*.

Mankiw, N. G. (1986): "The Equity Premium and The Concentration of Aggregate Shocks," *Journal of Financial Economics*, 17, 211–219.

Schmidt, L. (2015): "Climbing and Falling Off the Ladder : Asset Pricing Implications of Labor Market Event Risk," *Working Paper*, 1–74.

Stokey, N., R. Lucas, and E. with Prescott (1989): "Recursive Methods in Economic Dynamics," .

Storesletten, K., C. I. Telmer, and A. Yaron (2007): "Asset pricing with idiosyncratic risk and overlapping generations," *Review of Economic Dynamics*, 10, 519–548.

Thomas, J. and T. Worrall (1988): "Self Enforcing Wage Contracts," *The Review of Economic Studies*, 55, 541–553.

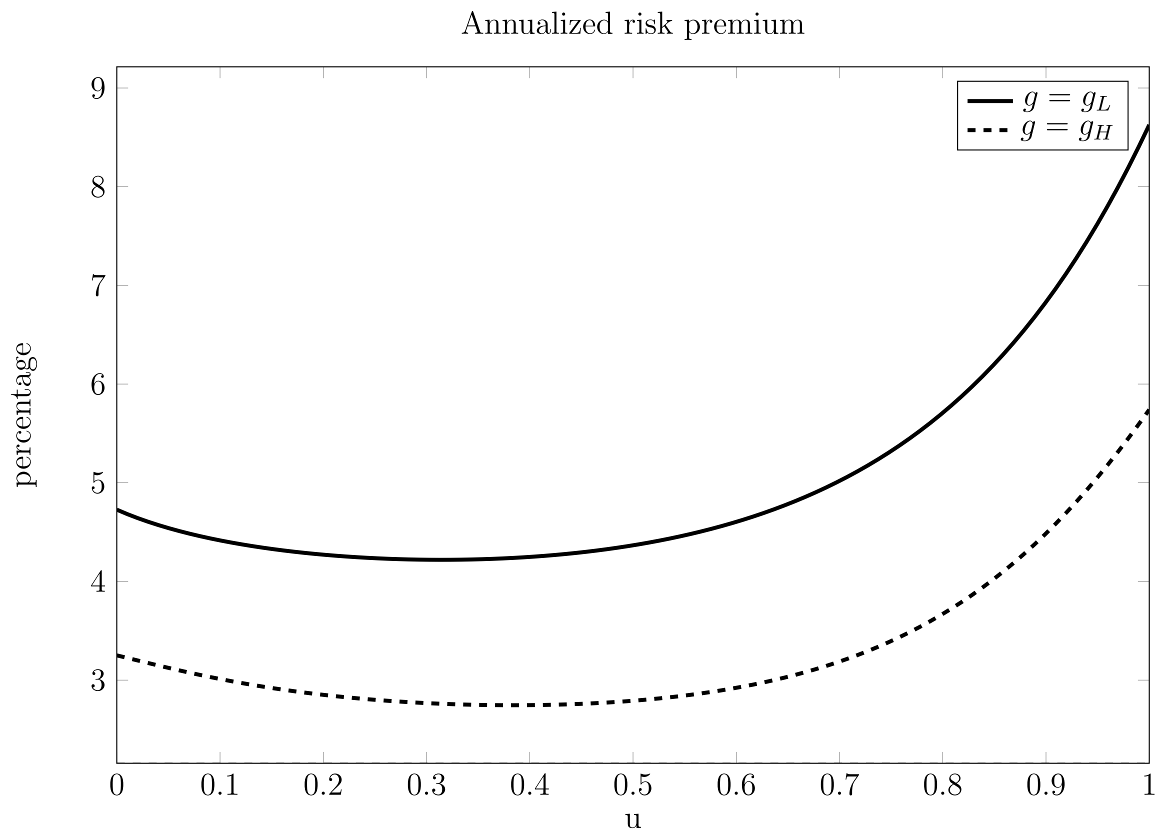

Figure 1: Annualized equity premium on firms that differ in u. The x-axis is normalized such that $\underline{u}(g) = 0$ and $\overline{u}(g) = 1$ for $g = g_L$and $g = g_H$.

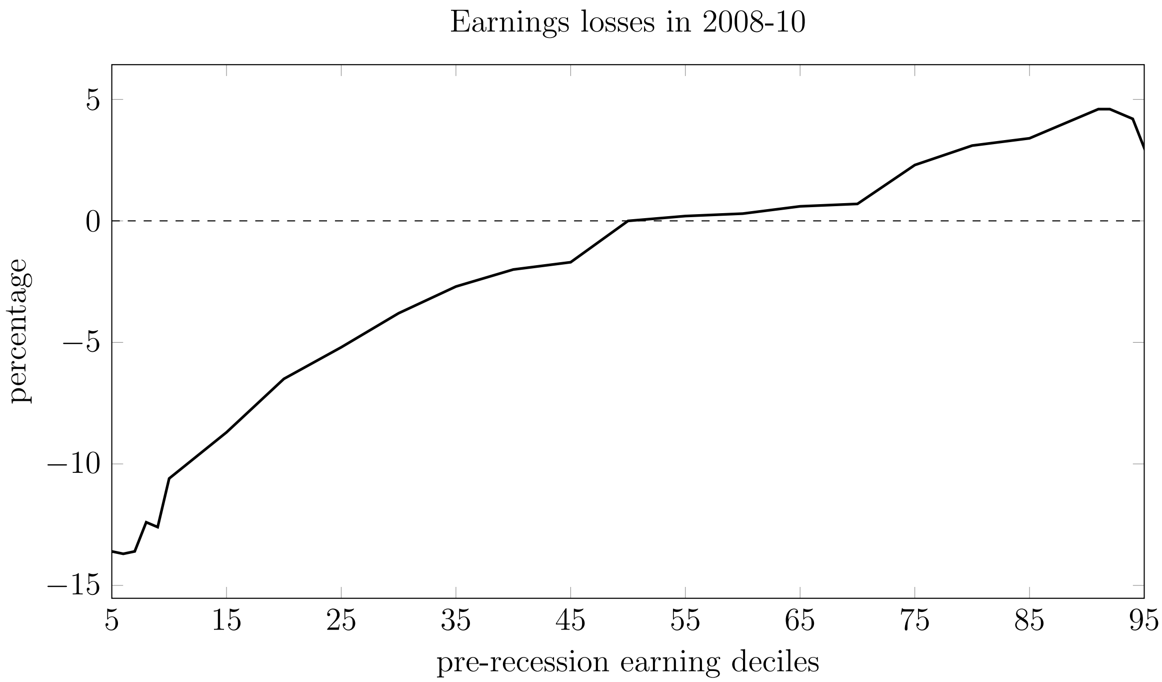

Figure 2: Demeaned earning losses in 2008-2010 recession [Data: Guvenen et al. (2014)]

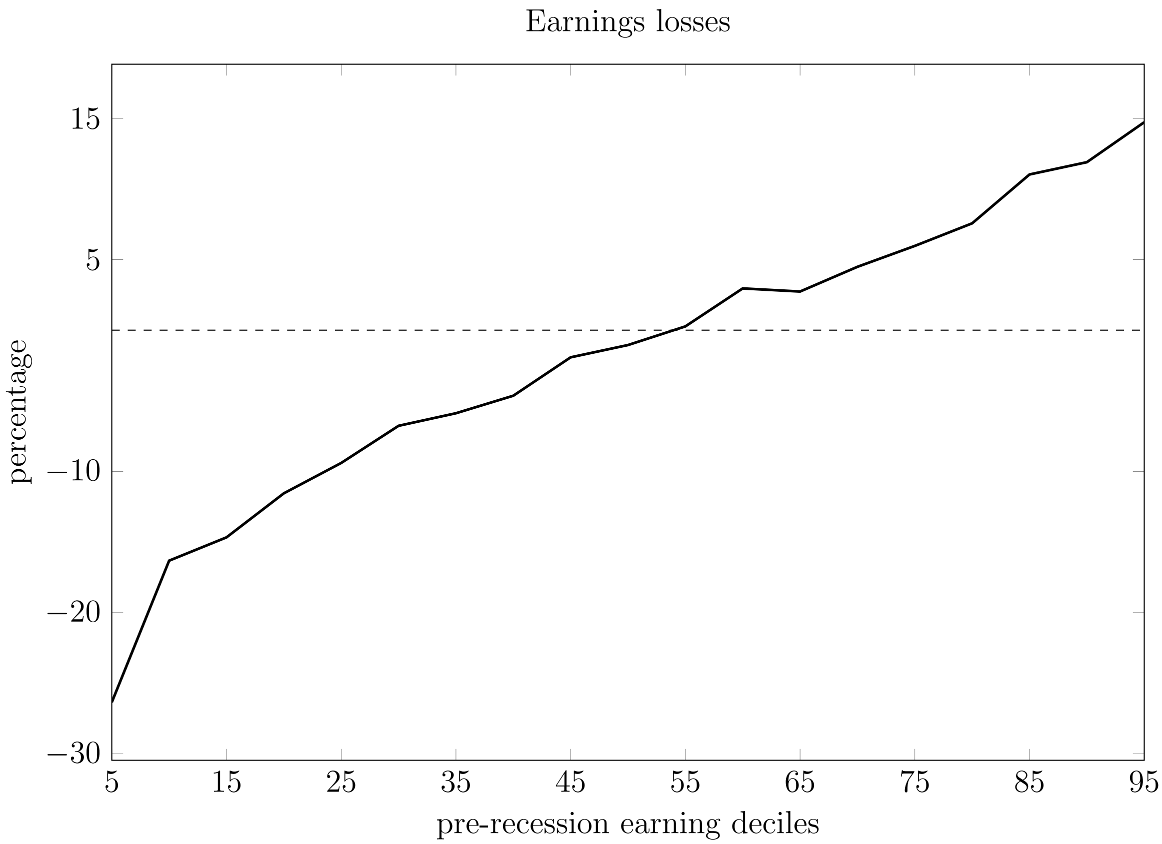

Figure 3: Median earnings losses for 12 quarters of g_L.

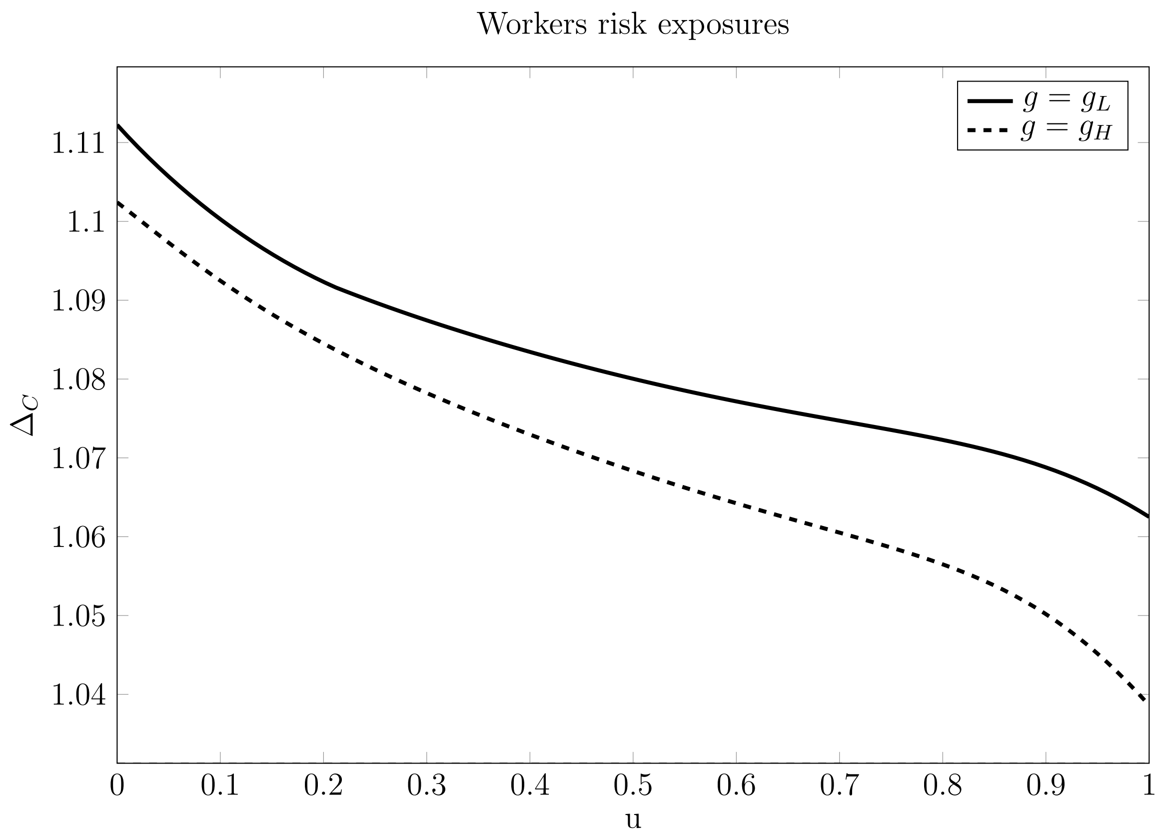

Figure 4: Workers' risk exposure measured using equation (20). The x-axis is normalized such that $\underline{u}(g) = 0$ and $\overline{u}(g) = 1$ for $g = g_L$and $g = g_H$.

Table 1: Targeted moments and model fit

Moments	Data		Model	
	SSA	PSID	SSA	PSID
Std. of 1 yr earnings growth in booms	0.53	0.31	0.55	0.33
Std. of 1 yr earnings growth in recessions	0.54	0.32	0.59	0.33
log 95 - log 5 earnings in booms	3.06	1.71	3.79	2.18
log 95 - log 5 earnings in recessions	3.05	1.62	3.85	2.28
log 90 - log 10 earnings in booms	2.12	1.254	2.89	1.70
log 90 - log 10 earnings in recessions	2.14	1.158	2.96	1.72
log 75 - log 25 earnings in booms	0.98	0.61	1.52	0.92
log 75 - log 25 earnings in recessions	0.98	0.57	1.45	0.89
Kelly skewness of 1 yr earnings in growth booms	3.5%	0.0%	0.0%	1.6%
Kelly skewness of 1 yr earnings growth in recessions	-8.9%	-5.9%	-4.9%	-3.0%
Average Labor share	70%		69.6%	68.4%
Std. of labor share	5%		4.8%	5.13%

The column "SSA" refers to moments and the calibration targeted to data reported in Guvenen et al. (2014) and the column "PSID" corresponds to the moments and calibration targeted to sample in Panel Study of Income Dynamics.

Table 2: Calibrated parameters

Parameters	Values	
Aggregate Risk		
g_H, g_L	0.7%, -0.3%	
$\sigma(g_H), \sigma(g_L)$	0.9%, 1.3%	
$\pi(g_H\|g_H), \pi(g_L\|g_L)$	0.98, 0.94	
Idiosyncratic Risk	SSA	PSID
σ_H, σ_L	11.30%, 10.27%	6.37%, 4.52%
λ, ρ	1.50%, 4.2	1.23%, 4.2
$\underline{\alpha}, \alpha^{new}$	0.16, 0.23	0.30, 0.40

The top panel lists the parameters governing the aggregate shocks and the bottom panel lists the parameters governing the idiosyncratic risk and agency frictions. "SSA" and "PSID" in the bottom panel refer to calibration targeted to data reported in Guvenen et al. (2014) and the sample in Panel Study of Income Dynamics respectively.

Table 3: Asset pricing implications

Moments	Full commitment	Limited commitment	
		SSA	PSID
Equity premium on Y_t	0.62%	3.39%	3.49%
Volatility of excess returns on Y_t	2.22%	9.27%	10.39%
Equity premium on $x_t Y_t$	0.62%	3.79%	3.76%
Volatility of excess returns on $x_t Y_t$	2.22%	10.76%	11.51%
Volatility of SDF $\sigma(\log \Lambda \vert g_L), \sigma(\log \Lambda \vert g_H)$	0.27, 0.15	0.50, 0.27	0.47, 0.28
Average risk free rate	4.55%	1.55%	1.48%
Volatility of risk free rate	0.44%	2.21%	3.01%

All moments are annualized. The columns “SSA” and “PSID” refer to results using the calibrations targeted to data reported in Guvenen et al. (2014) and the sample in Panel Study of Income Dynamics respectively.

Table 4: Return predictability

Horizon J	Model		Data	
(years)	β	R^2	β	R^2
1	-0.22	0.07	-0.12	0.09
2	-0.37	0.12	-0.20	0.15
3	-0.47	0.15	-0.27	0.20
4	-0.53	0.16	-0.32	0.23
5	-0.58	0.17	-0.41	0.27

Table 4 reports the coefficients and R^2 of the regressions $\sum_{j=1}^{J}(r_{t+j} - r_{f,t+j}) = \alpha + \beta(pd_t) + \epsilon_{t+j}$. The "Model" column uses simulated data and the "Data" column follows the construction in Beeler and Campbell (2012).

Appendix: Asset Pricing with Endogenously Uninsurable Tail Risks

Hengjie Ai and Anmol Bhandari

February 26, 2017

1 Normalized value functions

For a given stochastic process, $\{\Lambda_{t,t+1}\}_t$ for one-period ahead Arrow security prices, let $V_t(y, U)$ be the value attained by sequential problem (4) for some history of aggregate shocks g^t and initial conditions $y_0 = y$ and $U_0 = U$. It solves the following Bellman equation

$$V_t(y, U) = \max_{C,\{U'(z')\}} \left\{(y - C) + \mathbb{E}_t \Lambda_{t,t+1} V_{t+1}\left(y'(y, z'), U'\left(z'\right)\right)\right\} \tag{22}$$

$$U = \left[(1 - \beta) C^{1-\frac{1}{\psi}} + \beta \mathbb{M}_t \left(U'(z')\right)^{1-\frac{1}{\psi}}\right]^{\frac{1}{1-\frac{1}{\psi}}} \tag{23}$$

$$U'(z') \geq \underline{U}(z') \tag{24}$$

$$V_{t+1}\left(y'(y, z'), U'\left(z'\right)\right) \geq 0, \quad for \quad all \quad z'. \tag{25}$$

where the operator $\mathbb{M}_t (U'(z')) \equiv \left\{\mathbb{E}_t U'(z')^{1-\gamma}\right\}^{\frac{1}{1-\gamma}}$.

Guess that $V_{t+1}(y'(y, z'), U'(z'))$ admits a multiplicative decomposition

$$V_{t+1}\left(y'(y, z'), U'\left(z'\right)\right) = v_{t+1}\left(\frac{U(z')}{y'(y, z')}\right) y'(y, z'),$$

for some function $v_{t+1}(.)$. Substituting in the right hand side of (22) and using $y(y, z') =$

$ye^{g'+\varepsilon'}$ we can express the objective function as

$$y\left\{\left(1-\frac{C}{y}\right)+\mathbb{E}_t\Lambda_{t,t+1}v_{t+1}\left(\frac{U(z')}{y'(y,z')}\right)e^{g'+\varepsilon'}\right\}.$$

Since the $M_t(U')$ and hence the right hand side of (23) is a homogeneous of degree one in $C, U^{'}$ we can divide and multiply by y to obtain

$$\frac{U}{y}=\left[(1-\beta)\left(\frac{C}{y}\right)^{1-\frac{1}{\psi}}+\beta M_t\left(\frac{U'(z')}{y'(y,z')}e^{g'+\varepsilon'}\right)^{1-\frac{1}{\psi}}\right]^{\frac{1}{1-\frac{1}{\psi}}}.$$

Next assuming $\underline{U}(z')=\underline{u}(z')y'(z')$ we can scale both the constraints (24) and (22) with $y'(z')$. Using definitions

$$u=\frac{U}{y};\quad c=\frac{C}{y},$$

the objective function in (22) and constraints (23)-(25) can be expressed in normalized consumption and normalized promised values and we obtain the recursion for $v_t(\frac{U}{y})$ as in problem (7) in text.

2 Proof of Proposition 1

Standard arguments from Stokey et. al (1989) imply that the value function v is continuous, strictly decreasing, strictly concave and differentiable in the interior. Applying the envelope condition we obtain

$$\frac{\partial}{\partial u}v\left(u|\,\phi,g\right)=-\frac{1}{(1-\beta)\left(\frac{c(u|\phi,g)}{u}\right)^{-\frac{1}{\psi}}},\tag{26}$$

and the first order condition with respect to continuation utility implies

$$\begin{aligned}&\frac{\beta}{1-\beta}c\left(u|\,\phi,g\right)^{\frac{1}{\psi}}m\left(u|\,\phi,g\right)^{\gamma-\frac{1}{\psi}}\left(e^{g'+\varepsilon'}u'\left(u,g',\varepsilon'\right)\right)^{-\gamma}\\ \geq\;&-\Lambda\left(g'|\,\phi,g\right)\frac{\partial}{\partial u}v\left(u'\left(u,g',\varepsilon'\right)|\,\phi',g'\right),\end{aligned}\tag{27}$$

where $m\left(u|\,\phi,g\right)=\mathbb{M}_g e^{g'+\varepsilon'}u'\left(u,g',\varepsilon'\right)$. Monotonicity of the value function means that the limited commitment constraint (10) can be written as $u'\left(u,g',\varepsilon'\right)\leq\overline{u}\left(\phi',g'\right)$ for all

g', where $\overline{u}(\phi', g')$ is defined as a solution to $v(\overline{u}(\phi', g')|\,\phi', g') = 0$. Then for all possible realizations of (g', ε'), the above conditions can be expressed as

$$\left[\frac{e^{\varepsilon'} c\left(u'\left(u, g', \varepsilon'|\,\phi, g\right)|\,\phi', g'\right)}{c\left(u|\,\phi, g\right)}\right]^{-\frac{1}{\psi}} \left[\frac{e^{\varepsilon'} u'\left(u, g', \varepsilon'|\,\phi, g\right)}{m\left(u|\,\phi, g\right)}\right]^{\frac{1}{\psi}-\gamma} \geq \left[\frac{x\left(\phi', g'\right)}{x\left(\phi, g\right)}\right]^{-\frac{1}{\psi}} \left[\frac{w\left(\phi', g'\right)}{n\left(\phi, g\right)}\right]^{\frac{1}{\psi}-\gamma}. \tag{28}$$

The complementary slackness condition implies that (28) has to hold with equality whenever $u'(u, g', \varepsilon') \in (\underline{u}(\phi', g'), \overline{u}(\phi', g'))$.

Let $\mathrm{E} = \{\varepsilon'$ s.t the limited committment constraints (10) and (11) are slack$\}$. For a given $\tilde{m}$, using the first order condition in equation (28), for all $\varepsilon' \in \mathbb{E}$ define an implicit function $\tilde{u}(u, \varepsilon', \tilde{m}, g'|\phi, g)$ such that $z = \tilde{u}(u, \varepsilon', \tilde{m}|\phi, g)$ solves

$$\left[\frac{e^{\varepsilon'} c\left(z|\phi', g'\right)}{c\left(u|\,\phi, g\right)}\right]^{-\frac{1}{\psi}} \left[\frac{e^{\varepsilon'} z}{m\left(u|\,\phi, g\right)}\right]^{\frac{1}{\psi}-\gamma} = \left(\frac{x\left(\phi', g'\right)}{x(\phi, g)}\right)^{-\frac{1}{\psi}} \left(\frac{w\left(\phi', g'\right)}{n\left(\phi, g\right)}\right)^{\frac{1}{\psi}-\gamma}, \tag{29}$$

where $\phi' = \Gamma(g'|\phi, g)$. Note that u' and $\tilde{u}$ are related via

$$u'(u, g', \varepsilon'|\phi, g) = \tilde{u}\left(u, \varepsilon', m(u|\phi, g), g'|, \phi, g\right).$$

Strict concavity of $v(u)$ with respect to u and equation (26) implies $\frac{\partial c(\cdot|\phi,g)}{\partial u} \geq 0$. Together with equation (29) this yields $\frac{\partial \tilde{u}(u,\varepsilon',\tilde{m},g'|\phi,g)}{\partial \varepsilon'} \leq 0$ and $\frac{\partial \tilde{u}(u,\varepsilon',\tilde{m},g'|\phi,g)}{\partial u} \geq 0$. Since these properties hold for an arbitrary $\tilde{m}$ we conclude that the policy rule for continuation values $u'(u, g', \varepsilon'|\phi, g)$ is decreasing in ε' and increasing in initial promised values u.

Substituting the optimal policy for u', the left hand side of equation (29) is monotonic and continuous in ε'. Both $c(u|\phi, g)$ and $u'(u, g', \varepsilon'|\phi, g)$ are bounded and hence the left hand side approaches zero as $\varepsilon' \to \infty$ and approaches ∞ as $\varepsilon \to -\infty$. For a strictly positive $x(\phi, g)$, by the intermediate value theorem, there exists $\underline{\varepsilon}(u, g'|\phi, g)$ such that

$$\left[\frac{e^{\underline{\varepsilon}} c\left(\overline{u}(\phi', g')|\phi', g'\right)}{c\left(u|\,\phi, g\right)}\right]^{-\frac{1}{\psi}} \left[\frac{e^{\underline{\varepsilon}} \overline{u}(\phi', g')}{m\left(u|\,\phi, g\right)}\right]^{\frac{1}{\psi}-\gamma} = \left(\frac{x\left(\phi', g'\right)}{x(\phi, g)}\right)^{-\frac{1}{\psi}} \left(\frac{w\left(\phi', g'\right)}{n\left(\phi, g\right)}\right)^{\frac{1}{\psi}-\gamma},$$

and $\overline{\varepsilon}(u, g'|\phi, g)$ such that

$$\left[\frac{e^{\overline{\varepsilon}}c\left(\underline{u}(\phi', g')|\phi', g'\right)}{c\left(u|\,\phi, g\right)}\right]^{-\frac{1}{\psi}}\left[\frac{e^{\overline{\varepsilon}}\underline{u}(\phi', g')}{m\left(u|\,\phi, g\right)}\right]^{\frac{1}{\psi}-\gamma}=\left(\frac{x\left(\phi', g'\right)}{x(\phi, g)}\right)^{-\frac{1}{\psi}}\left(\frac{w\left(\phi', g'\right)}{n\left(\phi, g\right)}\right)^{\frac{1}{\psi}-\gamma}.$$

Monotonicity of u' with respect to ε' implies that the set $\mathrm{E} = [\underline{\varepsilon}(u|g'|\phi, g), \overline{\varepsilon}(u|g'|\phi, g)]$ and for $\varepsilon \notin \mathbb{E}$, the limited liability constraints (10) and (11) bind giving us $u'(u, g', \varepsilon'|\phi, g) = \overline{u}(\phi', g')$ for $\varepsilon' < \underline{\varepsilon}(u|g'|\phi, g)$ and $u'(u, g', \varepsilon'|\phi, g) = \underline{u}(\phi', g')$ for $\varepsilon' > \overline{\varepsilon}(u|g'|\phi, g)$.

Lastly we turn to the monotonicity of the threshold $\underline{\varepsilon}(u, g'|\phi, g)$ and $\overline{\varepsilon}(u, g'|\phi, g)$ with respect to u. The right hand side of equation (29) is independent of u but the equation ((29)) needs to be satisfied for all u. The definitions of $\underline{\varepsilon}(u|g'|\phi, g)$ and $\overline{\varepsilon}(u|g'|\phi, g)$ mean that $u'(u, g', \underline{\varepsilon}(u|g'|\phi, g)|\phi, g)$ is constant with respect to u. Hence we have

$$\frac{\partial u'(u, \underline{\varepsilon}(u|g'|\phi, g), g'|\phi, g)}{\partial u}+\frac{\partial u'(u, \underline{\varepsilon}(u|g'|\phi, g), \varepsilon'|\phi, g)}{\partial \varepsilon'}\frac{\partial \underline{\varepsilon}(u|g'|\phi, g)}{\partial u}=0$$

and

$$\frac{\partial u'(u, \overline{\varepsilon}(u|g'|\phi, g), g'|\phi, g)}{\partial u}+\frac{\partial u'(u, \overline{\varepsilon}(u|g'|\phi, g), \varepsilon'|\phi, g)}{\partial \varepsilon'}\frac{\partial \overline{\varepsilon}(u|g'|\phi, g)}{\partial u}=0.$$

Since u' is decreasing with respect to ε' and increasing with respect to u it follows that $\frac{\partial \underline{\varepsilon}(u|g'|\phi,g)}{\partial u} \geq 0$ and $\frac{\partial \overline{\varepsilon}(u|g'|\phi,g)}{\partial u} \geq 0$.

3 Details of the negative exponential distribution

We assume that the density function of the idiosyncratic shock $f\left(\varepsilon|g = g_L\right)$ takes the following form:

$$f\left(\varepsilon|\, g_L\right) = \begin{array}{cc} 0 & \varepsilon > \varepsilon_{MAX} \\ \lambda e^{\lambda(\varepsilon-\varepsilon_{MAX})} & \varepsilon \leq \varepsilon_{MAX} \end{array}$$

For later reference, we note that the moments of $f\left(\varepsilon|\, g_L\right)$ can be easily computed as:

$$\int_{-\infty}^{\varepsilon} e^{\theta t} f\left(t|\, g_L\right) dt = \frac{\lambda}{\lambda+\theta} e^{-\lambda\varepsilon_{MAX}+(\theta+\lambda)\varepsilon} \quad for \quad \lambda+\theta > 0. \tag{30}$$

It is straightforward to use (30) to show that the assumption $E\left[e^{\varepsilon}\right]=1$ amounts to the parameter restriction that $\varepsilon_{MAX}=\ln\frac{1+\lambda}{\lambda}$.

4 Proof for Proposition 2 and 3

We assume in Section 4 that the distribution $f\left(\varepsilon|\, g_H\right)$ is degenerate, and $f\left(\varepsilon|\, g_L\right)$ is the negative exponential distribution with parameter λ. We first introduce some simplifying notations. Let $\underline{\varepsilon}$=$\underline{\varepsilon}(u_0,g_L)$ be the cutoff level of ε such that any lower levels of ε will lead to a binding limited commitment constraint. We denote

$$\begin{aligned} u_H &= u'\left(u_0,g_H\right), \qquad u_L\left(\varepsilon\right)=u'\left(u_0,g_L,\varepsilon\right); \\ v_H &= v\left(u'\left(u_0,g_H\right)\right), \quad v_L\left(\varepsilon\right)=v\left(u'\left(u_0,g_L,\varepsilon\right)\right); \\ c_H &= c\left(u'\left(u_0,g_H\right)\right), \quad c_L\left(\varepsilon\right)=c\left(u'\left(u_0,g_L,\varepsilon\right)\right). \end{aligned}$$

Note that all above quantities are functions of u_0 and we will use the notation $\underline{\varepsilon}(u_0)$, $u_H\left(u_0\right)$, $u_L\left(u_0,\varepsilon\right)$, $v_H\left(u_0\right)$, $v_L\left(u_0,\varepsilon\right)$, $c_H\left(u_0\right)$, and $c_L\left(u_0,\varepsilon\right)$ to emphasize this dependence whenever necessary. Also, we use w_H and w_L to denote the capital owners' utility (normalized by aggregate output) at node H and L, respectively. In addition, let u_H^{FB} and u_L^{FB} denote the utility-to-consumption ratio of an agent who consumes the aggregate consumption in state g_H and g_L, respectively. That is, they are the normalized utility associated with full risk sharing. The first best u_H^{FB} and u_L^{FB} are determined by

$$u_H^{FB}=\left(e^{g_H}u_H^{FB}\right)^{\beta} \quad u_L^{FB}=\left(e^{g_L}u_L^{FB}\right)^{\beta}.$$

Also, we use u_L^{CD} to denote the normalized utility of an agent in an economy without risk sharing (the Constantinides-Duffie economy). That is, it is utility-consumption ratio of an agent who consumes y_t every period:

$$u_L^{CD}=\left(\int\left[e^{\{\varepsilon+g_L\}}u_L^{CD}\right]^{1-\gamma}f(\varepsilon|g_L)d\varepsilon\right)^{\frac{\beta}{1-\gamma}}. \tag{31}$$

It is straightforward to show that as $\gamma\rightarrow 1+\lambda$, $u_L^{CD}\rightarrow 0$.

Optimal risk sharing We first prove the following lemma that uses the optimal risk sharing condition (28) to express c_H as a function of $c_L(\underline{\varepsilon})$, the consumption of the marginal agent whose limited commitment constraint is just about to bind.

Lemma 1. *(FOC for the marginal agent)*

Given the consumption share of the capital owners, x_H and x_L, for all u_0, the normalized consumption of the marginal worker with $\varepsilon_1 = \underline{\varepsilon}(u_0)$ must satisfy:

$$\frac{c_H(u_0)}{e^{(1+\iota)\underline{\varepsilon}(u_0)} c_L(u_0, \underline{\varepsilon}(u_0))} \left[\frac{u_L^{FB}}{\xi u_L^{CD}}\right]^{\iota} = \frac{x_H}{x_L}. \tag{32}$$

Proof. By Proposition 1, the optimal risk sharing condition (28) must hold with equality for the marginal worker with the realization of $\underline{\varepsilon}$ at node L:

$$\left[\frac{c_H}{e^{\underline{\varepsilon}} c_L(\underline{\varepsilon})}\right]^{-1} \left[\frac{u_H}{e^{\underline{\varepsilon}} u_L(\underline{\varepsilon})}\right]^{1-\gamma} = \left[\frac{x_H}{x_L}\right]^{-1} \left[\frac{w_H}{w_L}\right]^{1-\gamma}. \tag{33}$$

We can use the promise keeping constraint to represent continuation utility as functions of consumption. For capital owners,

$$\begin{aligned} w_H &= x_H^{1-\beta} n_H^{\beta}, \quad \textit{where } n_H = (1-\alpha) e^{g_H} u_H^{FB}, \\ w_L &= x_L^{1-\beta} n_L^{\beta}, \quad \textit{where } n_L = (1-\alpha) e^{g_L} u_L^{FB}, \end{aligned} \tag{34}$$

where the computation of continuation utility n_H and n_L uses the fact that capital owners are not exposed to idiosyncratic risks and that together they consume $1-\alpha$ fraction of aggregate output. Because workers are not exposed to idiosyncratic risks at node H and consume α fraction of aggregate output, their continuation utility at node H can be computed similarly:

$$u_H = c_H^{1-\beta} m_H^{\beta}; \quad m_H = \alpha u_H^{FB} e^{g_H}.$$

At node L, workers consume αy_t for $t = 2, 3, \cdots$ and their normalized certainty equivalent in period 2 is

$$m_L(\varepsilon_1) = \left\{\int_{-\infty}^{\infty} \left[e^{g'+\varepsilon'} \left(\alpha u_L^{CD}\right)\right]^{1-\gamma} f\left(\varepsilon' \middle| g_L\right) d\varepsilon'\right\}^{\frac{1}{1-\gamma}} = \alpha \xi u^{CD} e^{g_L},$$

where we define $\xi \in (0,1)$ as:

$$\xi = \left\{ \int_{-\infty}^{\infty} e^{(1-\gamma)\varepsilon'} f\left(\varepsilon' \middle| g_L\right) d\varepsilon' \right\}^{\frac{1}{1-\gamma}}.$$

Therefore,

$$u_L\left(\underline{\varepsilon}\right) = c_L\left(\underline{\varepsilon}\right)^{1-\beta} m_L\left(\underline{\varepsilon}\right)^{\beta}; \quad m_L = \alpha \xi u_L^{CD} e^{g_L}, \tag{35}$$

Now we use expressions in (34) and (35) to replace the continuation utilities in (33) and simplify to get:

$$\left[\frac{c_H}{e^{[1+\iota]\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)}\right]^{-\Omega} \left[\frac{\xi u_L^{CD}}{u_L^{FB}}\right]^{-\beta(1-\gamma)} = \left[\frac{x_H}{x_L}\right]^{-\Omega}, \tag{36}$$

where to simplify notation, we denote

$$\Omega = 1 + (1-\beta)(\gamma - 1) > 0, \quad and \quad \iota = \frac{\beta(\gamma-1)}{\Omega(\gamma)}.$$

We can therefore obtain (32) by raising both sides of equation (33) to their $-\frac{1}{\Omega}th$ power. □

Market Clearing/Aggregation Next, we exploit the market clearing condition to express the total consumption of all workers as a function of $c_L\left(\underline{\varepsilon}\right)$. Our main result is summarized in the following lemma:

Lemma 2. *(Market clearing)*

Given the consumption share of the capital owners, x_H and x_L, for all u_0, the expected consumption of a worker with promised utility u_0 at node L is given by:

$$E\left[e^{\varepsilon'} c_L\left(u_0, \varepsilon'\right)\right] = e^{(1+\iota)\underline{\varepsilon}(u_0)} c_L\left(u_0, \underline{\varepsilon}\left(u_0\right)\right) \left\{\frac{\lambda}{1+\lambda} e^{-\lambda \varepsilon_{MAX} + (\lambda-\iota)\underline{\varepsilon}(u_0)} + \Phi\left(\underline{\varepsilon}\right)\right\}, \tag{37}$$

where the function $\Phi(\varepsilon)$ is defined as:

$$\Phi(\varepsilon) = \frac{\lambda}{\lambda - \iota}\left[e^{-\iota \varepsilon_{MAX}} - e^{-\lambda \varepsilon_{MAX} + (\lambda - \iota)\varepsilon}\right]. \tag{38}$$

Proof. The expected consumption of a worker with promised utility u_0 at node L can be computed as

$$\int_{-\infty}^{\underline{\varepsilon}} e^{\varepsilon'} c_L\left(\underline{\varepsilon}\right) f\left(\varepsilon' \middle| g_L\right) d\varepsilon' + \int_{\bar{\varepsilon}}^{\varepsilon_{MAX}} e^{\varepsilon'} c_L\left(\varepsilon'\right) f\left(\varepsilon' \middle| g_L\right) d\varepsilon'. \tag{39}$$

The first order condition (28) implies that for all $\varepsilon \geq \underline{\varepsilon}$,

$$e^{-\gamma\varepsilon} c_L\left(\varepsilon\right)^{-1} u_L\left(\varepsilon\right)^{1-\gamma} = e^{-\gamma\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)^{-1} u_L\left(\underline{\varepsilon}\right)^{1-\gamma}. \tag{40}$$

Using equation (35),

$$u_L\left(\varepsilon\right) = c_L\left(\varepsilon\right)^{1-\beta}\left[\alpha\xi u_L^{CD} e^{g_L}\right]^{\beta}. \tag{41}$$

We can combine equations (40) and (41) to get:

$$e^{-\gamma\varepsilon} c_L\left(\varepsilon\right)^{-1+(1-\gamma)(1-\beta)} = e^{-\gamma\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)^{-1+(1-\gamma)(1-\beta)}.$$

Raising both sides of the above equation to the $-\frac{1}{1-1+(1-\gamma)(1-\beta)}th$ power and using the definition of Ω and ι, we have, for all $\varepsilon \geq \underline{\varepsilon}$,

$$e^{\varepsilon} c_L\left(\varepsilon\right) = e^{-\iota\varepsilon} e^{(1+\iota)\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right). \tag{42}$$

Now, we compute the integral in (39) as:

$$\begin{aligned} & c_L\left(\underline{\varepsilon}\right)\int_{-\infty}^{\underline{\varepsilon}} e^{\varepsilon'} f\left(\varepsilon' \middle| g_L\right) d\varepsilon' + e^{(1+\iota)\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)\int_{\underline{\varepsilon}}^{\varepsilon_{MAX}} e^{-\iota\varepsilon} f\left(\varepsilon \middle| g_L\right) d\varepsilon \\ = \quad & \frac{\lambda}{1+\lambda} e^{-\lambda\varepsilon_{MAX}+(1+\lambda)\underline{\varepsilon}} c_L\left(\bar{\varepsilon}\right) + e^{(1+\iota)\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)\frac{\lambda}{\lambda-\iota}\left[e^{-\iota\varepsilon_{MAX}} - e^{-\lambda\varepsilon_{MAX}+(\lambda-\iota)\underline{\varepsilon}}\right], \end{aligned}$$

as needed. □

Equation (42) gives us that $C_L(\varepsilon) = e^{\varepsilon} c_L\left(\underline{\varepsilon}\right)$ for $\varepsilon < \underline{\varepsilon}$ and Proposition 2 follows immediately.

General Equilibrium In general equilibrium, $u_0 = \overline{u}_0$, and the market clearing conditions imply that $c_H = 1 - x_H$ and $E\left[e^{\varepsilon'} c_L\left(\varepsilon'\right)\right] = 1 - x_L$. Using Lemma 2, we can write

$$1 - x_L = E\left[e^{\varepsilon'} c_L\left(\varepsilon'\right)\right] = e^{(1+\iota)\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)\left\{\frac{\lambda}{1+\lambda} e^{-\lambda \varepsilon_{MAX} + (\lambda-\iota)\underline{\varepsilon}} + \Phi\left(\underline{\varepsilon}\right)\right\} \tag{43}$$

Using the above equation to replace $e^{(1+\iota)\underline{\varepsilon}} c_L\left(\underline{\varepsilon}\right)$ and using $1 - x_H$ to replace c_H, we can rewrite equation (32) as

$$\frac{x_H}{x_L}\frac{1-x_L}{1-x_H} = \left\{\frac{\lambda}{1+\lambda} e^{-\lambda \varepsilon_{MAX} + (\lambda-\iota)\underline{\varepsilon}} + \Phi\left(\underline{\varepsilon}\right)\right\}\left[\frac{u_L^{FB}}{\xi u_L^{CD}}\right]^{\iota}. \tag{44}$$

Note that if $\gamma = 1$, then $\iota = 0$, and using the definition of $\Phi\left(\underline{\varepsilon}\right)$ in (38),

$$\frac{\lambda}{1+\lambda} e^{-\lambda \varepsilon_{MAX} + (\lambda-\iota)\underline{\varepsilon}} + \Phi\left(\underline{\varepsilon}\right) = 1 - \frac{1}{1+\lambda} e^{\lambda\left(\underline{\varepsilon} - \varepsilon_{MAX}\right)} < 1.$$

It follows immediately that for $\gamma = 1$, $\frac{x_H}{x_L} < 1$ is counter-cyclical. As $\gamma \to 1 + \lambda$, $u_L^{CD} \to 0$, while all other terms on the right hand side of (44) remain bounded. We must have $\frac{x_H}{x_L} \to \infty$. By continuity, there exists $\hat{\gamma} \in (1, 1+\lambda)$ such that $\frac{x_H}{x_L} > 1$ if and only $\gamma > \hat{\gamma}$, as needed.

5 Proof of Proposition 4

Normalized value functions First, note that starting from period two, all workers consume a fraction α of firms' output, and as a result, firms' cash flow is $(1-\alpha) y_t$ for $t = 2, 3, \cdots$. In addition, capital owners consume $1 - \alpha$ fraction of aggregate output. Therefore, given the assumption of unit IES for capital owners, the price-to-dividend ratio of all firms' cash flow is a constant, $\frac{1}{1-\beta}$.

Therefore, at node H, the value of a firm (with promised utility u_0) normalized by output is

$$v_H\left(u_0\right) = 1 - c_H\left(u_0\right) + \Lambda_H\left[e^{g_H}\frac{1}{1-\beta}(1-\alpha)\right],$$

where we use Λ_H for the Arrow-Debreu price of one unit of consumption goods in period two measured in period-one numeraire. Because there is no uncertainty in period two, Λ_H

can be computed from the consumption growth rate of capital owners:

$$\Lambda_H = \beta \left[\frac{(1-\alpha)}{x_H} e^{g_H} \right]^{-1},$$

where x_H and $1-\alpha$ is the consumption share of capital owners at node H in period one and period two, respectively. We have:

$$v_H(u_0) = 1 - c_H(u_0) + \frac{\beta}{1-\beta} x_H. \tag{45}$$

Similarly,

$$v_L(u_0, \varepsilon) = 1 - c_L(u_0, \varepsilon) + \Lambda_L \left[e^{g_L} \frac{1}{1-\beta} (1-\alpha) \right],$$

where $\Lambda_L = \beta \left[\frac{(1-\alpha_H)}{x_L} g_L \right]^{-1}$. Therefore, similar to (45), we have

$$v_L(u_0, \varepsilon) = 1 - c_L(u_0, \varepsilon) + \frac{\beta}{1-\beta} x_L, \tag{46}$$

and

$$E\left[e^{\varepsilon} v_L(u_0, \varepsilon)\right] = 1 - E\left[e^{\varepsilon} c_L(u_0, \varepsilon)\right] + \frac{\beta}{1-\beta} x_L. \tag{47}$$

At $u_0 = \overline{u}_0$, using the market clearing condition, $1 - c_H(\overline{u}_0) = x_H$ and $1 - E\left[e^{\varepsilon} c_L(\overline{u}_0, \varepsilon)\right] = x_L$, equations (45) and (47) imply

$$\begin{aligned} v_H(\overline{u}_0) &= 1 - c_H(\overline{u}_0) + \frac{\beta}{1-\beta} x_H = \frac{1}{1-\beta} x_H, \\ E\left[e^{\varepsilon} v_L(\overline{u}_0, \varepsilon)\right] &= 1 - E\left[e^{\varepsilon} c_L(\overline{u}_0, \varepsilon)\right] + \frac{\beta}{1-\beta} x_L = \frac{1}{1-\beta} x_L. \end{aligned}$$

Clearly,

$$\frac{v_H(\overline{u}_0)}{E\left[e^{\varepsilon} v_L(\overline{u}_0, \varepsilon)\right]} = \frac{x_H}{x_L} > 1$$

if and only if $\gamma > \hat{\gamma}$.

Cross section of expected returns To characterize the dependence of $\frac{v_H(u_0)}{E[e^{\varepsilon} v_L(u_0,\varepsilon)]}$, note that in general,

$$c_H\left(u_0\right) = \frac{x_H}{x_L}\left[\frac{\xi u_L^{CD}}{u_L^{FB}}\right]^{\iota} e^{(1+\iota)\underline{\varepsilon}(u_0)} c_L\left(u_0, \underline{\varepsilon}\left(u_0\right)\right)$$

by Lemma 1 and

$$E\left[e^{\varepsilon} c_L\left(u_0, \varepsilon\right)\right] = e^{(1+\iota)\underline{\varepsilon}(u_0)} c_L\left(u_0, \underline{\varepsilon}\left(u_0\right)\right)\left\{\frac{\lambda}{1+\lambda} e^{-\lambda\varepsilon_{MAX}+(\lambda-\iota)\underline{\varepsilon}(u_0)} + \Phi\left(\underline{\varepsilon}\right)\right\}$$

by Lemma 2. Because at $\varepsilon = \underline{\varepsilon}(u_0)$, the limited commitment constraint, $v_L\left(u_0, \varepsilon\right) = 0$ binds, $c_L\left(u_0, \underline{\varepsilon}\left(u_0\right)\right) = 1 + \frac{\beta}{1-\beta} x_L$ by (46). To simplify notation, we denote $\theta_H = 1 + \frac{\beta}{1-\beta} x_H$ and $\theta_L = 1 + \frac{\beta}{1-\beta} x_L$. We then write $\frac{v_H(u_0)}{E[e^{\varepsilon} v_L(u_0,\varepsilon)]}$ as:

$$\frac{v_H\left(u_0\right)}{E\left[e^{\varepsilon} v_L\left(u_0, \varepsilon\right)\right]} = \frac{\theta_H - \phi\theta_L e^{(1+\iota)\underline{\varepsilon}(u_0)}}{\theta_L\left\{1 - e^{(1+\iota)\underline{\varepsilon}(u_0)} \omega\left(\underline{\varepsilon}\left(u_0\right)\right)\right\}},$$

where we denote $\phi = \frac{x_H}{x_L}\left[\frac{\xi u_L^{CD}}{u_L^{FB}}\right]^{\iota} \theta_L$, and $\omega\left(\underline{\varepsilon}\right) = \frac{\lambda}{1+\lambda} e^{-\lambda\varepsilon_{MAX}+(\lambda-\iota)\underline{\varepsilon}(u_0)} + \Phi\left(\underline{\varepsilon}\right)$ to simplify notation. By Proposition 1, $\underline{\varepsilon}(u_0)$ is a strictly increasing function of u_0. Therefore, to prove Proposition 4, it enough to show

$$\frac{\partial}{\partial\underline{\varepsilon}} \frac{\theta_H - \phi e^{(1+\iota)\underline{\varepsilon}}}{\left\{1 - e^{(1+\iota)\underline{\varepsilon}} \omega\left(\underline{\varepsilon}\right)\right\}} > 0,$$

which is given by the following lemma.

Lemma 3. *There exists $\tilde{\gamma} \in (1, 1+\lambda)$ such that $\gamma > \tilde{\gamma}$ implies that for all $\varepsilon \in (-\infty, \varepsilon_{MAX})$,*

$$\frac{\partial}{\partial\varepsilon}\left[\frac{\theta_H - \phi e^{(1+\iota)\varepsilon}}{1 - e^{(1+\iota)\varepsilon} \omega\left(\varepsilon\right)}\right] > 0. \tag{48}$$

Proof. We can compute (48) as :

$$\begin{aligned} & \frac{\partial}{\partial\varepsilon}\left[\frac{\theta_H - \phi e^{(1+\iota)\varepsilon}}{1 - e^{(1+\iota)\varepsilon} \omega\left(\varepsilon\right)}\right] \\ = \quad & \frac{-\phi e^{(1+\iota)\varepsilon}\left(1+\iota\right)\left[1 - e^{(1+\iota)\varepsilon} \omega\left(\varepsilon\right)\right] + \left[\theta_H - \Phi e^{(1+\iota)\varepsilon}\right] e^{(1+\iota)\varepsilon}\left[\left(1+\iota\right)\omega\left(\varepsilon\right) + \omega'\left(\varepsilon\right)\right]}{\left[1 - e^{(1+\iota)\varepsilon} \omega\left(\varepsilon\right)\right]^2}. \end{aligned}$$

We focus on the numerator and simplify:

$$\begin{aligned}
& -\phi e^{(1+\iota)\varepsilon}(1+\iota)\left[1-e^{(1+\iota)\varepsilon}\omega(\varepsilon)\right]+\left[\theta_H-\Phi e^{(1+\iota)\varepsilon}\right]e^{(1+\iota)\varepsilon}\left[(1+\iota)\omega(\varepsilon)+\omega'(\varepsilon)\right] \\
= \;& \theta_H\left[(1+\iota)\omega(\varepsilon)+\omega'(\varepsilon)\right]-\phi\left[(1+\iota)+e^{(1+\iota)\varepsilon}\omega'(\varepsilon)\right]
\end{aligned}$$

It is therefore enough to show

$$\theta_H\left[(1+\iota)\omega(\varepsilon)+\omega'(\varepsilon)\right]-\phi\left[(1+\iota)+e^{(1+\iota)\varepsilon}\omega'(\varepsilon)\right]>0 \tag{49}$$

Using the expression of $\omega(\varepsilon)$, we can compute

$$\begin{aligned}
(1+\iota)\omega(\varepsilon)+\omega'(\varepsilon) &= (1+\iota)\frac{\lambda}{\lambda-\iota}\left[e^{-\iota\varepsilon_{MAX}}-e^{-\lambda\varepsilon_{MAX}+(\lambda-\iota)\varepsilon}\right] \\
&= (1+\iota)\frac{\lambda}{\lambda-\iota}e^{-\iota\varepsilon_{MAX}}\left[1-e^{-(\lambda-\iota)\varepsilon_{MAX}+(\lambda-\iota)\varepsilon}\right] \\
&= (1+\iota)\frac{\lambda}{1+\lambda}e^{-\iota\varepsilon_{MAX}}\frac{1+\lambda}{\lambda-\iota}\left[1-e^{-(\lambda-\iota)\varepsilon_{MAX}+(\lambda-\iota)\varepsilon}\right] \\
&= (1+\iota)e^{-(1+\iota)\varepsilon_{MAX}}\frac{1+\lambda}{\lambda-\iota}\left[1-e^{-(\lambda-\iota)(\varepsilon_{MAX}-\varepsilon)}\right]>0,
\end{aligned}$$

where the last line uses the fact $\varepsilon_{MAX}=\ln\frac{1+\lambda}{\lambda}$. Also, the second term in equation (49) can be written as

$$\begin{aligned}
(1+\iota)+e^{(1+\iota)\varepsilon}\omega'(\varepsilon) &= (1+\iota)\left[1-\frac{\lambda}{1+\lambda}e^{-\lambda\varepsilon_{MAX}+(1+\lambda)\varepsilon}\right] \\
&= (1+\iota)\left[1-e^{-(1+\lambda)(\varepsilon_{MAX}-\varepsilon)}\right].
\end{aligned}$$

Therefore, to prove (49), it is enough to show that for all ε,

$$\theta_H e^{-(1+\iota)\varepsilon_{MAX}}\frac{1+\lambda}{\lambda-\iota}\left[1-e^{-(\lambda-\iota)(\varepsilon_{MAX}-\varepsilon)}\right]-\phi\left[1-e^{-(1+\lambda)(\varepsilon_{MAX}-\varepsilon)}\right]>0.$$

Because $\phi\to 0$ as $\gamma\to 1+\lambda$, we have $\theta_H e^{-(1+\iota)\varepsilon_{MAX}}>\phi$ for γ close enough to $1+\lambda$. We complete the proof by making the following observation:

Define

$$f(\varepsilon) = \frac{1+\lambda}{\lambda-\iota}\left[1-e^{-(\lambda-\iota)(\varepsilon_{MAX}-\varepsilon)}\right]$$
$$g(\varepsilon) = 1-e^{-(1+\lambda)(\varepsilon_{MAX}-\varepsilon)},$$

then $f(\varepsilon) > g(\varepsilon)$ for all $\varepsilon < \varepsilon_{MAX}$.

To see this, note that $f(\varepsilon_{MAX}) = g(\varepsilon_{MAX}) = 0$. Also, $f'(\varepsilon) < g'(\varepsilon)$ for all $\varepsilon < \varepsilon_{MAX}$, because

$$f'(\varepsilon) = -(1+\lambda)\, e^{-(\lambda-\iota)(\varepsilon_{MAX}-\varepsilon)}$$
$$g'(\varepsilon) = -(1+\lambda)\, e^{-(1+\lambda)(\varepsilon_{MAX}-\varepsilon)}.$$

$\square$

6 Computational Algorithm

Below we describe briefly our computation algorithm.

1. Start with an initial guess of the law of motion of x, $\Gamma_x(g'|g,x)$

$$\log x' = a(g,g') + b(g,g')\log x \tag{50}$$

2. Given the law of motion of x, compute the SDF $\Lambda\left(g',\eta'\middle|\, g,x\right)$

$$\Lambda\left(g',\eta'\middle|\, g,x\right) = \beta\left[\frac{x'\left(g'|g,x\right)e^{g'+\sigma(\eta')\eta'}}{x}\right]^{-\frac{1}{\psi}}\left[\frac{w\left(x',g'\right)e^{g'+\sigma(\eta')\eta'}}{\mathbb{M}_g e^{g'} w\left(\phi',g'\right)}\right]^{\frac{1}{\psi}-\gamma}$$

where w is defined in equation (15)

3. With the SDF, solve the following optimal contracting problem

$$v\left(u|g,x\right)=\max_{c,\{u'(z')\}}\left\{(1-c)+(1-\kappa)\int\Lambda\left(g',\eta'|g,x\right)e^{g'+\sigma(g')\eta'+\varepsilon'}v\left(u'\left(z'\right)|g',x'\right)\Omega(dz'|g)\right\}$$

$$s.t:\quad u=\left[(1-\beta)\,c^{1-\frac{1}{\psi}}+\beta\left(\mathbb{M}_g u'\right)^{1-\frac{1}{\psi}}\right]^{\frac{1}{1-\frac{1}{\psi}}},$$

$$\mathbb{M}_g u'=\left\{(1-\kappa)\int\left[e^{g'+\sigma(g')\eta'+\varepsilon'}u'\left(z'\right)\right]^{1-\gamma}d\Omega(z'|g)\right\}^{\frac{1}{1-\gamma}},$$

$$u'\left(z'\right)\geq\underline{u}\left(g'\right),\quad for\ all\ z',$$

$$v\left(u'\left(z'\right)\middle|\,g',x'\right)\geq 0,\ \ for\ \ all\ \ z'.$$

This step yields i) the value function $v\left(u|\,g,x\right)$ and the cutoff value of $\overline{u}\left(g,x\right)$ that satisfies $v\left(\overline{u}\left(g,x\right)|\,g,x\right)=0$; and ii) the policy functions $c\left(u|\,g,x\right)$ and $u'\left(u,g',\varepsilon'|\,g,x\right)$, which specifies the law of motion of u'. We solve the optimal contracting problem by value function iteration with 15 points for grid on u and 70 points for the grid on x by value function iteration.[1] To check the accuracy in computing the optimal contract, we plot a version of Euler equation errors in Figure 1. Fixing u,x,g and the aggregate state next period g', we draw 1000 idiosyncratic shocks ε' such that both agent and firm-side limited commitment constraints are not binding. We then use the maximum absolute log10 ratio of worker's MRS to owners' MRS across these shocks as our measure of Euler Equation Error. We repeat this procedure for different (u,x,g) and g' combinations with values of (u,x) that are not on the grid points where the value function is solved. The Euler equation errors computed this way has the magnitude of -4, which suggests that our approximation is reasonable.

4. We then use the optimal policies to simulate the model and update Γ_x. The details of the simulation procedure are given below:

[1] It turns out that the optimal policies for our problem can be obtained without any root solving by transforming the grid on u to a grid on the threshold $\underline{\varepsilon}$. We begin with a guess for $c(u|g,x)$ and $v(u|g,x)$. For a given g, x^j, and the threshold $\underline{\varepsilon}^j$, the compensation c^j, the optimal continuation values $u^j(\varepsilon')$ can be solved analytically. The promised keeping constraint then yields corresponding u^j associated $\underline{\varepsilon}^j$. Then new $c(u|g,x)$ and $v(u|g,x)$ are obtained by interpolation. We then iterate until the value function and consumption functions both converge with a tolerance of $1e-7$ under a sup norm.

(a) Let $\phi(t)$ denote the summary measure at time t. In simulations, we approximate the continuous distribution $\phi(t)$ by a finite-point distribution. A density ϕ is characterized by a set of grid points $\{u^*[n](t)\}_{n=1}^{N+2}$ and corresponding weights $\{\phi[n](t)\}_{n=1}^{N+2}$ such that

- $u^*[1]$ and $u^*[N+1]$ are the boundaries where the limited commitment constraint binds: $u^*[1] = \underline{u}(g_t)$ and $u^*[N+1] = \overline{u}(g_t, x_t)$. $u^*[N+2] = \alpha^{new} u^{FB}(g_t)$ is the restarting utility.
- $\{u^*[j]\}_{j=2,3,\cdots N}$ are the interior points: $u^*[j] \in (u_{j-1}, u_j)$, for $j = 2, 3 \cdots N$, are chosen appropriately to minimize the error that one could obtain by using a finite-point approximation of a continuous density ϕ. We explain this procedure in step 5. below
- $\phi[1]$ and $\phi[N+1]$ are the income shares of agents with a binding limited commitment constraint at $u^*[1]$ and $u^*[N+1]$, respectively.
- $\{\phi[j]\}_{j=2,3,\cdots N}$ are the income shares of agents in the interior and $\phi[N+2]$ is the income share of agents who entry the economy.

Clearly, we should have $\sum_{j=1}^{N+2} \phi[j](t) = 1$ for all t. In our application we use $N = 50$

(b) Starting with an initial distribution of u, denoted $\phi_0(u)$. For example, we can choose a point mass at $u_0(g)$ with $\underline{u}(g) < u(g) < \overline{u}(g, x)$ for every x, g, solve the following equation for x_0:

$$\int \phi_0(u)\, c(u|\, g, x_0)\, du = 1 - x_0$$

Now we have an initial condition (ϕ_0, x_0), and ϕ_0 is represented as $\{\phi[n](0), u^*[n](0)\}_{n=1}^{N+1}$, such that $\phi[m](0) = 1$, and $u^*[m](0) = u_0(g_0)$, where m is defined by $u_0(g) \in (u_{m-1}, u_m)$.

(c) Having solved x_0, we use the law of motion of $u'(u, g', \varepsilon'|\, g, x)$ to compute ϕ_1. Here we describe a general procedure to solve for $\{\phi[n](t+1);\ u^*[n](t+1);\ x_{t+1}\}_{n=1}^{N+2}$ given $\{\phi[n](t);\ u^*[n](t);\ x_t\}_{n=1}^{N+2}$. Note that the assumed law of motion gives a natural candidate for x_{t+1}. We denote

$x_{t+1} = \Gamma\left(x_t | \, g_t, g_{t+1}\right)$.

i. First, we approximate the distribution $f\left(\varepsilon | \, g\right)$ by a finite dimensional distribution such that $\sum_j^J f_g\left[j\right] = 1$ and $\sum_j^J e^{\varepsilon_j} f_g\left[j\right] = 1$, for $g = g_H, g_L$. In our application we use $J = 50$.

ii. Given $\left\{\phi\left[n\right]\left(t\right), u^*\left[n\right]\left(t\right)\right\}_{n=1}^{N+2}$ for period t, conditioning on the realization of aggregate state g_{t+1} , for each $n = 1, 2, \cdots, N+2$, we compute $\left\{\phi_{t+1}\left[n, j\right]\right\}_{n,j}$. The interpretation is that $\phi_{t+1}\left[n, j\right]$ is the total measure of income share that comes from agents with $u^*\left[n\right]\left(t\right)$ and with realization of ε_j, which is given by:

$$\phi_{t+1}\left[n, j\right] = (1-\kappa) f_{g_{t+1}}\left[j\right] \phi_t\left[n\right] e^{\varepsilon_j}, \quad j = 1, 2 \cdots, J.$$

Note that the continuation utility of these agents is $u'\left(u^*\left[n\right]\left(t\right), g_{t+1}, \varepsilon_j | \, g_t, x_t\right)$, a fact that we will use below.

iii. We now compute $\left\{\phi_{t+1}\left[m\right]\right\}_{m=1,2,\cdots N+2}$ for the next period.

$$\phi_{t+1}\left[1\right] = \sum_{n=1}^{N+2} \sum_{j=1}^{J} \phi_{t+1}\left[n, j\right] I_{\left\{u'\left(u^*[n](t), g_{t+1}, \varepsilon_j | g_t, x_t\right) \leq \underline{u}(g_{t+1})\right\}},$$

$$\phi_{t+1}\left[2\right] = \sum_{n=1}^{N+2} \sum_{j=1}^{J} \phi_{t+1}\left[n, j\right] I_{\left\{u'\left(u^*[n](t), g_{t+1}, \varepsilon_j | g_t, x_t\right) \in \left(u_1^{(t+1)}, u_2^{(t+1)}\right)\right\}},$$

$$\phi_{t+1}\left[m\right] = \sum_{n=1}^{N+2} \sum_{j=1}^{J} \phi_{t+1}\left[n, j\right] I_{\left\{u'\left(u^*[n](t), g_{t+1}, \varepsilon_j | g_t, x_t\right) \in \left[u_{m-1}^{(t+1)}, u_m^{(t+1)}\right)\right\}}, \quad m = 3, \cdots N$$

$$\phi_{t+1}\left[N+1\right] = \sum_{n=1}^{N+2} \sum_{j=1}^{J} \phi_{t+1}\left[n, j\right] I_{\left\{u'\left(u^*[n](t), g_{t+1}, \varepsilon_j | g_t, x_t\right) \geq \overline{u}(g_{t+1}, x_{t+1})\right\}},$$

$$\phi_{t+1}\left[N+2\right] = \kappa.$$

The interpretation is again that $\phi\left[1\right]$ and $\phi\left[N+1\right]$ are the income shares of agents with a binding limited commitment constraint at $u^*\left[1\right]$ and $u^*\left[N+1\right]$, respectively, and $\left\{\phi\left[j\right]\right\}_{j=2,3,\cdots N}$ are the income shares of agents in the interior. $\phi\left[N+2\right]$ is the income share of agents who entry the economy.

iv. We need to update the vector normalized utilities $\left\{u^*\left[n\right]\left(t+1\right)\right\}_{n=1}^{N+2}$.

Clearly, we should have $u^*[1](t+1) = \underline{u}(g_{t+1})$, $u^*[N+1](t+1) = \overline{u}(g_{t+1}, x_{t+1})$ and $u^*[N+2](t+1) = \alpha^{new} u^{FB}(g_{t+1})$. For $m = 2, \cdots, N$, we choose $u^*[m](t+1) \in \left[u_{m-1}^{(t+1)}, u_m^{(t+1)}\right)$ such that the resource constraint holds exactly for $u \in \left[u_{m-1}^{(t+1)}, u_m^{(t+1)}\right)$. That is, we pick $u^*[m](t+1)$ to be the solution (denoted u^*) to

$$\sum_{n=1}^{N+2} \sum_{j=1}^{J} \phi_{t+1}[n,j]\, c\left(u'\left(u^*[n](t), g_{t+1}, \varepsilon_j \middle| g_t, x_t\right) \middle| g_{t+1}, x_{t+1}\right) I_{\left\{u'\left(u^*[n](t), g_{t+1}, \varepsilon_j | g_t, x_t\right) \in \left[u_{m-1}^{(t+1)}, u_m^{(t+1)}\right)\right\}}$$

$$= c(u^* | g_{t+1}, x_{t+1}) \phi_{t+1}[m].$$

(d) Up to now, we have described a procedure to simulate forward the economy. This allows us to compute the "market clearing" $\left\{x_{t+1}^{MC}\right\}_{t=0}^{\infty}$ as follows:

$$x_{t+1}^{MC} = 1 - \sum_{m=1}^{N+2} c(u^*[m](t+1) | g_{t+1}, x_{t+1}) \phi_{t+1}[m]. \tag{51}$$

Given the sequence of $\{g_t\}_{t=1}^{T}$, we simulate the economy forward for T periods to obtain $\left\{x_t^{MC}\right\}_{t=0}^{T}$. We divide the sample into four cases: $g_H \to g_H$, $g_H \to g_L$, $g_L \to g_H$, $g_L \to g_L$ and use regressions (50) to update the law of motion of x. We go back to step 1 to iterate until the unconditional R^2 approaches 99.9%.

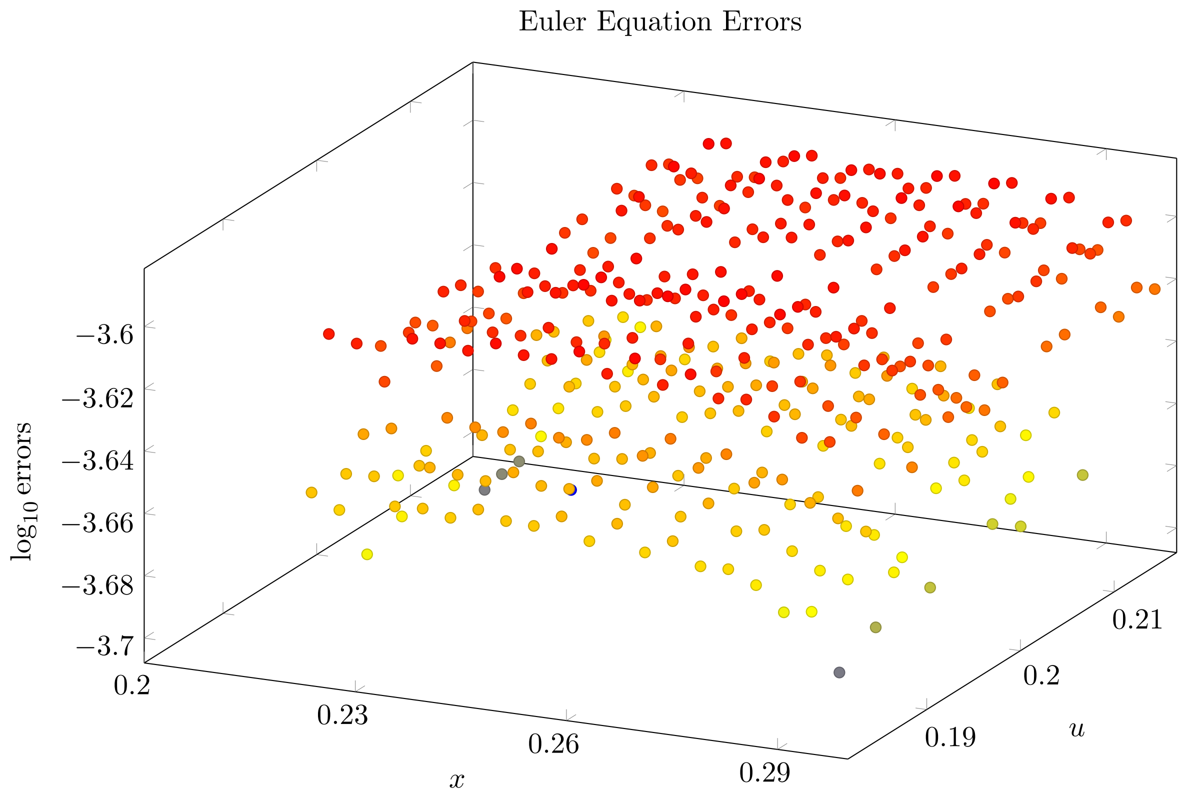

Figure 1: Euler equation errors for $g = g_L$ and $g' = g_L$.