Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 146 (2011) 845–862
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Self-insurance vs. self-financing: A welfare analysis of the persistence of shocks ☆

Francisco J. Buera [a,b], Yongseok Shin [c,d,*]

[a] University of California at Los Angeles, United States
[b] NBER, United States
[c] Washington University in St. Louis, United States
[d] Federal Reserve Bank of St. Louis, United States

Received 15 September 2009; final version received 28 September 2010; accepted 21 November 2010

Available online 2 February 2011

Abstract

We study the welfare cost of market incompleteness in a generalized Bewley model where idiosyncratic risk takes the form of entrepreneurial productivity shocks. Market incompleteness in our framework has two dimensions. First, in the Bewley tradition, only a limited set of instruments for consumption smoothing is available. Second, entrepreneurs' capital rental is subject to collateral constraints. As is well known, it is harder to self-insure against more persistent shocks, and the welfare cost of missing consumption insurance *increases* with shock persistence. On the other hand, with collateral constraints, an increase in shock persistence leads to better allocation of production factors through entrepreneurs' self-financing, and the welfare cost of imperfect capital rental markets *decreases* with shock persistence. The overall welfare cost of market incompleteness can be increasing, decreasing, or even non-monotone in shock persistence, depending on the relative strengths of its two components—the cost of missing insurance and the cost of imperfect capital markets.

JEL classification: D31; D52; D58; D91; E21; E44; L26

Keywords: Incomplete market; Self-insurance; Self-financing; Shock persistence

☆ We thank the associate editor, three anonymous referees, and the seminar participants at the Federal Reserve Bank of Minneapolis for many useful suggestions. Roberto Fattal Jaef provided excellent research assistance. Financial support of the National Science Foundation (SES-0820318 and 0946647) is gratefully acknowledged. The usual disclaimer applies.
* Corresponding author at: Washington University in St. Louis, United States.
E-mail addresses: fjbuera@econ.ucla.edu (F.J. Buera), yshin@wustl.edu (Y. Shin).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2011.01.003

Does market incompleteness matter for welfare? In this paper, we answer this question with a generalized Bewley model. Let us first be clear about what we mean by Bewley models. We follow Ljungqvist and Sargent [19] and use the term "Bewley models" to refer to the large—and still expanding—class of incomplete-market heterogeneous-agent models, inclusive of both exchange-economy and production-economy models.

Conventional Bewley models share a common structure. The model builder exogenously limits the set of instruments that individuals can use to smooth their consumption against idiosyncratic income fluctuations. The unavailability of explicit consumption insurance is often justified perfunctorily on the grounds of enforcement or information problems. However, on the production side—if there is one, all factor and output markets operate efficiently, and a representative firm is assumed. One can easily imagine the same enforcement or information problems that preclude consumption insurance contracts disrupting the production side of the economy as well.

We further develop this idea and introduce collateral constraints on capital rental that arise from imperfect enforceability of contracts. Now, our model has two distinct dimensions of market incompleteness: missing consumption insurance—as in conventional Bewley models—and imperfect capital rental markets.

Ours is a model with individual-specific technologies (entrepreneurship), where idiosyncratic risk takes the form of shocks to individuals' entrepreneurial productivity. We note that our model nests the conventional Bewley models: With perfectly-functioning capital rental markets, our economy is isomorphic to the one in Aiyagari [1].

We treat the persistence of idiosyncratic shocks as a parameter, and ask about the welfare cost of market incompleteness as the persistence varies, holding fixed the unconditional distribution of the idiosyncratic shocks. In our model, the welfare cost of market incompleteness is the sum of the cost of missing consumption insurance and the cost of imperfect capital markets.

In the absence of contracts contingent on their income risk, individuals accumulate a buffer stock of savings to smooth their consumption. This behavior is known as *self-insurance*. It is harder to self-insure against more persistent shocks, and hence the welfare cost of missing consumption insurance is *increasing* in shock persistence. This is a well-known result from the large literature building on Bewley [4]: See, for example, Deaton [9], Huggett [16], Aiyagari [1], and Levine and Zame [18]. This literature uses exchange economy models or models with frictionless production, and can only address the welfare cost of missing consumption insurance, which primarily manifests itself through the variance of individual consumption.

In our generalized Bewley model, we obtain a new result: The welfare cost of imperfect capital rental markets is *decreasing* in shock persistence. The reason is as follows. Being subject to collateral constraints in capital rental markets, entrepreneurs with profitable projects need to self-finance their production. When entrepreneurial productivity shocks are persistent, a constrained entrepreneur with a profitable project keeps it for longer on average, and is given the time to save up enough collateral and self-finance his growing, profitable operation. From the macroeconomic perspective, with persistent shocks, *self-financing* is an effective substitute for well-functioning capital rental markets in terms of allocating production factors. The higher allocative efficiency owing to more persistence leads to more aggregate output for a given level of aggregate capital—i.e., the aggregate total factor productivity (TFP) of the economy is higher. The mean consumption is commensurately higher, and the welfare cost of imperfect capital rental markets is lower. By contrast, when the entrepreneurial productivity shocks are not very persistent, constrained entrepreneurs with good projects are more likely to lose their profitable production opportunities before they can save up enough capital for self-financing. For the macroeconomy, without self-financing effectively making up for the imperfect capital market, the welfare cost of imperfect capital rental markets remains large.

How the overall welfare cost—the sum of the cost of missing insurance and the cost of imperfect capital markets—responds to shock persistence depends on the relative strengths of these two countervailing forces.

When the capital rental market functions perfectly, the overall welfare cost is equal to the cost of missing consumption insurance. Thus, the welfare cost of market incompleteness is increasing in shock persistence, re-affirming the well-known result in the literature on self-insurance. On the other hand, when the collateral constraints on capital rental are restrictive enough, the welfare cost of imperfect capital markets dominates the cost of missing insurance: The overall welfare cost of market incompleteness is decreasing in shock persistence. With intermediate degrees of collateral constraints on capital rental, the overall welfare cost of market incompleteness has a U shape when plotted against the persistence of idiosyncratic shocks. The possibility that the welfare cost of market incompleteness can be decreasing in shock persistence has hitherto not been obtained within the standard Bewley framework.

In Section 2.2.4, we also consider a narrower but more practical notion of the welfare cost of missing consumption insurance. Starting with an economy with missing consumption insurance, we measure the welfare gains from the introduction of perfect "ex post" consumption insurance that respects and hence perpetuates the initial inequality. We find that the welfare gains from ex post insurance are much smaller than those from the "veil of ignorance" insurance.

The early debates on the welfare cost of uninsurable idiosyncratic risk have given rise to a large literature focusing on the persistence of individual income processes.[1] Our analysis shows that the persistence of shocks has different welfare implications depending on whether income itself is the shock or income has an endogenous component that responds to underlying shocks on productivity or opportunities. While the primary goal of our paper is to provide a clear exposition of the welfare consequences of shock persistence in generalized Bewley models, one can also find a practical implication: If shocks are highly persistent, the provision of consumption insurance should be the priority for policymakers aiming to improve welfare; if shocks are not that persistent, they will need to first address the frictions in the production side capital markets.

0.1. Related literature

Given the vastness of the literature using Bewley models and also the purpose of our paper, we do not attempt to review all contributions in this area. We instead refer interested readers to the excellent survey of Heathcote et al. [14].

In the traditional Bewley literature, the source of heterogeneity and hence inequality is invariably labor income, which is assumed to be purely exogenous. Recently, researchers have emphasized that individual income dynamics contains endogenous components that reflect individual decisions on occupation, job search, labor supply, and the accumulation of assets and human capital. In this context, we highlight the individual-specific technologies (entrepreneurship) in our model, which depend not only on individuals' shock realizations but also on their saving decisions when capital rental markets are imperfect. Furthermore, we abstract from individual labor productivity shocks. This abstraction, while remediable in a straightforward manner, is not at all important for the main point of our paper—the fact that the welfare cost of missing

[1] The debates have not been completely resolved, and there seem to remain substantial disagreements regarding the persistence of shocks. For a recent coverage, see Guvenen [11,12] and references therein.

consumption insurance and the welfare cost of imperfect capital rental markets respond to shock persistence in opposite directions.

The papers that are most closely related to ours are Quadrini [21] and Cagetti and De Nardi [6]. They incorporate collateral constraints into models with entrepreneurship, and show that these elements can explain the observed wealth inequality in the United States. Intuitively, if there are collateral constraints, talented entrepreneurs will self-finance and hold a large ownership stake in their own businesses, which translates into a fat right tail of the wealth distribution. We build on their analysis and study the welfare cost of market incompleteness in this class of models, with a particular emphasis on the role of the persistence of idiosyncratic shocks.[2]

Yet another paper deserves further discussion. Heathcote et al. [13] show that an increase in idiosyncratic labor productivity risk may have a positive welfare effect in an incomplete market model with endogenous labor supply, as it presents an opportunity to raise aggregate productivity by concentrating market work among more productive workers. Despite a superficial similarity, there are a number of important differences. Foremost, Heathcote et al.'s model considers a permanent, "fixed-effect" shock that is uninsurable whatsoever and a transitory shock that is completely insured.[3] As such, there is no self-insurance. More to the point, their welfare analysis focuses on the variance of these shocks, but is not applicable to shock persistence. By contrast, our analysis focuses on the persistence of the one and only shock in the model, which is uninsured except by self-insurance, while holding fixed its entire unconditional distribution.[4]

1. Model

We propose a model with individual-specific technologies (entrepreneurship) and imperfect capital rental markets. Individuals are heterogeneous with respect to their entrepreneurial productivity and wealth. In each period, individuals choose either to operate an individual-specific technology or to work for a wage. Imperfection in the capital rental market is modeled with a collateral constraint that is proportional to an individual's financial wealth. As in the conventional Bewley models, there is no explicit consumption insurance, and individuals can only use a risk-free asset to self-insure against idiosyncratic risk.

1.1. Heterogeneity and demographics

Individuals live indefinitely, and are heterogeneous with respect to their wealth a and their entrepreneurial productivity e, with the former being chosen endogenously by forward-looking saving decisions. An individual's entrepreneurial productivity follows a stochastic process. One can think of the entrepreneurial productivity shock as an arrival of a new technology making previous production processes obsolete or as a shift in relative demand.

[2] We here use the framework developed in Buera and Shin [5], which studies the macroeconomic transition dynamics following a large-scale reform that creates the need for the reallocation of resources among heterogeneous producers, with an emphasis on the role of production-side financial frictions in delaying such reallocation. That paper also provides more technical details on the model properties and computation.

[3] In their model, for standard utility specifications and parameter values, only the increase in the completely-insured transitory risk has the welfare-enhancing effect. See Propositions 1 and 1a of Heathcote et al. [13].

[4] In addition, in their model, labor is the only input of production, and the production side is frictionless. One important advantage of their analysis is that it permits a closed-form expression for the welfare cost. Our richer environment—with asset accumulation, collateral constraints, and occupation choice—denies us such tractability and lucidity.

In our model, an individual with a given e at the beginning of the period faces a non-stochastic return to his entrepreneurial activity for the period, although his entrepreneurial productivity may change at the end of the period. That is, the e-shock only affects the future profitability of his entrepreneurial activities, and not the return from the current capital and labor input choices. We abstract from high-frequency, temporary shocks that would affect the risk of entrepreneurial investment but not its expected profitability.[5] In addition, we abstract from multi-stage investment projects of Holmström and Tirole [15], in whose model transitory liquidity shocks can adversely affect the ability of constrained entrepreneurs to complete a project.

We denote by $P(e, e')$ the transition function of the Markov process governing the evolution of entrepreneurial productivity. We denote by $G_t(a, e)$ the cumulative density function for the joint distribution of entrepreneurial productivity and wealth at the beginning of period t. The population size of the economy is normalized to one, and there is no population growth.

1.2. Preferences

Individuals discount their future utility using the same discount factor β. The preferences over contingent plans for the consumption sequence from the point of view of an individual in period t are represented by $\mathbb{E}_t \sum_{s=t}^{\infty} \beta^{s-t} u(c_s)$.

1.3. Technologies

In any given period, individuals can choose either to work for a wage or to operate an individual-specific technology. We label the latter option as entrepreneurship. We assume that an entrepreneur with productivity e who uses k units of capital and hires l units of labor produces according to a production function $f(e, k, l)$, which is assumed to be strictly increasing in all arguments and strictly concave in capital and labor, with $f(0, k, l) = 0$.

1.4. Financial markets

A risk-free one-period bond and physical capital are the only financial and real assets in the economy. A perfectly-competitive financial intermediary sells bonds backed by the physical capital stock to households, and rents out capital to entrepreneurs. Individuals may also issue bonds (private IOUs) to other individuals subject to quantity restrictions. This quantity limit can be zero (no borrowing allowed for intertemporal consumption smoothing), and actually will be assumed to be zero in our analysis. It is assumed that all the bonds are repaid with probability one, and that their payoff is not state-contingent, making them risk-free. In equilibrium, the net demand for bonds by individuals is equal to the stock of physical capital. The risk-free return on bonds—the risk-free interest rate in the economy—is r. The zero-profit condition of the intermediary implies that the rental cost of capital is $r + \delta$, where δ is the depreciation rate.

We assume that entrepreneurs' capital rental k is limited by a collateral constraint $k \leqslant \lambda a$, where a is individual financial wealth and λ measures the degree of credit frictions, with $\lambda = \infty$ corresponding to perfect capital rental markets, and $\lambda = 1$ to financial autarky where all capital

[5] See Covas [8] and Angeletos [2] for models with transitory, uninsurable investment risks.

needs to be self-financed by entrepreneurs. The same λ applies to everyone in a given economy.[6] Our specification captures the common prediction from models of limited contract enforcement: The amount of credit is limited by entrepreneurs' wealth. This specification has been widely used in the literature on financial frictions and entrepreneurship [10].

We note that ours is a model with fungible capital. Apart from the collateral constraint, there is no other friction in capital rental, such as irreversibility or adjustment costs of investment. With $\lambda = 1$, however, there is no rental market, and one can interpret this financial autarky alternatively as a model with individual-specific capital. In this case, unused capital—as it cannot be rented out to producers—simply depreciates at rate δ.[7]

In this paper we focus on within-period borrowing, or capital rental, for production purposes. We do not allow borrowing for intertemporal consumption smoothing in our model, which translates into $a \geqslant 0$. This no-borrowing constraint—in the language of Bewley models—will only bind for individuals who choose to be workers, and has no direct bearing on the behavior of entrepreneurs, who will need to hold financial assets to overcome the collateral constraint. By not allowing individuals to borrow for consumption smoothing, we are obtaining the upper bound on the welfare cost of missing consumption insurance.

1.5. Individuals' problem

The recursive representation of the problem of an individual is given by the following Bellman equation:

$$v(a,e) = \max_{c,a' \geqslant 0} u(c) + \beta \mathbb{E}\big[v\big(a',e'\big)\big|e\big]$$
$$\text{s.t. } c + a' \leqslant \max\big\{w, \pi(a,e;w,r)\big\} + (1+r)a, \tag{1}$$

where the equilibrium wage w and interest rate r are given, and $\pi(a,e;w,r)$ is the profit from operating an individual technology. This indirect profit function is defined as:

$$\pi(a,e;w,r) = \max_{l,k \leqslant \lambda a} \big\{f(e,k,l) - wl - (\delta + r)k\big\}.$$

The input demand functions are denoted by $l(a,e;w,r)$ and $k(a,e;w,r)$, and the collateral constraint $k \leqslant \lambda a$ is taken into account.

The max operator in the budget constraint stands for the occupation choice. An individual with current wealth a and entrepreneurial productivity e will choose to be an entrepreneur only if profits as an entrepreneur, $\pi(a,e;w,r)$, exceed labor income as a wage earner, w. More formally, an individual with productivity e decides to be an entrepreneur if his current wealth a is higher than the threshold wealth $\underline{a}(e)$, where $\underline{a}(e)$ solves:

[6] Our collateral constraint can be derived from the following limited enforcement problem. Consider an individual with financial wealth $a \geqslant 0$ deposited in the financial intermediary at the beginning of a period. Assume that he rents k units of capital. Then he can abscond with fraction $1/\lambda$ of the rented capital. The only punishment is that he will lose his financial wealth a deposited in the intermediary. In particular, he will not be excluded from any economic activity in the future: He is even allowed to instantaneously deposit the stolen capital k/λ and continue on as a worker or an entrepreneur. (This assumption is essential for obtaining the simple static collateral constraint. If there are any dynamic considerations, the constraint will also depend on the shock realization and its persistence.) In the equilibrium, the financial intermediary will rent capital only to the extent that no individual will renege on the rental contract: $k/\lambda \leqslant a$.

[7] We have worked out numerical exercises with $\lambda = 1$. The results are qualitatively similar to what is reported in Section 2.2 ($\lambda = 1.5$), although the magnitude of the effect of market incompleteness is substantially larger with $\lambda = 1$.

$$\pi\big(\underline{a}(e), e; w, r\big) = w.$$

That is, with imperfect capital rental markets, individuals with a given entrepreneurial productivity will choose to become entrepreneurs only if they are wealthy enough to overcome the collateral constraint and run their businesses at a profitable scale.[8] Note that, with perfect capital rental markets ($\lambda = \infty$), entrepreneurial productivity e is the sole determinant of an individual's occupation.

As we make the assumption that individuals are equally productive as workers, the occupation choice serves as an avenue of insurance from the downside risk of the stochastic entrepreneurial productivity. However, in spite of the occupation choice, the welfare cost of missing consumption insurance in our model is quite large—especially when shocks are highly persistent—compared to the results in the existing literature (Section 2.2.1). In any case, this assumption is not at all important for our main result that the welfare cost of missing consumption insurance and the welfare cost of imperfect capital rental markets respond to shock persistence in opposite directions.

1.6. Stationary competitive equilibrium

A stationary competitive equilibrium consists of the invariant distribution of entrepreneurial productivity and wealth $G_\infty(a, e)$, policy functions $\{c(a, e), a'(a, e), l(a, e), k(a, e)\}$, and prices $\{w, r\}$ such that:

1. $\{c(a, e), a'(a, e), l(a, e), k(a, e)\}$ solve the individuals' problem in (1), given $\{w, r\}$.
2. The labor, capital, and goods markets all clear:

$$\iint_{\underline{a}(e)}^{\infty} l(a, e) G_\infty(da, de) = \int G_\infty\big(\underline{a}(e), de\big), \qquad \text{(Labor Market)}$$

$$K \equiv \iint_{\underline{a}(e)}^{\infty} k(a, e) G_\infty(da, de) = \iint a G_\infty(da, de), \qquad \text{(Capital Market)}$$

$$\iint c(a, e) G_\infty(da, de) + \delta K = \iint_{\underline{a}(e)}^{\infty} f\big(e, k(a, e), l(a, e)\big) G_\infty(da, de). \qquad \text{(Goods Market)}$$

3. The invariant distribution $G_\infty(a, e)$ satisfies the equilibrium mapping:

$$G_\infty(a, e) = \iint_{a'(\hat{e}, \hat{a}) \leqslant a} \int_{\tilde{e} \leqslant e} P(\hat{e}, d\tilde{e}) G_\infty(d\hat{a}, d\hat{e}).$$

[8] Obviously, an individual's e may be so low that he will never choose to be an entrepreneur. In this case, $\underline{a}$ should be thought of as ∞.

2. Quantitative exploration

In quantifying our theory, we interpret our entrepreneurial technology as an establishment in the data. We assume that the US economy is the benchmark with perfect capital markets ($\lambda = \infty$), and pin down parameter values using relevant moments of the US data. In particular, the distribution of entrepreneurial productivity can be inferred from the size distribution of establishments. Section 2.1 describes in detail our calibration strategy.

We then ask about the welfare cost of market incompleteness as we vary the persistence of shocks, holding fixed their unconditional distribution. Section 2.2 reports the results.

2.1. Calibration

We first describe the parametrization of the model, and then discuss the calibration of the parameters. For the sake of clarity, we choose a parsimonious parametrization that follows as much as possible the standard practices in the literature.

We choose a period utility function that has a constant relative risk aversion:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}.$$

We assume that an entrepreneur with productivity level e who hires k units of capital and l units of labor produces according to the following production function:

$$f(e,k,l) = e\left(k^{\alpha} l^{1-\alpha}\right)^{1-\nu},$$

where ν is the share of output going to the entrepreneur, and $1-\nu$ is the span-of-control parameter [20]. Accordingly, $1-\nu$ represents the share of output going to the variable factors. Out of this, fraction α goes to capital, and $1-\alpha$ goes to labor.

The entrepreneurial productivity e is assumed to follow a discretized version of an autoregressive process with normal innovations, $\log e' = \rho \log e + \varepsilon$, with $\varepsilon \sim N(0,\sigma^2)$. In particular, we approximate this autoregressive process with a 40-state Markov chain following the procedure of Tauchen [22]. For our welfare analysis, we vary the shock persistence parameterized by ρ, while we hold fixed the invariant distribution of this Markov process.

We now need to specify six parameter values: technological parameters α, ν, and the depreciation rate δ; the parameter governing the dispersion of entrepreneurial productivity σ; the subjective discount factor β and the coefficient of relative risk aversion γ.

We let $\gamma = 1.5$ following the standard practice. The one-year depreciation rate is set at $\delta = 0.06$. We impose $\alpha(1-\nu) = 0.3$ to match the aggregate income share of capital.

We are thus left with three parameters: ν, σ, and β. We choose them so that the model economy with perfect capital rental markets ($\lambda = \infty$) matches relevant moments of the US data on income concentration, establishment size (defined as the number of employees), and real interest rate.

We pick $\nu = 0.2$ and $\sigma^2/(1-\rho^2) = 0.39$—the unconditional variance of the log entrepreneurial productivity—to match the earnings share of the top five per cent of earners in the US, 0.30 [6], and the fraction of labor employed by the largest ten per cent of establishments, 0.63 according to the US Census. In particular, we can infer ν from the fraction of total income accounted for by the top five per cent of earners in the population: Top earners are mostly entrepreneurs both in the data and in our model, and ν controls the share of income going to the entrepreneurial input.[9] Given the span-of-control parameter $1-\nu$, the dispersion of the entrepreneurial productivity can be chosen to match the right tail of the establishment size distribution.

Finally, we choose $\beta = 0.92$ so that the annual interest rate is four per cent.[10]

2.2. Results

We now explore the impact of the persistence of shocks—holding fixed their unconditional distribution—on the welfare cost of market incompleteness. In our generalized Bewley model, the overall welfare cost of market incompleteness has two components: the cost of missing consumption insurance and the cost of imperfect capital rental markets. It will be shown that the former is increasing in shock persistence—as is well known in the literature on self-insurance, while the latter is decreasing in shock persistence.

To isolate the welfare cost of missing consumption insurance, we start with an economy with perfect capital rental markets ($\lambda = \infty$) where individuals can only use a risk-free asset to self-insure against the idiosyncratic shock. The welfare difference between this economy and the complete market benchmark—both with perfect capital rental markets—will be interpreted as the welfare cost of missing consumption insurance, which, we show, increases with shock persistence. In making this comparison, the two economies have the same initial condition, and our welfare calculations correctly account for the transition phase.[11]

To compute the welfare cost of imperfect capital rental markets, we construct an economy with restrictive collateral constraints ($\lambda = 1.5$).[12] The welfare difference between this economy and the economy with perfect capital rental markets ($\lambda = \infty$)—both lacking consumption insurance—will be interpreted as the welfare cost of imperfect capital rental markets, which, we show, decreases with shock persistence. Again, these two economies start from the same initial condition, and our welfare calculations correctly account for the transition phase.

The incomplete market economy with $\lambda = 1.5$ will also be compared to the complete market benchmark with perfect capital rental markets, and the welfare difference between the two will be interpreted as the overall welfare cost of market incompleteness. How this overall welfare cost varies with shock persistence depends on the relative strengths of the two countervailing forces—the cost of missing consumption insurance and the cost of imperfect capital rental markets.

---

9 In our calibration, slightly more than seven per cent of the individuals are entrepreneurs. This turns out to be quite close to the eight per cent in the data [6].

10 To pin down β that corresponds to a given interest rate target, we need to specify a value for ρ. Here, we picked $\rho = 0.98$. This is in accordance with the consensus within the standard Bewley literature that idiosyncratic income shocks are quite persistent, although there is no reliable estimate for the persistence of entrepreneurial income shocks that we can use specifically for our model. Anyways, the choice of $\rho = 0.98$ here does not affect our quantitative analysis. Note that in our analysis ρ is the parameter that we experiment with to figure out its impact on welfare. We also considered an alternative calibration strategy where β is re-calibrated for each value of ρ. Our main results were not sensitive to this alternative strategy at all. In the end, we decided to keep β constant in order to make the whole exercise more transparent and comparable.

11 We consider two alternative ways of introducing consumption insurance. First, we can provide perfect consumption insurance from behind the veil of ignorance, which involves redistribution and the elimination of inequality. Second, we can provide perfect ex post consumption insurance that respects and hence perpetuates the existing inequality. We first consider the veil of ignorance insurance as it is more closely related to the earlier tradition of the Bewley literature. In Section 2.2.4, we consider the ex post insurance.

12 We pick $\lambda = 1.5$ as it yields an equilibrium external finance to GDP ratio of a typical developing economy, 0.6–0.8 according to Beck et al. [3].

Our metric of welfare is the perfectly-egalitarian measure of economy-wide wellbeing, which is a natural choice given that all individuals have the same preferences. The welfare costs we report are in units of permanent consumption compensation that is necessary to make an individual indifferent between inhabiting an economy with improved capital rental and/or consumption insurance markets and remaining in the particular stationary equilibrium. Appendix A explains in more detail our welfare cost formulae.

2.2.1. Welfare cost of missing consumption insurance

We construct an economy with perfect capital rental markets ($\lambda = \infty$). There is no explicit consumption insurance, and individuals can only self-insure using a risk-free asset. With perfect capital rental markets, individuals' occupations and earnings are determined only by their exogenously-evolving entrepreneurial productivity, and our model can be recast as an economy of Aiyagari [1], whose production side is frictionless.

Starting from the stationary equilibrium of this incomplete market economy, we then compute the welfare gains from eliminating all consumption inequality once and for all. This welfare difference, measured in units of permanent consumption compensation, is interpreted as the welfare cost of missing consumption insurance. In fact, here we are doing more than merely completing the markets: We eliminate all consumption inequality through redistribution, as if the consumption insurance were provided behind the veil of ignorance. While this "ex ante" insurance provision helps us better relate to the existing literature on consumption insurance, it clearly is an upper bound on the gains from consumption insurance. In Section 2.2.4, we consider the provision of consumption insurance that respects the existing wealth inequality, and examine its welfare impact as we vary shock persistence.

Next, we vary the persistence of shocks—holding fixed their unconditional distribution—and trace the welfare cost of missing consumption insurance, depicted with a dotted line in the left panel of Fig. 1. It is increasing in shock persistence. This is a reproduction of the well-known result that the welfare loss from missing consumption insurance is larger for more persistent shocks because they are more difficult to self-insure against [9,16,1,18]. Reflecting this difficulty, the dispersion of individual consumption in the stationary equilibrium of this incomplete market economy increases with shock persistence, as shown in the center panel of Fig. 1 with a dotted line.

Note that the welfare cost of missing consumption insurance in our model is much larger than in the typical Bewley models, despite a measure of insurance against the downside risk provided by individuals' occupation choice. One reason is that we ruled out borrowing for consumption purposes ($a \geqslant 0$). Another is our calibration strategy focusing on the right tail of the earnings distribution that mostly captures entrepreneurial income. Our implied earnings inequality is somewhat larger than estimates based on labor income alone.[13]

The stationary equilibrium of a standard incomplete market model has a larger aggregate capital stock and hence larger aggregate consumption than its complete market counterpart—as long as the capital stock is below the golden-rule level. Aiyagari [1] is a well-known example, and Huggett and Ospina [17] provide a thorough analysis on this phenomenon. The dotted line in the right panel of Fig. 1 is the mean consumption in the incomplete market stationary equilibrium with $\lambda = \infty$ relative to the mean consumption in the complete market stationary equilibrium for given shock persistence. The dotted line lies above one.

[13] See, for example, Castañeda et al. [7] for more on this point.

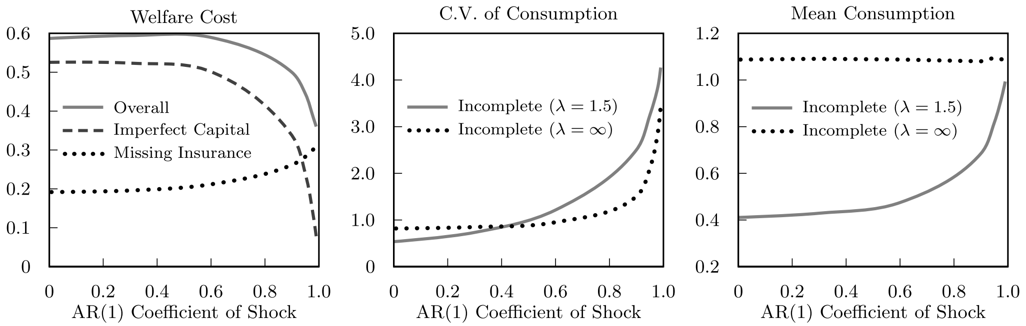

Fig. 1. In the left panel, the solid line is the welfare difference between the $\lambda = 1.5$ incomplete market economy and the complete market economy (overall cost of market incompleteness). The dashed line is the welfare difference between the $\lambda = 1.5$ economy and the incomplete market economy with perfect capital markets (cost of imperfect capital rental markets). The dotted line is the welfare difference between the incomplete market economy with perfect capital markets and the complete market economy (cost of missing consumption insurance). The center panel reports the coefficient of variation of consumption in the stationary equilibria of two incomplete market economies: one with $\lambda = 1.5$ and the other with perfect capital rental markets. The right panel shows their mean consumption relative to the complete market economy. All quantities are computed for given levels of shock persistence (horizontal axis).

2.2.2. *Welfare cost of imperfect capital rental markets*

We now construct an economy with a high degree of imperfections in the capital rental market: $\lambda = 1.5$. Individuals can still only self-insure using a risk-free asset. Starting from the stationary equilibrium of this economy, we then compute the welfare gains of making the capital rental market perfect (i.e., $\lambda = \infty$) once and for all, correctly accounting for the transition phase. We leave the missing consumption insurance as is. This welfare difference is interpreted as the welfare cost of imperfect capital rental markets.

We compute the welfare cost of imperfect capital rental markets for various levels of shock persistence—holding fixed the unconditional distribution—and plot it with a dashed line in the left panel of Fig. 1. We first note that the welfare cost of imperfect capital rental markets can be substantially larger than the welfare cost of missing consumption insurance. Furthermore, it is decreasing in shock persistence.

Why does the adverse impact of capital rental market imperfections decrease with more persistent shocks?

With restrictive collateral constraints on capital rental, entrepreneurs with profitable projects need to self-finance their production. For constrained entrepreneurs, accumulated assets not only help smooth their consumption over time, but also serve as collateral that enables them to rent more capital and increase their scale of production and profits over time. The persistent shock process implies that, probabilistically speaking, these entrepreneurs will keep their profitable projects for a prolonged period, and are likely to successfully increase their production scale and profits over time through self-financing. From the macroeconomic perspective, when entrepreneurial productivity shocks are persistent, there are more productive entrepreneurs operating at more efficient scales than will be the case with low persistence shocks. That is, more persistent shocks lead to better allocation of entrepreneurial talent and capital for production purposes.

This higher allocative efficiency shows up as a higher level of aggregate productivity in the economy. The left panel of Fig. 2 shows that the measured aggregate TFP in the stationary equilibrium of the $\lambda = 1.5$ incomplete market economy increases with shock persistence. As a result,

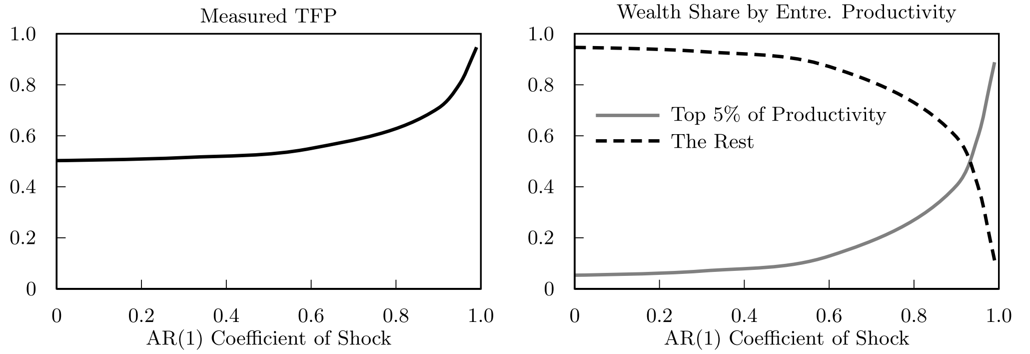

Fig. 2. The left panel shows the measured TFP in the stationary equilibrium of the $\lambda = 1.5$ incomplete market economy for given levels of shock persistence. TFP is computed as $Y/(K^{0.3}L^{0.7})$, and then normalized by the TFP of the complete market benchmark. In the right panel, for each $\lambda = 1.5$ stationary equilibrium, we report what fraction of wealth is held by those in the top five per cent of the entrepreneurial productivity distribution (solid line). By construction, the solid and the dashed lines add up to one.

aggregate output and consumption are commensurately higher when shocks are more persistent. See the solid line in the right panel of Fig. 1, which is the mean consumption in the stationary equilibrium of the $\lambda = 1.5$ economy relative to the mean consumption with complete markets.

In this context, one can conclude that, when shocks are persistent, *self-financing* is an effective substitute for efficient capital rental markets for the purpose of allocating production factors, and the welfare cost of imperfect capital rental markets is small.

By contrast, when the shocks are not very persistent, constrained entrepreneurs with good projects are more likely to lose their profitable production opportunities before they can save up enough assets for self-financing. From the macroeconomic perspective, without self-financing effectively making up for inefficient capital rental markets, the welfare cost of imperfect capital rental markets remains large.

We now provide more direct and quantitative evidence on the more efficient resource allocation—through self-financing—with more persistent shocks. We compute what fraction of the aggregate wealth is held by those belonging to the top five per cent of the entrepreneurial productivity distribution in the stationary equilibrium of the $\lambda = 1.5$ economy. As shown in the right panel of Fig. 2 (solid line), the more persistent the shocks are, the more of the aggregate wealth belongs to this group. When shocks are very persistent, these individuals are most likely to be active entrepreneurs, and self-finance with their own wealth up to two-thirds ($1/\lambda$) of their operation.[14] Their wealth enables them to scale up their production and operate closer to the maximal-profit scale. With less persistent shocks, these high-productivity individuals own much smaller fractions of the aggregate wealth. This implies that they are more likely to be constrained and operate at inefficient scales, if at all—they are much less likely to be active entrepreneurs when shocks are not very persistent.

We now ask what explains such vast differences in the wealth accumulation of those in the top five per cent of the entrepreneurial productivity distribution across the $\lambda = 1.5$ stationary

[14] Recall that, in our model with imperfect capital rental markets, an individual's occupation choice depends on his wealth as well as productivity. With perfect capital rental markets, his entrepreneurial productivity alone determines his occupation.

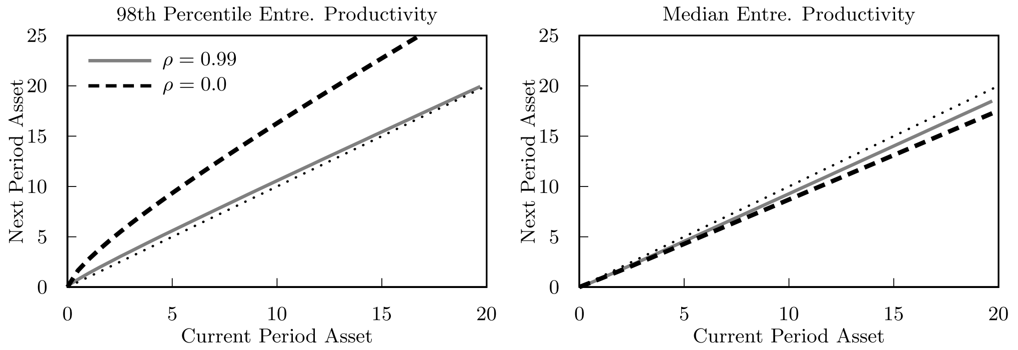

Fig. 3. The left panel shows the asset accumulation decision rule of an individual in the ninety-eighth percentile of the entrepreneurial productivity distribution. The solid line is for the $\lambda = 1.5$ stationary equilibrium with highly persistent shocks ($\rho = 0.99$), and the dashed line is for the one with serially uncorrelated shocks ($\rho = 0.0$). The right panel is for an individual with the median entrepreneurial productivity. The dotted lines have 45-degree slopes.

equilibria indexed by shock persistence. Is it primarily because more persistent shocks allow on average longer durations of profitable entrepreneurship opportunities? Or, is it because these individuals save a larger fraction of their income when shocks are more persistent? The answer turns out to be the former.

The left panel of Fig. 3 is the asset accumulation decision rule of an individual in the ninety-eighth percentile of the entrepreneurial productivity distribution. His decision rule maps the current asset level (horizontal axis) into the next period asset level (vertical axis), in the $\lambda = 1.5$ stationary equilibrium with $\rho = 0.99$ (solid line) and with $\rho = 0$ (dashed line). The dotted line has a 45-degree slope. As is clearly shown, this individual actually saves more when shocks are less persistent. This is primarily explained by the permanent income hypothesis: Consumption smoothing dictates that one should save a larger fraction of a transitory positive income shock than of a persistent positive income shock.[15]

We conclude that it is the law of probability, not the differences in saving behavior per se, that explains the wealth accumulation and successful self-financing of the highly-productive entrepreneurs in economies with persistent shocks.[16]

In summary, with imperfect capital rental markets, persistent shocks give individuals with profitable projects long enough a window of opportunities to overcome the collateral constraints through self-financing and operate their businesses at more efficient scales over time. From the macroeconomic perspective, with persistent shocks, self-financing is effectively substituting for

[15] There are other forces—of secondary importance—that lie behind this result, and they do not necessarily point to the same direction. The equilibrium wage and capital rental rate are lower in the stationary equilibrium with less persistent shocks ($\rho = 0$). This implies that an active entrepreneur with given productivity and wealth will earn more profits, although he earns less interest income. In addition, the lower interest rate somewhat mitigates, via the standard intertemporal substitution channel, the saving efforts.

[16] Fig. 3 also shows the decision rule of an individual of the median entrepreneurial productivity. He dis-saves by more when shocks are less persistent (dashed line). Again, consumption smoothing dictates that one should not reduce consumption as much in response to a transitory negative income shock as in response to a persistent one. In addition, the lower equilibrium wage and interest rate implies that this individual—who is most likely to be a worker—makes lower labor and interest income than in an economy with more persistent shocks. The lower interest rate also abets his dis-saving via the standard intertemporal substitution channel.

well-functioning capital rental markets in terms of resource allocation, and thereby reduces the welfare cost of imperfect capital rental markets.

2.2.3. Overall welfare cost of market incompleteness

In our generalized Bewley model, the overall cost of market incompleteness is the sum of the cost of missing consumption insurance and the cost of imperfect capital rental markets. In Sections 2.2.1 and 2.2.2, we find that, while persistent shocks undermine individuals' ability to self-insure, they foster allocative efficiency through self-financing. With the cost of missing consumption insurance and the cost of imperfect capital rental markets responding in opposite directions to shock persistence, it is the relative strengths of these two countervailing forces that determine how the overall welfare cost varies with shock persistence.

In our framework, their relative strengths can be transparently parameterized by λ. For example, if λ is a very large number, and hence the capital rental markets are nearly perfect, the overall welfare cost is almost identical to the cost of missing consumption insurance, and the overall welfare cost of market incompleteness will be increasing in shock persistence. On the other hand, if λ is very close to one—that is, capital rental markets are almost shut down, the cost of imperfect capital markets dominates, and the overall welfare cost of market incompleteness will be decreasing in shock persistence. Indeed, with $\lambda = 1.5$, the overall welfare cost is already decreasing in shock persistence, as shown with a solid line in the left panel of Fig. 1.[17]

With intermediate degrees of frictions in the capital rental market, the effect of shock persistence on the overall welfare cost of market incompleteness can be non-monotone. Indeed, for $\lambda = 7.5$, we find that the overall welfare cost has a U shape when plotted against the persistence of idiosyncratic shocks (not shown). The right and upward arm of the U is driven by the difficulty of self-insurance against very persistent shocks, while the left and downward arm is driven by the difficulty of overcoming collateral constraints on capital rental through self-financing when individuals' profitable entrepreneurial opportunities are short-lived.

2.2.4. Consumption insurance that preserves initial inequality

In Sections 2.2.1 and 2.2.3, consumption insurance eliminated all inequality as if the insurance were provided behind the veil of ignorance. Without redistributive policies, however, complete markets can only provide insurance conditional on the initial state of individuals—i.e., their wealth and entrepreneurial productivity.

We first construct the stationary equilibrium of the $\lambda = \infty$ incomplete market economy, and introduce ex post consumption insurance that respects (and hence perpetuates) the existing inequality based on individuals' wealth and entrepreneurial productivity. The welfare gains are shown in Fig. 4 with a dotted line. This is the cost of missing ex post consumption insurance. Contrary to our results in Section 2.2.1, the cost of missing insurance is now decreasing in shock persistence. Recall that we hold fixed the unconditional distribution of shocks as we vary shock persistence. This implies that the conditional variance of shocks is getting smaller as shocks become more persistent. Therefore, conditioning on the initial state, there is less risk to be in-

[17] To compute the overall welfare cost of market incompleteness, we start from the stationary equilibrium of the $\lambda = 1.5$ incomplete market economy. We then compute the welfare gains of making the capital rental market perfect and also eliminating consumption inequality once and for all (ex ante insurance), correctly accounting for the transition phase. We repeat this calculation for various degrees of shock persistence, holding constant the unconditional distribution of the shocks.

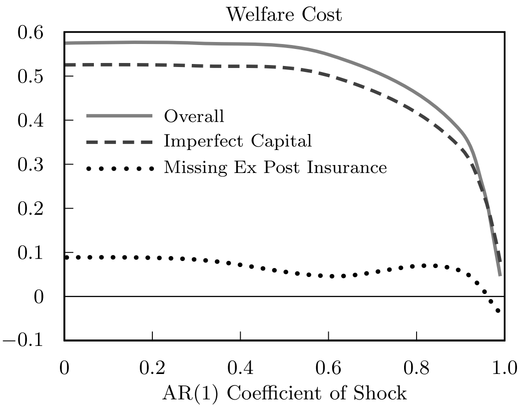

Fig. 4. The solid line is the welfare difference between the $\lambda = 1.5$ incomplete market economy and the complete market with ex post insurance and perfect capital markets (overall cost of market incompleteness). The dashed line is the cost of imperfect capital rental markets from the left panel of Fig. 1. The dotted line is the welfare difference between the $\lambda = \infty$ incomplete market economy and the complete market with ex post insurance (cost of missing ex post insurance).

sured in the short term—although farther out into the future the risk asymptotically attains its magnitude in the unconditional distribution—and ex post insurance becomes less valuable.[18]

The cost of imperfect capital rental markets remains the same as in Section 2.2.2, and is duly reproduced in Fig. 4 with a dashed line. With the welfare cost of missing insurance being much smaller than in Section 2.2.1, the overall welfare cost of market incompleteness (solid line) is almost equal to the cost of imperfect capital rental markets. As such, the overall welfare cost is decreasing in shock persistence, even for a very large λ.

3. Concluding remarks

Bewley models are a workhorse of modern macroeconomics. They have proven useful for understanding income and wealth inequality in the data, and for assessing the welfare consequences of market incompleteness. Numerous papers have relied on these models to evaluate the macroeconomic and distributional effects of various economic policies, such as income taxation, pensions, and health insurance.

In this paper, we raise two issues with the conventional Bewley models. One is the asymmetric assumptions of market incompleteness on the consumer side and the production side. We explore the idea that the same enforcement or information problems that preclude consumption insurance may as well disrupt the functioning of capital rental markets. The other pertains to the way individual income processes are modeled. Instead of specifying the income process directly, we go one level deeper and consider how entrepreneurial productivity or business opportunities translate into income realizations.

[18] Note that our welfare measure attaches equal weights to all individuals. With initial inequality, the Pareto weights in the planner problem for the complete market economy are not equal across individuals. Obviously, were it not for this discrepancy between these two weighting schemes, the welfare cost of missing ex post insurance would always be non-negative.

Our paper demonstrates that some standard model implications might be overturned once we enrich Bewley models to address these additional issues. We envision many other useful ways of laying the foundations of Bewley models, and leave them for future research.

Appendix A. Welfare cost calculations

We measure the welfare cost of market incompleteness by comparing the welfare of the stationary equilibrium of an incomplete market economy with a given λ to the welfare of three alternative market arrangements: (i) An economy that has perfect capital rental markets ($\lambda = \infty$) but no consumption insurance; (ii) a complete market economy where existing inequality is eliminated (ex ante insurance, Section 2.2.1); (iii) a complete market economy where the initial inequality—based on wealth and entrepreneurial productivity—is preserved (ex post insurance, Section 2.2.4). The last two economies have perfect capital rental markets.

Our welfare calculations correctly account for the transitions, which start from the stationary equilibrium of the incomplete market economy with a given λ. Our metric of welfare is the perfectly-egalitarian measure of economy-wide wellbeing.

The welfare $\mathcal{W}_\infty(\lambda)$ of the stationary equilibrium of an incomplete market economy with a given λ is computed by integrating the individual value functions—defined as Eq. (1) in Section 1—with respect to the invariant distribution of wealth and entrepreneurial productivity:

$$\mathcal{W}_\infty(\lambda) = \int v(a, e; \lambda) G_\infty(da, de; \lambda).$$

Similarly, the welfare of this incomplete market economy with λ that starts from a given joint distribution of wealth and entrepreneurial ability $G_0(a, e)$, is given by:

$$\mathcal{W}_0(G_0; \lambda) = \int v_0(a, e; \lambda) G_0(da, de).$$

The integration is with respect to the initial distribution of wealth and entrepreneurial productivity, and $v_0(a, e)$ is the individual value function that takes into account the aggregate transition from G_0 to G_∞. See Buera and Shin [5] for a formal definition.

The welfare $\mathcal{V}_0^{CM}(K_0)$ in a complete market economy with no inequality (ex ante insurance) that starts with aggregate capital K_0 solves the following Bellman equation:

$$\mathcal{V}_t^{CM}(K_t) = \max_{C_t, K_{t+1} \geqslant 0} u(C_t) + \beta \mathcal{V}_{t+1}^{CM}(K_{t+1})$$

$$\text{s.t. } C_t + K_{t+1} \leqslant A K_t^{\alpha(1-\nu)} + (1-\delta) K_t,$$

where A is the TFP in an economy with perfect capital rental markets,

$$A = \max_{\underline{e}} \left(\int_{e \geqslant \underline{e}} e^{\frac{1}{\nu}} \mu(de) \right)^{\nu} \mu(e)^{\alpha(1-\nu)},$$

and $\mu(e)$ is the c.d.f. of the unconditional entrepreneurial productivity distribution.

The welfare $\mathcal{W}_0^{CM}$ of a complete market economy where the initial inequality G_0 is respected (ex post insurance) is:

$$\mathcal{W}_0^{CM}(G_0) = \mathcal{V}_0^{CM}(K_0) \int \omega(a, e)^{1-\gamma} G_0(da, de),$$

where $\omega(a,e)$ is the consumption share of an individual with initial wealth a and entrepreneurial productivity e, given by:

$$\omega(a,e)=\frac{\hat{\omega}(a,e)}{\int\hat{\omega}(a,e)G(da,de)},$$
$$\hat{\omega}(a,e)=(1+r_0)a+\mathbb{E}\left[\sum_{t=0}^{\infty}\frac{\max\{\pi^u(\tilde{e};w_t,r_t),w_t\}}{(1+r_t)^t}\middle|e\right],$$

with $\pi^u(e;w,r)$ denoting the profits of an entrepreneur with productivity e unconstrained by collateral constraints.

The welfare cost of imperfect capital rental markets we report is in units of permanent consumption compensation necessary to make an individual indifferent between the status quo—the stationary equilibrium of the economy with λ—and an economy with better capital markets indexed by λ' (i.e., $\lambda'>\lambda$), accounting for the transition phase:

$$\chi(\lambda,\lambda')=1-\left(\frac{\mathcal{W}_\infty(\lambda)}{\mathcal{W}_0(G_\infty(a,e;\lambda);\lambda')}\right)^{\frac{1}{(1-\gamma)}}.$$

The welfare cost relative to a particular complete market economy is defined in a similar fashion.

References

[1] S.R. Aiyagari, Uninsured idiosyncratic risk and aggregate saving, Quart. J. Econ. 109 (3) (1994) 659–684.
[2] G.-M. Angeletos, Uninsured idiosyncratic investment risk and aggregate saving, Rev. Econ. Dynam. 10 (1) (2007) 1–30.
[3] T. Beck, A. Demirgüç-Kunt, R. Levine, A new database on the structure and development of the financial sector, World Bank Econ. Rev. 14 (3) (2000) 597–605.
[4] T.F. Bewley, Stationary monetary equilibrium with a continuum of independently fluctuating consumers, in: W. Hildenbrand, A. Mas-Colell (Eds.), Contributions to Mathematical Economics in Honor of Gérard Debreu, North-Holland, Amsterdam, 1986, pp. 79–102.
[5] F.J. Buera, Y. Shin, Financial frictions and the persistence of history: A quantitative exploration, Working Paper 16400, National Bureau of Economic Research, 2010.
[6] M. Cagetti, M. De Nardi, Entrepreneurship, frictions, and wealth, J. Polit. Economy 114 (5) (2006) 835–870.
[7] A. Castañeda, J. Díaz-Giménez, J.-V. Ríos-Rull, Accounting for the U.S. earnings and wealth inequality, J. Polit. Economy 111 (4) (2003) 818–856.
[8] F. Covas, Uninsured idiosyncratic production risk with borrowing constraints, J. Econ. Dynam. Control 30 (11) (2006) 2167–2190.
[9] A. Deaton, Saving and liquidity constraints, Econometrica 59 (5) (1991) 1221–1248.
[10] D.S. Evans, B. Jovanovic, An estimated model of entrepreneurial choice under liquidity constraints, J. Polit. Economy 97 (4) (1989) 808–827.
[11] F. Guvenen, Learning your earnings: Are labor income shocks really very persistent?, Amer. Econ. Rev. 97 (2007) 687–712.
[12] F. Guvenen, An empirical investigation of labor income processes, Rev. Econ. Dynam. 12 (1) (2009) 58–79.
[13] J. Heathcote, K. Storesletten, G.L. Violante, Insurance and opportunities: A welfare analysis of labor market risk, J. Monet. Econ. 55 (3) (2008) 501–525.
[14] J. Heathcote, K. Storesletten, G.L. Violante, Quantitative macroeconomics with heterogeneous households, Ann. Rev. Econ. 1 (1) (2009) 319–354.
[15] B. Holmström, J. Tirole, Private and public supply of liquidity, J. Polit. Economy 106 (1) (1998) 1–40.
[16] M. Huggett, The risk-free rate in heterogeneous-agent incomplete-insurance economies, J. Econ. Dynam. Control 17 (5–6) (1993) 953–969.
[17] M. Huggett, S. Ospina, Aggregate precautionary savings: When is the third derivative irrelevant?, J. Monet. Econ. 48 (2) (2001) 373–396.

[18] D.K. Levine, W.R. Zame, Does market incompleteness matter?, Econometrica 70 (5) (2002) 1805–1839.
[19] L. Ljungqvist, T.J. Sargent, Recursive Macroeconomic Theory, second edition, MIT Press, Cambridge, MA, 2004.
[20] R.E. Lucas Jr., On the size distribution of business firms, Bell J. Econ. 9 (2) (1978) 508–523.
[21] V. Quadrini, Entrepreneurship, saving, and social mobility, Rev. Econ. Dynam. 3 (1) (2000) 1–40.
[22] G. Tauchen, Finite state Markov chain approximations to univariate and vector autoregressions, Econ. Letters 20 (2) (1986) 177–181.