Throwing Good Money After Bad*

Matthew Ryan[†] Rhema Vaithianathan[‡] Luca Rigotti[§]

January 2014

Abstract

An "investment bubble" is a period of "excessive, and *predictably* unprofitable, investment" (DeMarzo, Kaniel and Kremer, 2007, p.737). Such bubbles most often accompany the arrival of some new technology, such as the tech stock boom and bust of the late 1990's and early 2000's. We provide a rational explanation for investment bubbles based on the dynamics of learning in highly uncertain environments. Objective information about the earnings potential of a new technology gives rise to a set of priors, or a *belief function*. A generalised form of Bayes' Rule is used to update this set of priors using earnings data from the new economy. In each period, agents – who are heterogeneous in their tolerance for ambiguity – make optimal occupational choices, with wages in the new economy set to clear the labour market. A preponderance of bad news about the new technology may nevertheless give rise to increasing firm formation around this technology, at least initially. To a frequentist outside observer, the pattern of adoption appears as an investment bubble.

JEL Codes: C11, D51, D81, J24, O31

Keywords: Ambiguity, Belief Function, Investment Bubble, Inference.

---

*Previous versions of this paper have circulated under the titles "Ambiguity Aversion, Entrepreneurship and Innovation" and "Ambiguity Aversion, Innovation and Investment Bubbles".

[†]Department of Economics, University of Auckland, New Zealand. Email: `m.ryan@auckland.ac.nz`

[‡]Department of Economics, AUT, New Zealand and Singapore Management University, Singapore. Email: `rhema.vaithianathan@aut.ac.nz`

[§]Department of Economics, University of Pittsburgh, USA. Email: `luca@pitt.edu`

1 Introduction

DeMarzo, Kaniel and Kremer (2007) describe a period of "excessive, and *predictably* unprofitable, investment" (p.737; emphasis in original) as an "investment bubble". They observe that such bubbles are often associated with new technologies, and cite the tech stock boom of the late 1990's as an example. They quote as follows from a 2003 report by the Joint Economic Committee of the United States Congress:

> "[B]usiness investment and aggregate after-tax corporate profits diverged from 1997 to 2000. Entrepreneurs and firm managers succumbed to the same 'irrationality' regarding their decisions to invest in capital assets that stock market participants were suffering regarding their decisions to purchase equities."(p.736)

In episodes such as this, the market appears to throw good money after bad. However, it is less clear that genuine "irrationality" is involved. In this paper we show that investment bubbles may arise quite naturally in the presence of ambiguity about the prospects of a new technology.

In our model, market participants must update ambiguous prior beliefs on the basis of earnings data from a new technology sector. We show that the evolving equilibrium may follow a boom-bust path, even if the market data convey "bad news" on average. If the market had held *precise* prior beliefs, a monotonic decline of the new sector would have been observed, but ambiguity may support an initial period of significant growth.

Our model is based on that in Rigotti, Ryan and Vaithianathan (2011; henceforth, RRV). Unlike RRV, which studied the diffusion profile of successful innovations, we focus here on the adoption of *unsuccessful* technologies. We also depart from RRV by incorporating an explicit learning process into the model. Market participants hold ambiguous prior beliefs, so learning involves a non-standard inference problem. We apply the theory of *belief functions* (Dempster, 1967; Shafer, 1976) to solve this inference problem.

In the model, an investment bubble arises within a *stable* economic environment with *public* information. There are no exogenous shocks and no informational asymmetries. Two effects drive the learning dynamics. First, the content of the earnings data (i.e., "good news" versus "bad news") causes agents to adjust their profit expectations in the obvious direction. This is what we call the *news effect.* Second, *any* news, good or bad, reduces the level of uncertainty around these profit expectations. This *ambiguity reduction* effect does not arise in conventional Bayesian inference from a precise prior. The ambiguity-reduction effect ameliorates the impact of bad news – and amplifies the impact of good news – on the lower bound on profit expectations. Data which contain predominantly bad news may therefore improve the lower bound. This is what drives the boom phase of the investment bubble in the model.

Note that ours is not a story of excessive exuberance, irrational or otherwise. Entrepreneurial innovators, who focus on the *upper* bound on profit expectations, are increasingly discouraged by the poor earnings data. It is their conservative suppliers of capital – human and financial – who are emboldened to increase investment by the improvement in worst-case scenario outcomes.

The paper proceeds as follows. Section 2 introduces our notion of ambiguity. We also describe how market participants make statistical inferences when they have ambiguous priors. In Section 3 we develop a simple economic model with one good and two production technologies – an old technology which is well understood and a new technology of uncertain productivity. Agents, who differ in their tolerance for ambiguity, make rational occupational choices, with wages and the number (mass) of new technology firms determined endogenously in competitive equilibrium. Section 4 studies learning dynamics for this economy. Each period, all agents observe the earnings of the new technology firms and use this public information to update beliefs about its expected profitability. Using a simulated numerical example, we demonstrate the possibility of positive diffusion despite the predominance of bad news. The size of the new technology sector initially rises before the shake-out begins, even though the new technology is inferior to the old, as a frequentist analysis of the data would lead one to conclude. We offer some concluding remarks in Section 5.

2 Ambiguity and inference

Consider a finite state (or parameter) space $\Theta = \{\theta_1, ..., \theta_n\}$. Dempster (1967) provides a convenient structure for characterising ambiguous beliefs over Θ. Dempster supposes that these beliefs are induced by a *source* (S, Σ, μ, Γ), where (S, Σ, μ) is a measure space and $\Gamma : S \to 2^{\Theta} \diagdown \{\varnothing\}$ is a measurable mapping from S to the non-empty subsets of Θ. This source provides information about Θ as follows: if $s \in S$ is realised, then the true θ value must lie in the set $\Gamma(s) \subseteq \Theta$. Knowledge of the measure μ on (S, Σ) therefore provides probabilistic information about θ via the *information mapping* Γ. For example, if $\Gamma(s)$ is a singleton for every $s \in S$, then the source induces a probability on Θ. Otherwise, the source may provide only partially specified probabilistic information about θ.

To quantify this information, we associated to each source (S, Σ, μ, Γ) a *belief function* $v : 2^{\Theta} \to [0, 1]$, defined as follows:

$$v(E) = \mu(\{s \in S \mid \Gamma(s) \subseteq E\}).$$

The quantity $v(E)$ is a lower bound on the probability of E implied by the source. Letting $\Delta(\Theta)$ denote the set of all probability measures on Θ,

$$\Pi = \{\pi \in \Delta(\Theta) \mid \pi(E) \geq v(E) \quad \text{for all } E \subseteq \Theta\}$$

is the collection of probability measures consistent with the source. Importantly, the set Π is non-empty (Dempster, 1967). It is clearly also closed and convex. Economists refer to Π as the *core* of v.[1]

Shafer (1976) discusses alternative characterisations of belief functions, and argues for their usefulness as measures of degrees of belief.

Many situations in which information is ambiguous, or imprecisely specified, may be represented using Dempster's construction. A notable example is furnished by Ellsberg's (1961) three-colour experiment.

The Ellsberg Experiment. Consider an urn containing 90 balls. The following information is given: 30 of the balls are red (r), while each of the other 60 balls is either black (b) or green (g). A ball has been drawn from the urn at random. Let $\theta \in \Theta = \{r, b, g\}$ denote the unknown colour of the ball. The information given does not allow us to assign precise probabilities to each element of Θ, though we can certainly assign probability $\frac{1}{3}$ to r. The information can, however, be described using a source defined as follows: $S = \{s', s''\}$ with $\Sigma = 2^S$, $\Gamma(s') = \{r\}$, $\Gamma(s'') = \{b, g\}$, $\mu(s') = \frac{1}{3}$ and $\mu(s'') = \frac{2}{3}$. In other words, with probabilty $\frac{1}{3}$ the ball is red, and with probability $\frac{2}{3}$ it is either black or green. From this source we obtain the following belief function:

$$v(E) = \begin{cases} 1 & \text{if } E = \Theta \\ \frac{2}{3} & \text{if } E = \{b, g\} \\ \frac{1}{3} & \text{if } r \in E \neq \Theta \\ 0 & \text{otherwise} \end{cases}$$

and associated set of probabilities:

$$\Pi = \left\{ \pi \in \Delta(\Theta) \,\middle|\, \pi(r) = \frac{1}{3} \right\} \tag{1}$$

The set Π contains exactly those probabilities on Θ that are consistent with the information given.

[1] There is some potential for confusion, however, as Shafer (1976, p.40) uses the term "core" to refer to a different property of belief functions. We will conform to the economists' usage here. Note that v may be recovered as the lower envelope of Π:

$$v(E) = \min_{\pi \in \Pi} \pi(E).$$

Importantly, this example illustrates that ambiguous beliefs may arise from *objective* information. Anyone provided with the information above would arrive at the same set (1) of compatible probabilities.

In the following section, we consider an economy in which the earnings of firms using a new technology are described by a discrete random variable whose associated probability function depends on the unknown value of the parameter $\theta \in \Theta = \{\theta_1, ..., \theta_n\}$. Prior information about θ can be summarised in the form of a belief function, v.

If v is not a probability, how should market participants update their beliefs on the basis of earnings data? What is the appropriate rule of inference?

The statistics literature does not offer a consensus view on this question. Several proposals have been studied. We adopt the method originally proposed by Shafer (1976, Chapter 11), which generalises Bayesian inference. In particular, it coincides with standard Bayesian inference when the belief function is a probability (Π is a singleton). The details of Shafer's Method (SM) are given in Appendix A.[2] We here describe its salient features.

Recall that the belief function v assigns a lower bound to the probability of each $E \subseteq \Theta$. In other words, prior beliefs determine a *probability interval*

$$[v(E),\ 1 - v(\Theta \setminus E)] \tag{2}$$

for each event $E \subseteq \Theta$. The probability interval for E may be usefully characterised by its *mid-point*

$$\frac{1}{2} v(E) \ + \ \frac{1}{2} [1 - v(\Theta \setminus E)] \tag{3}$$

and its *width*

$$1 - v(\Theta \setminus E) - v(E) \tag{4}$$

The process of inference updates the mid-point and the width. We shall refer to the data's impact on the mid-point as the *news effect* and its impact on the width as the *ambiguity-reduction effect*.

The interaction of these two effects will be important for our analysis. The data are said to be "good news" for E if the mid-point (3) increases after updating, and "bad news" if it decreases. It is also natural to expect that new data will reduce the width (4) by reducing ambiguity. Such is the case under SM updating for our model. If so, then the ambiguity-reduction effect ameliorates the effect of good news on $v(E)$ and amplifies its effect on $1 - v(\Theta \setminus E)$, and conversely for bad news. See Figure 1.

Suppose we receive a series of data which convey predominantly bad news for E, in the sense that the mid-point (3) is lower after updating on the basis of this data. The upper end-point of the interval (2) will also certainly have fallen. However, the lower end-point may have risen if the ambiguity-reduction effect is sufficiently strong.

[2] We also discuss two alternative methods of inference in Appendix A.

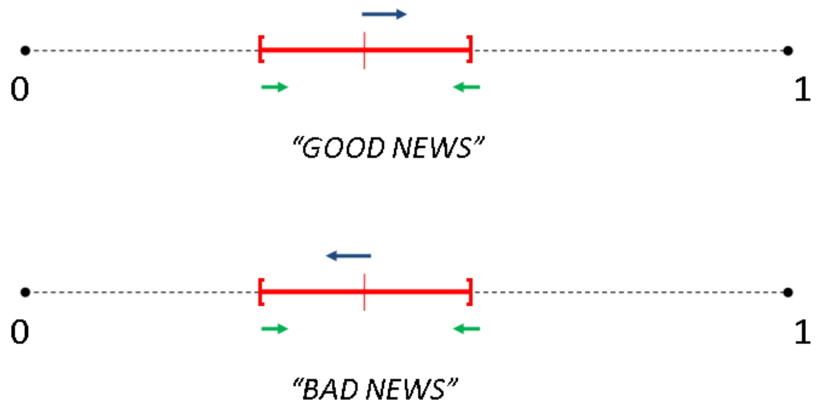

Figure 1: News and ambiguity-reduction effects on probability intervals

It is this latter possibility which creates the potential for the market to throw good money after bad. Suppose $\theta \in E$ implies high earning potential for the new technology and $\theta \in \Theta \diagdown E$ implies low earning potential. Relatively pessimistic agents – those who pay relatively close attention to the lower end-point of the probability interval (2) – may therefore become increasingly sanguine about the new technology, at least initially, even if most of the early news is bad. Depending on the aggregate degree of pessimism in the population, one might observe an "investment bubble" of the sort described by DeMarzo, Kaniel and Kremer (2007).

The remainder of the paper illustrates this possibility in a simple model. Note that it is a possibility, not a certainty. There must be enough agents who are sufficiently pessimistic to be positively affected when news is bad on average, yet sufficiently bold to be willing to shift to the "new economy" from their safe jobs in the "old economy".

3 A model of firm formation

Consider an economy with a single consumption good and two technologies for producing it: an established technology (α) and a new innovation (β). Each technology requires the input of two (full-time) agents. Both technologies are freely available – one may think of them as different techniques for deploying the human capital of the firm, rather than technologies embodied in capital goods. A firm is formed when two agents join forces to produce the consumption good using one of the available technologies. In particular, each agent must commit all of her human capital to a single firm. We will speak of "α firms" and "β firms", with the obvious meanings.

Since there is only one consumption good, the partners in any given firm consume what they produce according to an agreed sharing rule – there is no trade between firms. Our main interest is in the process of firm formation and the level of adoption of the new technology β.

We make the following assumption about technologies α and β.

Assumption 1 *Each technology produces a stochastic output: either M or $m \in (0, M)$. Technology α produces M with probability*

$$\frac{2K - m}{M - m}$$

(where $K > 0$), while technology β produces M with probability $\theta \in \Theta = \{\theta_1, \theta_2\}$, where

$$0 \ < \ \theta_1 \ < \ \frac{2K - m}{M - m} \ < \ \theta_2 \ < \ 1.$$

Note that technology α has an expected output of $2K$ units, or K units per capita. Technology β either first-order stochastically dominates technology α (if $\theta = \theta_2$), or is first-order stochastically dominated by it (if $\theta = \theta_1$). In particular:

$$\theta_2 M + (1 - \theta_2)\, m \ > \ 2K \ > \ \theta_1 M + (1 - \theta_1)\, m \tag{5}$$

We say that the new technology is *superior* to the old if $\theta = \theta_2$ and *inferior* if $\theta = \theta_1$.[3]

There is uncertainty about which θ value governs the new technology's production, and no agent possesses any private information about θ. All agents share common beliefs about θ based on public information.

Assumption 2 *There is ambiguous public information about the true value of θ, described by the belief function v on Θ with associated (non-singleton) core Π.*

Firms use Pareto efficient sharing rules to share the stochastic output between the partners. *All agents are risk neutral in the model* – see (6) below – *so any sharing rule is Pareto efficient for a firm using technology α.* It is therefore natural to make the following:

[3] These stochastic dominance relationships are significant. Two well-known explanations of investment bubbles rely on risk aversion to conclude that the investment is "irrational" (DeMarzo, Kaniel and Kremer, 2007; Pástor and Veronesi, 2006). In our model, investing in an inferior technology is irrational irrespective of the investor's risk attitude.

Assumption 3 *Partners in firms using technology α agree to share output equally.*

For firms using technology β, it is not straightforward to determine the Pareto efficient sharing rules.

Suppose a partner in such a firm receives share $s_M \in [0, M]$ when realised output is M and $s_m \in [0, m]$ when realised output is m. We assume that she uses the following *Arrow-Hurwicz criterion* (Arrow and Hurwicz, 1972; Luce and Raiffa, 1957, p.282)[4] to evaluate the contract $s = (s_M, s_m) \in [0, M] \times [0, m]$:

$$U(s;\lambda) = \lambda \left[\max_{\pi\in\Pi} \sum_{\theta\in\Theta} \pi(\theta) f(\theta;s) \right] + (1-\lambda) \left[\min_{\pi\in\Pi} \sum_{\theta\in\Theta} \pi(\theta) f(\theta;s) \right] \qquad (6)$$

where

$$f(\theta;s) = \theta s_M + (1-\theta) s_m$$

is the expected payment in state θ and $\lambda \in [0,1]$ the agent's *ambiguity tolerance* parameter. The value of λ is the only dimension along which agents will be differentiated in the model.

We shall impose the restriction $\lambda \leq \frac{1}{2}$ for every agent – see Assumption 4 below – in order to make use of the following result (which may be of independent interest):[5]

Theorem 3.1 *Consider two agents with respective ambiguity tolerance parameters λ_1 and λ_2 satisfying $\lambda_1 < \lambda_2 \leq \frac{1}{2}$. Suppose they are partners in a β firm. Let $s_M^k \in [0, M]$ and $s_m^k \in [0, m]$ be the shares of M and m (respectively) promised to the partner with parameter λ_k ($k \in \{1,2\}$). (Hence $s_M^1 + s_M^2 = M$ and $s_m^1 + s_m^2 = m$.) This sharing rule is Pareto efficient iff $s_m^1 = \min\{s_M^1, m\}$.*

In a Pareto efficient sharing rule, the less ambiguity-tolerant partner is offered a fixed "wage" ($w = s_M^1$), which is paid to the extent resources allow (i.e., $s_m^1 = \min\{w, m\}$). The partner with the higher tolerance for ambiguity becomes the residual output claimant. Such firms have a familiar "owner-worker" structure. We assume that all β firms pay the same wage w.[6] This w will be chosen to equilibrate the labour market for the β-technology sector.

Assumption 4 *There is a continuum of agents with unit mass. Ambiguity tolerance is distributed in the population according to the differentiable distribution function H. We assume that H is satisfies $H(0) = 0$ and $H\left(\frac{1}{2}\right) = 1$, and $H'(x) > 0$ whenever $H(x) \in (0,1)$.*

[4] Jaffray (1989, 1991, 1994), Hendon *et al.* (1994) and Jaffray and Wakker (1994) provide axiomatic foundations for this model of decision-making.

[5] All proofs are in Appendix B.

[6] If wages differed across such firms, the owner of a high-wage β firm and the worker from a low-wage β firm could both be better off by forming a new β firm paying an intermediate wage.

Given $w \leq M$, each agent chooses an optimal occupation from the set $\mathcal{O} = \{A, B_L, B_O\}$, where A denotes an occupation in an α firm, B_L denotes wage-labour in a β firm and B_O denotes owning a β firm. Occupation A yields payoff K (Assumption 3); occupation B_L yields payoff (6) with $s_M = w$ and $s_m = \min\{w, m\}$; and occupation B_O yields payoff (6) with $s_M = M - w$ and $s_m = \max\{m - w, 0\}$.

Let us define $BR(w; \lambda) \subseteq \mathcal{O}$ to be the set of utility-maximising occupations for type λ given w.

Lemma 3.1 *If $w < \frac{1}{2}(M + m)$, there exist unique values*

$$\underline{\lambda}(w) \in \left[0, \frac{1}{2}\right]$$

and

$$\overline{\lambda}(w) \in \left[\underline{\lambda}(w), \frac{1}{2}\right]$$

such that

$$BR(w;\lambda) = \begin{cases} \{B_O\} & \textit{if } \lambda > \overline{\lambda}(w) \\ \{B_L\} & \textit{if } \lambda \in \left(\underline{\lambda}(w),\ \overline{\lambda}(w)\right) \textit{ and } (w > m \textit{ or } w \neq K) \\ \{A, B_L\} & \textit{if } \lambda \in \left(\underline{\lambda}(w),\ \overline{\lambda}(w)\right) \textit{ and } w = K \leq m \\ \{A\} & \textit{if } \lambda < \underline{\lambda}(w) \end{cases}$$

If $w \geq \frac{1}{2}(M + m)$ the worker in a β firm receives a strictly higher share of any realised output than the owner. No-one would choose to own a β firm at such a wage, so such firms could not exist in equilibrium (Definition 1). When $w < \frac{1}{2}(M + m)$ Lemma 3.1 implies a natural ordering of occupations by ambiguity: the most ambiguity-tolerant agents choose to own β firms, the least ambiguity tolerant are partners in α firms, and a middle group supply labour to β firms.[7]

Definition 1 *We say that w is an* equilibrium *wage if there exist $\lambda^*, \lambda^{**} \in \left[0, \frac{1}{2}\right]$ with $\lambda^* \leq \lambda^{**}$ such that*

(i) $A \in BR(w, \lambda)$ *for all* $\lambda < \lambda^*$

(ii) $B_L \in BR(w, \lambda)$ *for all* $\lambda \in (\lambda^*, \lambda^{**})$

(iii) $B_O \in BR(w, \lambda)$ *for all* $\lambda > \lambda^{**}$

(iv) $H(\lambda^{**}) - H(\lambda^*) \ = \ 1 - H(\lambda^{**})$

[7]Note, however, that Lemma 3.1 does not guarantee that each of these sets is non-empty.

At an equilibrium wage it is possible to allocate agents to utility-maximising occupations in such a way that the β-sector labour market clears – condition (iv).[8] The quantity $1 - H(\lambda^{**})$ is the mass of β firms which form in the equilibrium.

Theorem 3.2 *An equilibrium exists and the mass of β firms is the same at any equilibrium.*

4 Learning

The model of the preceding section provides the simplest possible environment for illustrating an investment bubble.

If β firms are formed in equilibrium, then realised output from these firms generates (public) information about θ. Agents use this information to update their common belief function (according to the SM rule). Updating disturbs the existing equilibrium – the utilities associated with occupations in the β-technology sector change – so a new equilibrium will be established. The learning process therefore determines a stochastic path for the mass of β firms.[9] We will show that the *average* path may rise and then fall when β is an *inferior* technology ($\theta = \theta_1$). Of course, data from the β-technology sector will convey bad news (on average) when β is inferior to α, so this path resembles the boom and bust of an investment bubble.

Let us therefore introduce a time index $t \in \{1, 2, ...\}$. Technology β first becomes available at the start of period $t = 1$. In period t, agents base their decisions on a common belief function $v = v_t$ with associated (non-singleton) core Π_t. The belief function v_1 is exogenously given, and describes the public information about β available prior to its adoption. All β firms formed in period t receive the same output realisation $y_t \in \{M, m\}$.[10] The belief function v_{t+1} is formed by updating v_t based on the observation y_t using the SM rule.

[8] If $w \geq \frac{1}{2}(M+m)$ we have $BR(w, \lambda) \subseteq \{A, B_L\}$ for every λ (recall the discussion following Lemma 3.1) so equilibrium requires $A \in BR(w, \lambda)$ for all λ (i.e., $\lambda^* = \lambda^{**} = \frac{1}{2}$).

[9] We assume that technology choices are freely reversible and that the consumption good perishes after one period, so the optimal choice of occupation depends only on current period returns. Hence, the learning process determines a sequence of one-shot equilibria of the sort defined in the previous section. Of course, if no β firms form in period t, then agents learn nothing further about the β technology and the equilibrium does not change thereafter. This may create incentives for agents to experiment with the β technology even if it reduces current utility, in order to produce information that will alter the future values of equilibrium variables. We shall sidestep this potential source of temporal dependence in decision-making by supposing that a measure zero set of "non-rational" agents always form β firms, or that a public research institute produces new data on technology β in each period (as in Jensen, 1982).

[10] If each firm received an independent draw from the Bernoulli random variable then a continuum of observations would be generated each period and learning would cease after one round of updating.

The system therefore evolves according to the following stochastic process. Nature draws a sequence $\{y_1, y_2, ...\}$ of independent observations on the Bernoulli random variable that takes value M with probability θ and m with probability $1-\theta$. From this sequence (and the exogenously given v_1) we determine a sequence $\{v_1, v_2, ...\}$ of belief functions via SM inference. In period t, the equilibrium wage w_t and associated mass of β firms, denoted by δ_t, is determined using v_t. Our interest focusses on the shape of the average $\{\delta_1, \delta_2, ...\}$ path when $\theta = \theta_1$ (i.e., when β is an inferior technology).

Figure 2 illustrates two such paths, for two particular cases of our model. The horizontal axis measures t and the vertical measures δ_t. For each path, we have set $M = 70$, $m = 10$, $\theta_1 = 0.45$, $\theta_2 = 0.5$, $K = 19.15$ and $v_1(\{\theta_1\}) = v_1(\{\theta_2\}) = 0.1$.[11] The two paths are generated using two different H distributions, one left-skewed and one right-skewed.[12] A left-skewed distribution represents a more ambiguity-tolerant population than a right-skewed distribution. For each case, we plot the average $\{\delta_1, \delta_2, ..., \delta_{50}\}$ path over 1000 simulations.[13] Both cases exhibit an initial upswing in this averaged path. The upswing is most pronounced for the left-skewed H distribution.

It is important to recall that Figure 2 reports *averaged* diffusion paths over 1000 trials, so these shapes are not the artefacts of a particular sample path for y_t. On average, the data favour θ_1. If v_1 was a probability, so all agents are conventional Bayesians, the average posterior probability on θ_1 would increase monotonically, giving a monotonically declining path for (average) δ_t. The ambiguity-reduction effect is clearly at work during the upswing in Figure 2.

What is going on in these examples? Let us define

$$\underline{p}_t = \min_{\pi \in \Pi_t} \sum_{\theta \in \Theta} \pi(\theta)\,\theta = v_t(\{\theta_2\})\,\theta_2 + [1 - v_t(\{\theta_2\})]\,\theta_1$$

$$\overline{p}_t = \max_{\pi \in \Pi_t} \sum_{\theta \in \Theta} \pi(\theta)\,\theta = v_t(\{\theta_1\})\,\theta_1 + [1 - v_t(\{\theta_1\})]\,\theta_2$$

Then $\left[\underline{p}_t\ ,\ \overline{p}_t\right]$ is the probability interval for $y_t = M$ based on the information available at the start of period t. Initially, we have $\left[\underline{p}_1\ ,\ \overline{p}_1\right] = [0.455,\ 0.495]$ with mid-point

$$\frac{1}{2}\left(\underline{p}_1 + \overline{p}_1\right) = 0.475.$$

On average, the news is "bad" for this event, since $\theta_1 = 0.45$ is the true parameter value: the mid-point declines (on average). However, the lower bound on this interval rises initially

[11] Note that v_1 is a belief function (a source is easily constructed) and Assumption 1 is satisfied.

[12] For the right-skewed distribution, we used the Beta$(25, 15)$, and for the left-skewed distribution the Beta$(15, 25)$. We re-scaled each distribution so it is supported on $\left[0, \frac{1}{2}\right]$.

[13] We have truncated the simulation at 50 periods to focus on the paradoxical upswing, but the δ_t paths do eventually converge to zero for both distributions.

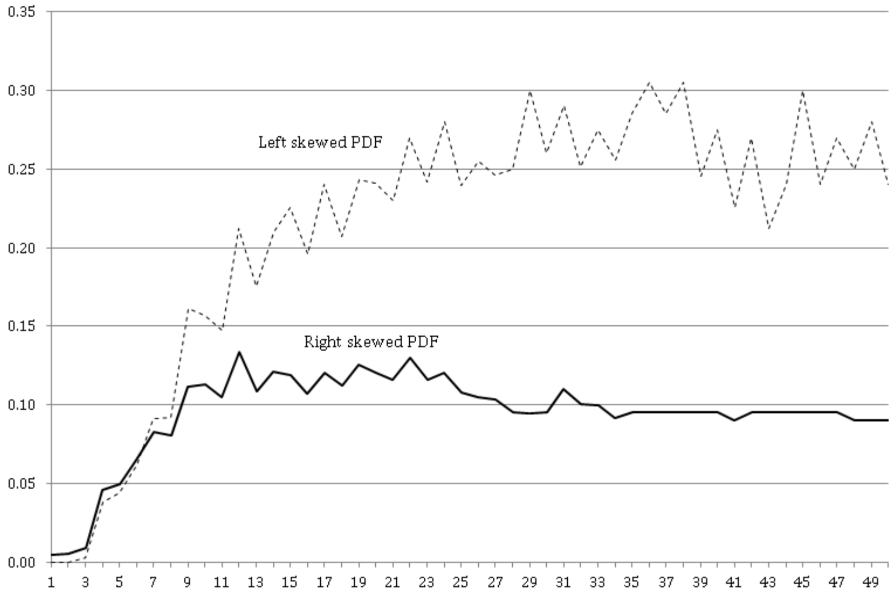

Figure 2: Diffusion profile for $\theta = \theta_1$ with right vs left skewed λ distribution

due to the ambiguity-reduction effect. This rising lower bound drives the upswing in the adoption of the new technology. Figure 3 illustrates. It traces the mid-point and lower bound for the average probability interval $\left[\underline{p}_t\ ,\ \overline{p}_t\right]$, using v_1, θ_1 and θ_2 as for Figure 2.

The driving force behind the upswing is not the increased exuberance of β firm owners. A maximally ambiguity-tolerant decision-maker, one with $\lambda = \frac{1}{2}$, evaluates β sector occupations using the *mid-point* of the probability interval $\left[\underline{p}_t\ ,\ \overline{p}_t\right]$: recall (6) and note that

$$\sum_{\theta\in\Theta} \pi(\theta) f(\theta; s) \;=\; s_m \;+\; (s_M - s_m) \sum_{\theta\in\Theta} \pi(\theta)\,\theta.$$

This mid-point is falling (on average), so types with λ near $\frac{1}{2}$ are increasingly pessimistic about the prospects of the β technology. On the other hand, the most ambiguity-*intolerant* decision-maker, one with $\lambda = 0$, evaluates β sector occupations using the *lower bound* of the probability interval $\left[\underline{p}_t\ ,\ \overline{p}_t\right]$, which is initially rising (on average). In general, the lower the value of λ (i.e., the less ambiguity tolerant the decision-maker), the more attention is paid to this lower bound. Recalling Lemma 3.1, it is clear that the main driver of the

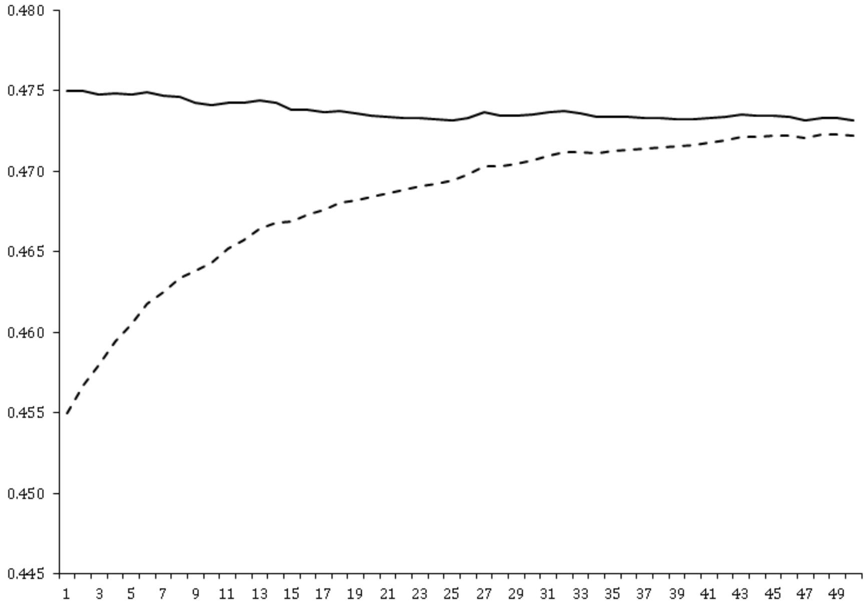

Figure 3: Evolution of lower bound $\underline{p}_t$ (dashed) and mid-point $\frac{1}{2}\left(\underline{p}_t + \overline{p}_t\right)$ (solid)

upswing is an increase in the supply of labour to β firms from agents with relatively low λ values. This pushes down equilibrium wages and encourages β firm formation, despite the increasing pessimism of many β firm owners.

5 Concluding remarks

In ambiguous environments, there is a sense in which "any news is good news" for pessimists and "any news is bad news" for optimists. All news reduces ambiguity, so optimistic expectations about best-case-scenarios are dented, while pessimists become less gloomy about worst-case-scenarios. The latter effect drives the upswing in investment during the initial phase of the investment bubble. Since equilibrium forces select (relative) optimists into entrepreneurship, it is the pessimism of their suppliers of labour which is instrumental in explaining the boom.

One can easily imagine similar dynamics motivated by the relative pessimism of financiers to new technology entrepreneurs. Our wage contracts are analogous to standard debt contracts, so we would expect the suppliers of finance to be more pessimistic – in equilibrium – than the entrepreneurial borrowers.[14] In particular, risks are technology-specific

[14]Several papers study the financing of "optimistic" entrepreneurs by "realistic" lenders, though the

rather than firm-specific, so lenders cannot diversify against the uncertain performance of the new technology.

Our bubble features the "predictably unprofitable" investment of the DeMarzo, Kaniel and Kremer (2007) definition. An outside observer applying a frequentist lens to the data would (on average) conclude that the new technology is less profitable than – indeed, first-order stochastically dominated by – the old technology, and they would do so well before the path of investment turns downward. Likewise, a Bayesian observer with a probabilistic prior, whatever it might be, would (on average) have a monotonically declining expectation of the returns to the new technology. These expectations are eventually confirmed, lending an appearance of irrationality to the original run-up in investment by the market.

Appendix A

Inference with Multiple Priors[15]

The SM method constructs a *likelihood-based belief function* to encapsulate the new information from the sample data, then combines this with the prior belief function using *Dempster's rule of combination.* We shall discuss the italicised terms in reverse order.

Dempster's rule is a natural method for combining information about a common state space Θ from two independent sources, say (S, Σ, μ, Γ) and $(S', \Sigma', \mu', \Gamma')$. One first constructs a combined source

$$(S \times S', \Sigma \times \Sigma', \mu^*, \Gamma^*),$$

where $\Gamma^*(s, s') = \Gamma(s) \cap \Gamma'(s')$ contains the Θ-states consistent with both $s \in S$ and $s' \in S'$, and μ^* is the product measure $\mu \times \mu'$ conditioned on

$$E^* \;=\; \{(s, s') \in S \times S' \mid \Gamma^*(s, s') \neq \emptyset\}.$$

In other words, E^* is the event that the two sources deliver non-contradictory information. Next, one constructs the belief function generated by $(S \times S', \Sigma \times \Sigma', \mu^*, \Gamma^*)$. If v is the belief function induced by (S, Σ, μ, Γ) and v' the belief function induced by $(S', \Sigma', \mu', \Gamma')$, then the belief function associated with $(S \times S', \Sigma \times \Sigma', \mu^*, \Gamma^*)$ is denoted $v \oplus v'$ and satisfies

$$\begin{aligned}(v \oplus v')(E) \quad &= \quad \frac{(\mu \times \mu')(\{(s, s') \in S \times S' \mid \Gamma^*(s) \subseteq E\})}{(\mu \times \mu')(E^*)} \\ &= \quad \mu^*(\{(s, s') \in S \times S' \mid \Gamma^*(s) \subseteq E\})\end{aligned}$$

definitions of optimism in these papers typically differ from ours: see, for example, De Meza and Southey (1996), Manove and Padilla (1999) and Dushnitsky (2010).

[15] The interested reader may consult Wasserman (1990) for more details.

for each $E \subseteq \Theta$. The operator $\oplus$ is called the *orthogonal sum.*

The *likelihood-based belief function* is the belief function induced by the source

$$\left([0,1]\,,\ \mathcal{B}\left([0,1]\right),\ \text{Leb},\ \Gamma^{l}\right),$$

where $\mathcal{B}\left([0,1]\right)$ are the Borel subsets of $[0,1]$, Leb is Lebesgue measure, and

$$\Gamma^{l}(s) \;=\; \left\{\theta\in\Theta \;\middle|\; \frac{l(\theta)}{\max_{\theta'\in\Theta} l(\theta')} \geq s\right\}.$$

Thus, $\Gamma^{l}(s)$ is a likelihood upper-contour set. The belief function induced by this source has the following specification:

$$v^{l}(E) \;=\; 1 \;-\; \frac{\max_{\theta\in\Theta\diagdown E} l(\theta)}{\max_{\theta'\in\Theta} l(\theta')} \tag{7}$$

To help interpret the likelihood-based belief function (7), observe that $v^{l}(E)$ is the contribution of E to achieving the maximum likelihood: $100\left[1-v^{l}(E)\right]\%$ of the maximum likelihood can be achieved within $\Theta\diagdown E$. Wasserman (1990) discusses axiomatic foundations for (7).

In our model, we use the SM rule to update beliefs about the "success" probability of a Bernoulli random variable, so let's consider that inference problem.

Let Y be a Bernoulli random variable with success probability $\theta\in\Theta=\{\theta_1,\theta_2\}$, where $\theta_1<\theta_2$. The prior is described by the belief function v_1. We observe a sequence $\{y_t\}_{t=1}^{m}$ of random draws from Y, where $y_t\in\{M,m\}$ with $y_t=M$ indicating a "success" in period t. After each draw we apply the SM procedure to update the belief function. Let v_t denote the belief function prevailing at the start of period t. To simplify notation, we write $v_{t,i}$ for $v_t\left(\{\theta_i\}\right)$.

By direct calculation, the SM rule gives the following update formulae:

$$v_{t+1}\left(\{\theta_1\}\mid y_t=M\right) \;=\; \frac{v_{t,1}\theta_1}{v_{t,1}\theta_1+\left(1-v_{t,1}\right)\theta_2} \;\leq\; v_{t,1} \tag{8}$$

$$v_{t+1}\left(\{\theta_2\}\mid y_t=M\right) \;=\; \frac{v_{t,2}\theta_2+\left(1-v_{t,1}-v_{t,2}\right)\left(\theta_2-\theta_1\right)}{v_{t,1}\theta_1+\left(1-v_{t,1}\right)\theta_2} \;\geq\; v_{t,2} \tag{9}$$

$$v_{t+1}\left(\{\theta_1\}\mid y_t=m\right) \;=\; \frac{v_{t,1}\left(1-\theta_1\right)+\left(1-v_{t,1}-v_{t,2}\right)\left(\theta_2-\theta_1\right)}{v_{t,2}\left(1-\theta_2\right)+\left(1-v_{t,2}\right)\left(1-\theta_1\right)} \;\geq\; v_{t,1} \tag{10}$$

$$v_{t+1}\left(\{\theta_2\}\mid y_t=m\right) \;=\; \frac{v_{t,2}\left(1-\theta_2\right)}{v_{t,2}\left(1-\theta_2\right)+\left(1-v_{t,2}\right)\left(1-\theta_1\right)} \;\leq\; v_{t,2}. \tag{11}$$

These coincide with the standard Bayesian updating formulae if

$$v_{t,1} \;+\; v_{t,2} \;=\; 1$$

(i.e., if there is no ambiguity at the start of period t).

Observe also that

$$v_{t+1}\left(\{\theta_1\} \mid y_t\right) + v_{t+1}\left(\{\theta_2\} \mid y_t\right) \;\; \geq \;\; v_{t,1} \; + \; v_{t,2}$$

for any $y_t \in \{M, m\}$. The probability interval for the event $\{\theta_2\}$ is $[v_{2,t},\; 1 - v_{1,t}]$ in period t, so updating reduces the width of this interval for any y_t. (The same is true of the interval for $\{\theta_1\}$.) This is the *ambiguity reduction* effect.

As noted in Section 2, the SM approach is not the only method of inference available when the prior is described through a belief function. Let us briefly mention two others.

Robust Bayesian Approach. Robust Bayesian (RB) inference applies Bayes' Rule to every prior in Π_t to construct a set of posteriors. (This procedure can be applied whether or not Π_t is the core of a belief function.) Importantly, the SM posterior probability interval for any event is always contained in the RB posterior interval (Wasserman, 1990, Theorem 7). Roughly speaking, SM inference resolves ambiguity more rapidly than RB inference. Since the ambiguity-reduction effect is critical to an investment bubble, SM inference is more conducive to a bubble than RB inference. In fact, we have not been able to obtain an investment bubble in any simulation in which agents perform inference according to the RB procedure.

Indeed, many scholars – including Shafer (1982, p.327) – have noted with concern the radical lack of ambiguity-resolution when applying the RB method. It is well-known, for example, that RB inference may exhibit *dilation* – the paradoxical possibility that the posterior interval for an event, conditional on *any* cell in a partition of the parameter space, strictly contains the prior interval (Seidenfeld and Wasserman, 1993).[16]

Shafer's Alternative Method. Shafer himself had misgivings about SM inference and subsequently developed some alternatives in Shafer (1982).[17] He there proposes

[16] A simple example is the following (Seidenfeld and Wasserman, 1993, p.1140). Two coins are to be flipped and the outcomes recorded. Each coin is known to be fair but there is complete uncertainty about the independence or otherwise of the outcomes. Thus, $\Theta = \{HH, HT, TH, TT\}$ and the prior set is

$$\Pi \;=\; \left\{ \pi \in \Delta(\Theta) \;\middle|\; \pi\left(\{HH, HT\}\right) = \pi\left(\{HH, TH\}\right) = \frac{1}{2} \text{ and } 0 \leq \pi\left(\{HH\}\right) \leq \frac{1}{2} \right\}$$

The prior interval for the event of tossing a Head on the first coin is $\left\{\frac{1}{2}\right\}$, but conditional on the outcome of the second coin toss – whatever it might be – the posterior interval is $[0, 1]$.

[17] His misgivings arose from the fact that the SM approach may give different posterior belief functions depending on whether updating is done one data point at a time, or the data sequence summarised in a single likelihood-based belief function. This problem was observed in early reviews of Shafer (1976) – see Shafer (1982, p.338).

that the method of inference should be specific to the context that gives rise to the model – he rejects the notion of a "one size fits all" approach to inference. Shafer (1982) analyses three specific contexts and proposes a different mode of inference for each. None coincides with SM. The context of our model is best captured by what Shafer calls "Models Composed of Independent Frequency Distributions" (Shafer, 1982, Section 3.1).[18] If we were to use the method proposed by Shafer (1982) for this context, instead of SM, our conclusions would be strengthened, since Shafer's alternative method resolves ambiguity even more rapidly than the SM approach. Details are available from the authors on request.

Appendix B
Proofs

It will be convenient to define the following upper and lower bounds on the probability of realising output M from technology β:

$$\underline{p} = \min_{\pi \in \Pi} \sum_{\theta \in \Theta} \pi(\theta)\,\theta$$

$$\overline{p} = \max_{\pi \in \Pi} \sum_{\theta \in \Theta} \pi(\theta)\,\theta$$

Note that $0 < \underline{p} < \overline{p} < 1$ by Assumptions 1 and 2. We also define

$$p^{\lambda} = \lambda \overline{p} + (1 - \lambda)\,\underline{p}$$

which is strictly increasing in λ and satisfies $p^{\lambda} \in (0, 1)$ for all $\lambda \in \left[0, \frac{1}{2}\right]$.

Proof of Theorem 3.1. It is easily verified that

$$U(s; \lambda) = \min_{p \in \left[p^{\lambda},\ p^{1-\lambda}\right]} p s_M + (1 - p)\, s_m \tag{12}$$

when $\lambda \leq \frac{1}{2}$.

We first prove necessity. We shall suppose that $s_m^1 \neq \min\{s_M^1, m\}$ and prove that a Pareto improvement is feasible.

[18] This method (as Shafer observes) was originally proposed by Smets (1978).

Let $u_1 = U(s^1; \lambda_1)$ be the certainty equivalent of s^1 for type λ_1. In particular,

$$u_1 \in \left[\min\left\{s_m^1, s_M^1\right\},\ \max\left\{s_m^1, s_M^1\right\}\right].$$

Since $s_m^1 \neq \min\{s_M^1, m\}$ and $s_m^1 \leq m$ we must have

$$s_m^1 < m \leq s_M^1 \leq M$$

when $\min\{s_M^1, m\} = m$ or

$$\min\left\{s_m^1, s_M^1\right\} < \max\left\{s_m^1, s_M^1\right\} \leq m < M$$

when $\min\{s_M^1, m\} = s_M^1$. It follows that

$$\alpha(u_1, u_1) + (1-\alpha)\left(s_M^1, s_m^1\right) \leq (M, m) \tag{13}$$

for $\alpha > 0$ sufficiently close to zero. It is therefore possible to move the λ_1 type's contract marginally in the direction of its certainty equivalent without violating feasibility. We shall show that this yields a Pareto improvement.

Choose $\alpha \in (0,1)$ small enough to satisfy (13) and define the new sharing rule

$$\begin{aligned} \hat{s}^1 &= \alpha(u_1, u_1) + (1-\alpha)s^1 \\ \hat{s}^2 &= (M, m) - \hat{s}^1. \end{aligned}$$

Then

$$\begin{aligned} U\left(\hat{s}^1; \lambda_1\right) &= \min_{p \in \left[p^{\lambda_1},\ p^{1-\lambda_1}\right]} p\hat{s}_M^1 + (1-p)\hat{s}_m^1 \\ &= \alpha u_1 + (1-\alpha)\left[\min_{p \in \left[p^{\lambda_1},\ p^{1-\lambda_1}\right]} ps_M^1 + (1-p)s_m^1\right] \\ &= u_1 \end{aligned}$$

Since p^λ is strictly increasing in λ, we have $p^{\lambda_2} > p^{\lambda_1}$ and $p^{1-\lambda_2} < p^{1-\lambda_1}$. Hence, using Aubin (1998, Proposition 4.4) we deduce:

$$\begin{aligned} \lim_{\alpha \downarrow 0} \frac{U\left(\hat{s}^2; \lambda_2\right) - U\left(s^2; \lambda_2\right)}{\alpha} &= \min_{p \in \left[p^{\lambda_2},\ p^{1-\lambda_2}\right]} p\hat{s}_M^1 + (1-p)\hat{s}_m^1 - u_1 \\ &> \min_{p \in \left[p^{\lambda_1},\ p^{1-\lambda_1}\right]} p\hat{s}_M^1 + (1-p)\hat{s}_m^1 - u_1 \\ &= 0 \end{aligned}$$

Therefore, if $\alpha \in (0,1)$ is small enough, we have a Pareto improvement. This proves the necessity part of the Theorem.

To prove sufficiency, we use an Edgeworth Box to depict all feasible sharing rules – see Figure 4. The contours of (12) are piecewise linear, with a kink at the decision-maker's certainty line. To establish the slopes of the various pieces, let $q^{\lambda} = p^{1-\lambda}$ and note that

$$\frac{q^{\lambda_1}}{1-q^{\lambda_1}} > \frac{q^{\lambda_2}}{1-q^{\lambda_2}} > \frac{p^{\lambda_2}}{1-p^{\lambda_2}} > \frac{p^{\lambda_1}}{1-p^{\lambda_1}}.$$

Figure 4 depicts one sharing rule with wage $w = s_M^1 = s_m^1$ and another with wage $w' = s_M^1 > m = s_m^1$. It is easy to see that each scenario is Pareto efficient. □

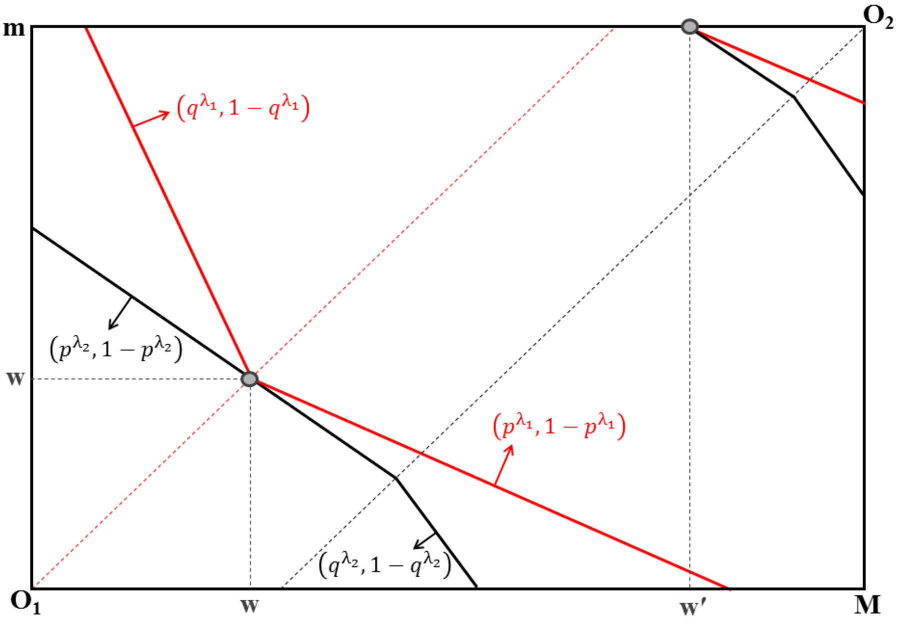

Figure 4: Pareto efficiency in an Edgeworth Box

Proof of Lemma 3.1. Consider a β firm paying wage $w < \frac{1}{2}(M+m)$. Let $s^{O,w} = (M-w,\ \max\{m-w,0\})$ be the owner's share and $s^{L,w} = (w,\ \min\{w,m\})$ be the worker's share. Defining $s^* = (M,m)$ we observe that

$$f(\theta; s^*) = f\left(\theta; s^{O,w}\right) + f\left(\theta; s^{L,w}\right)$$

for each $\theta \in \Theta$. Since $f\left(\cdot; s^{O,w}\right) : \Theta \to \mathbb{R}$ and $f\left(\cdot; s^{L,w}\right) : \Theta \to \mathbb{R}$ are *comonotone*, we have (Schmeidler, 1986):

$$U\left(s^*; \lambda\right) \;=\; U\left(s^{O,w}; \lambda\right) + U\left(s^{L,w}; \lambda\right). \tag{14}$$

The right-hand side of (14) is the sum, for an agent of type λ, of the utility from owning a β firm and the utility from working in one. These utilities may be expressed as follows:

$$U\left(s^{O,w}; \lambda\right) = \begin{cases} p^\lambda M + \left(1 - p^\lambda\right) m - w & \text{if } w \leq m \\ p^\lambda \left(M - w\right) & \text{if } w > m \end{cases} \tag{15}$$

and

$$U\left(s^{L,w}; \lambda\right) = \begin{cases} w & \text{if } w \leq m \\ p^\lambda w + \left(1 - p^\lambda\right) m & \text{if } w > m \end{cases} \tag{16}$$

For given w, both functions are linear in λ. The former is also strictly increasing in λ, while the latter is non-decreasing. From equation (14), the *average* of these two functions is equal to $\frac{1}{2}U\left(s^*; \lambda\right)$, so $U\left(s^*; \lambda\right)$ is strictly increasing in λ.

We next observe that the difference

$$U\left(s^{O,w}; \lambda\right) \;-\; U\left(s^{L,w}; \lambda\right) \;=\; \begin{cases} m \;-\; 2w \;+\; p^\lambda \left(M - m\right) & \text{if } w \leq m \\ \\ -m \;+\; p^\lambda \left[\left(M + m\right) - 2w\right] & \text{if } w > m \end{cases} \tag{17}$$

is strictly increasing in λ, since p^λ is strictly increasing in λ and

$$w \;<\; \frac{1}{2}\left(M + m\right).$$

Hence, for each $w < \frac{1}{2}\left(M + m\right)$, there exists a *unique* real number $a\left(w\right)$ (not necessarily in $\left[0, \frac{1}{2}\right]$) such that

$$U\left(s^{O,w};\; a\left(w\right)\right) \;=\; U\left(s^{L,w};\; a\left(w\right)\right) \;=\; \frac{1}{2}U\left(s^*;\; a\left(w\right)\right).$$

We have

$$U\left(s^{O,w}; z\right) \;>\; U\left(s^{L,w};\; z\right)$$

for any $z > a\left(w\right)$ and

$$U\left(s^{O,w}; z\right) \;<\; U\left(s^{L,w};\; z\right)$$

for any $z < a\left(w\right)$.

Finally, consider the piecewise linear function

$$G_w\left(z\right) \;=\; \max\left\{U\left(s^{O,w};\; z\right), U\left(s^{L,w};\; z\right)\right\}.$$

This gives the maximum return available to a type z agent from β occupations, given w. It is strictly increasing above $a\left(w\right)$ and weakly increasing below it. Figure 5 illustrates.

To complete the proof of the Lemma, we compare K with $G_w\left(z\right)$. There are two cases to consider.

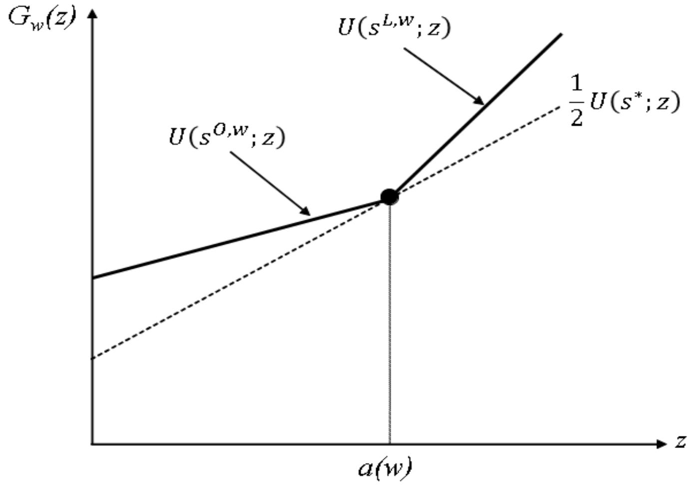

Figure 5: The piecewise linear function $G_w(z)$

Case I: $K \leq w \leq m$**.** In this case, $G_w(z) = w \geq K$ when $z \leq a(w)$, and $G_w(z) > w \geq K$ otherwise: wage w is not exposed to default, and no type λ strictly prefers to be employed in the α sector. See Figure 6.

Therefore, $\underline{\lambda}(w) = 0$ and $\overline{\lambda}(w) = \max\left\{0, \min\left\{a(w), \frac{1}{2}\right\}\right\}$. If $w = K$, then all types with $\lambda < \overline{\lambda}(w)$ are indifferent between working in a β firm and an occupation in α; otherwise, such types have a strict preference for working in a β firm.

Case II: $w > m$ **or** $w < K$**.** In this case, $G_w(z)$ is either strictly increasing or else $w < K$. It follows that there exists a *unique* $b(w)$ such that $K = G_w(b(w))$. Moreover, $G_w(z) > K$ for $z > b(w)$ and $G_w(z) < K$ for $z < b(w)$. Figure 7 illustrates a scenario with $w > m$.

If $b(w) > a(w)$, then $\overline{\lambda}(w) = \underline{\lambda}(w) = \max\left\{0, \min\left\{b(w), \frac{1}{2}\right\}\right\}$, while if $b(w) \leq a(w)$ we have $\overline{\lambda}(w) = \max\left\{0, \min\left\{a(w), \frac{1}{2}\right\}\right\}$ and $\underline{\lambda}(w) = \min\left\{\max\{0, b(w)\}, \frac{1}{2}\right\}$.

This completes the proof of Lemma 3.1. $\square$

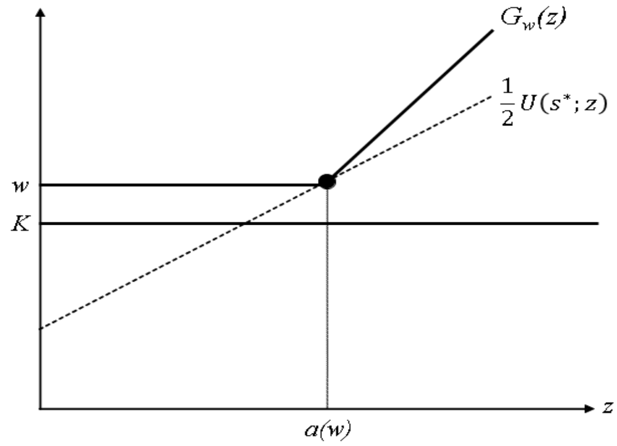

Figure 6: Case I with $w > K$

Proof of Theorem 3.2. The essence of the result is easily grasped using Lemma 3.1. From (15) and (16) we observe that

$$w > w' \quad \Rightarrow \quad U\left(s^{O,w}; z\right) > U\left(s^{O,w'}; z\right) \quad \text{for all } z \tag{18}$$

and

$$w > w' \quad \Rightarrow \quad a(w) > a(w') \tag{19}$$

It follows that an increase in w will strictly and continuously reduce net excess demand in the β labour market when types choose utility-maximising occupations.[19] Since excess demand is clearly positive when $w = 0$ and negative as $w \to \frac{1}{2}(M+m)$, it is easy to obtain a triple $(w, \lambda^*, \lambda^{**})$ that satifies (i)–(iv). The uniqueness properties follow directly from (18) and (19). Figure 8 illustrates an equilibrium.

A more formal argument runs as follows. We will construct an excess labour demand correspondence $\Lambda : \left[0, \frac{1}{2}(M+m)\right] \twoheadrightarrow [-1, 1]$ for the β labour market, and confirm that $0 \in \Lambda(w^*)$ for some

$$w^* \in \left[0, \frac{1}{2}(M+m)\right].$$

[19]Excess labour demand can be multi-valued, so these statements are somewhat loose, but can be made precise without altering their spirit.

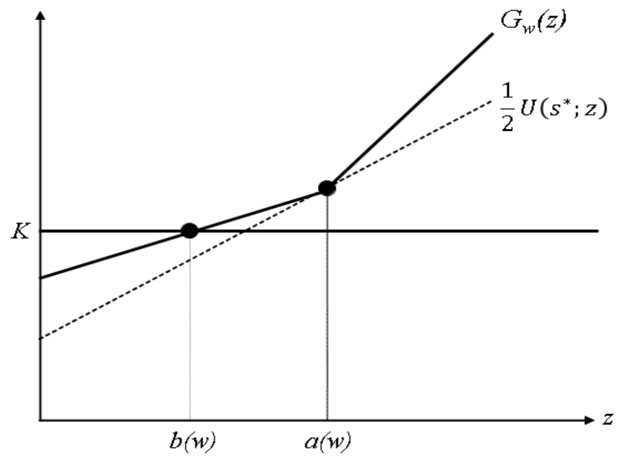

Figure 7: Case II with $w > m$

Define Λ as follows:

$$\Lambda(w) = \begin{cases} \left\{1 - 2H\left(\overline{\lambda}(w)\right) + H\left(\underline{\lambda}(w)\right)\right\} & \text{if } U\left(s^{O,w}; \overline{\lambda}(w)\right) > K \\ \left\{1 - 2H\left(\overline{\lambda}(w)\right) + H(\lambda) \mid \lambda \in \left[\underline{\lambda}(w), \overline{\lambda}(w)\right]\right\} & \text{otherwise} \end{cases}$$

Standard arguments confirm that Λ has closed graph, and that

$$\Lambda^{-1}(z) = \left\{ w \in \left[0, \frac{1}{2}(M+m)\right] \middle| z \in \Lambda(w) \right\}$$

is convex. Applying von Neumann's Intersection Lemma (Border, 1985, p.75) to the graph of Λ and the set

$$\left[0, \frac{1}{2}(M+m)\right] \times \{0\},$$

it follows that there exists a $w^* \in \left[0, \frac{1}{2}(M+m)\right]$ such that $0 \in \Lambda(w^*)$.

Using the monotonicity of H, we deduce that there exists a unique $\lambda^* \in \left[\underline{\lambda}(w^*), \overline{\lambda}(w^*)\right]$ such that $A \in BR(w, \lambda)$ for all $\lambda < \lambda^*$, $B_L \in BR(w, \lambda)$ for all $\lambda \in \left(\lambda^*, \overline{\lambda}(w^*)\right)$ and $B_O \in BR(w, \lambda)$ for all $\lambda > \overline{\lambda}(w^*)$.

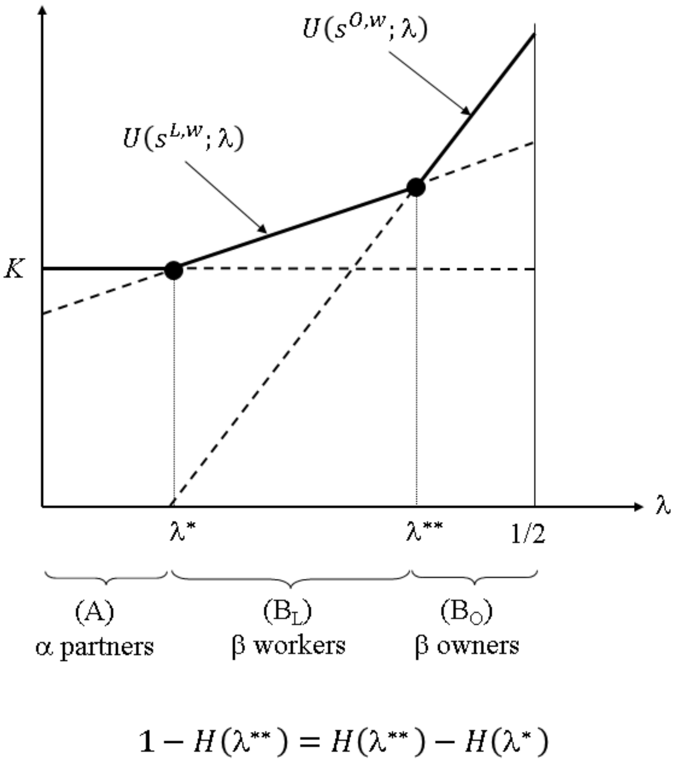

Figure 8: Equilibrium

Finally, let $\left(w^*, \lambda^*, \lambda^{**} = \overline{\lambda}\left(w^*\right)\right)$ describe an equilibrium as above, and assume it is *not* the case that $\lambda^{**} = \lambda^* = \frac{1}{2}$ (i.e., assume that a non-zero density of β firms operate in equilibrium). It suffices to show that there is no other w satisfying $0 \in \Gamma(w)$. Since β firms operate in the equilibrium described by $(w^*, \lambda^*, \lambda^{**})$, we have $\lambda^* < \lambda^{**} < \frac{1}{2}$. But the utility difference (17) is strictly decreasing in w and strictly increasing in λ, so any change to the wage rate must upset labour market equilibrium in the β sector. $\square$

References

Arrow, K. J. and L. Hurwicz (1972). "An Optimality Criterion for Decision-Making Under Ignorance." In *Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle*, C. F. Carter and J. L. Ford (eds). Oxford: Basil Blackwell.

Aubin, J.-P. (1998). *Optima and Equilibria: An Introduction to Nonlinear Analysis* (2nd edition). Berlin: Springer-Verlag.

Border, K. (1985). *Fixed Point Theorems with Application to Economics and Game Theory.* Cambridge: CUP.

DeMarzo, P., R. Kaniel and I. Kremer (2007). "Technological Innovation and Real Investment Booms and Busts," *Journal of Financial Economics* **85**, 735–754.

De Meza, D. and C. Southey (1996). "The Borrower's Curse: Optimism, Finance and Entrepreneurship," *Economic Journal* **106**, 375-386.

Dempster, A.P. (1967). "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics* **38**, 325–339.

Dushnitsky, G. (2010). "Entrepreneurial Optimism in the Market for Technological Inventions" *Organization Science* **21(1)**, 150-167.

Ellsberg, D. (1961). "Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics* **75**, 643-669.

Hendon, E., H.J. Jacobsen, B. Sloth and T. Tranæs (1994). "Expected Utility with Lower Probabilities," *Journal of Risk and Uncertainty* **8(2)**, 197-216.

Jaffray, J.-Y. (1989). "Linear Utility Theory for Belief Functions," *Operations Research Letters* **8**, 107-112.

Jaffray, J.-Y. (1991). "Linear Utility Theory and Belief Functions: A Discussion," in A. Chikán (ed.) *Progress in Decision, Utility and Risk Theory.* Dordrecht: Kluwer Academic.

Jaffray, J.-Y. (1994). "Dynamic Decision Making with Belief Functions", in R. R. Yager, J. Kacprzyk, & M. Fedrizzi (Eds.), *Advances in the Dempster-Shafer Theory of Evidence* (pp. 331–352). Wiley, New York.

Jaffray, J.-Y. and P. Wakker (1994). "Decision Making with Belief Functions: Compatibility and Incompatibility with the Sure-Thing Principle," *Journal of Risk and Uncertainty* **8**, 255–271.

Jensen, R. (1982). "Adoption and Diffusion of an Innovation of Uncertain Profitability," *Journal of Economic Theory* **27**, 182–193.

Luce, R.D. and H. Raiffa (1957). *Games and Decisions: Introduction and Critical Survey.* New York: John Wiley and Sons.

Manove M., and A.J. Padilla (1999): “Banking (Conservatively) with Optimists.” *RAND Journal of Economics*, 30(2), 324-350, Summer.

Pástor, L. and P. Veronesi (2006): “Was There a Nasdaq Bubble in the Late 1990’s?”, *Journal of Financial Economics* **81**, 61–100.

Rigotti, L., M.J. Ryan and R. Vaithianathan (2011). “Optimism and Firm Formation,” *Economic Theory*, **46(1)**: 1-38.

Schmeidler D. (1986). “Integral Representation without Additivity”, *Proceedings of the American Mathematical Society*, **97**: 255-261.

Seidenfeld, T. and L. Wasserman (1993). “Dilation for Sets of Probabilities”, *Annals of Statistics* **21(3)**, 1139-1154.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press.

Shafer, G. (1982). “Belief Functions and Parametric Models,” *Journal of the Royal Statistical Association, Series B* **44(3)**, 322-352.

Smets, P. (1978). “Un Modèle Mathématico-Statistique Simulant le Processus du Diagnostic Médical”, Doctoral dissertation, Free University of Brussels.

Wasserman, L.A. (1990): “Belief Functions and Statistical Inference”, *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* **18**, 183-196.