Ambiguity Aversion, Innovation and Investment Bubbles

Luca Rigotti* Matthew Ryan† Rhema Vaithianathan‡

DRAFT: PRELIMINARY AND INCOMPLETE

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: D5, D8, L2, J24, M13, O31

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

*Department of Economics, University of Pittsburgh, 4907 W.W. Posvar Hall, 230 S. Bouquet Street, Pittsburgh, PA 15260, USA. Email: luca@pitt.edu

†Department of Economics, University of Auckland, PB 92019 Auckland, New Zealand. Email: m.ryan@auckland.ac.nz

‡Department of Economics, University of Auckland, PB 92019 Auckland, New Zealand. Email: r.vaithianathan@auckland.ac.nz

1 Introduction

DeMarzo, Kaniel and Kremer (2007) define an "investment bubble" as a period of "excessive, and *predictably* unprofitable, investment" (p.737; emphasis in original). They observe that such bubbles are often associated with the arrival of some new technology, and cite the tech stock boom of the late 1990's – followed shortly thereafter by the "tech wreck" – as a prominent example. Why, they ask, do markets throw good money after bad in these circumstances?

To answer this question one must explain bubbles in *real* investment, as opposed to financial asset prices, and why such bubbles tend to be associated with new technologies. We provide such an explanation here. We show that investment bubbles may arise from learning dynamics in environments characterised by ambiguous prior information. Since new technologies confront decision-makers with imprecisely quantified risks, they provide particularly fertile ground for investment bubbles. Pástor and Veronesi (2006) provide evidence that the Nasdaq boom of the late 1990's was associated with particularly high uncertainty about the prospects of new technology firms.

If the market has an ambiguous prior on the returns to the new technology – prior beliefs cannot be described using a single probability – then what inferences should the market draw from the performance of the early adopters? This is not a standard statistical exercise and a generalisation of conventional Bayesian inference is required. Several generalisations have been discussed in the literature, including one introduced by Shafer (1976). We show that if all market participants use Shafer's method of inference, there may be increasing adoption of the new technology even when market data convey bad news. By "bad news", we mean that a frequentist analysis of the data would suggest that the new technology has inferior returns to the old. Moreover, any Bayesian who analysed the data using a precise prior probability would be increasingly pessimistic about the prospects for the new technology. Eventually, of course, the weight of evidence tilts boom into bust.

In our model, the investment bubble arises entirely from the process of learning from *public* information within a stable environment – no exogenous shocks or asymmetric information are required. There are two effects which drive the learning dynamics. First, the content of the news (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*, and is quite familiar. 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. The interaction of these two effects is sufficient to generate an investment bubble. In particular, ambiguity reduction ameliorates the impact of bad news and amplifies the impact of good news on "worst-case scenario" expectations; conversely for "best-case scenarios". Because of this asymmetry, data which contain predominantly bad news may nevertheless improve the lower bound on expected earnings. This underlies the boom phase of the investment bubble.

It follows that the boom in our model is not driven by the increasing exuberance of entrepreneurs, since entrepreneurial innovators focus on "best-case scenarios". Instead, it is their more conservative suppliers of capital – human and financial – who are emboldened to increase investment. If the ambiguity reduction effect dominates the news effect then these types become less reluctant to invest in the new technology, which fuels the boom.

Of course, ours is not the first explanation of investment bubbles. However, the mechanism that we propose is both novel and parsimonious. It also establishes a clear connection between investment bubbles and innovation. The next section reviews related literature and alternative mechanisms for investment bubbles. Section 3 introduces our model of ambiguity. We also describe how market participants draw statistical inferences when they have ambiguous priors, and how they make economic decisions in environments of ambiguous (or imprecise) risk. In section 4 we introduce a simple economic model with one good and two production technologies – an old technology which is well understood and a new technology of uncertain value. Agents in the model must choose an occupation. Wages and the number (actually, mass) of new technology adopters are determined endogenously in competitive equilibrium. Section 5 studies learning dynamics for this economy. Each period, the market observes the earnings of the new technology sector and uses this (public) information to update beliefs about its profitability. Using a simulated numerical example, we demonstrate the possibility of an investment bubble. 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 6. Two appendices contain the proofs of all results, plus supplementary material on statistical inference with ambiguous priors.

2 Related literature

In the model of DeMarzo, Kaniel and Kremer (2007), investment bubbles arise because producers self-insure against future price risk through over-investment in a risky new technology. They have a two-period model in which a homogeneous good is produced according to one of two technologies – a safe (old) technology or a risky (new) one – in the first period, then goods producers trade with service providers in the second period. The equilibrium relative price of services depends positively on aggregate output from the new technology sector. If agents are sufficiently risk-averse, this generates a positive correlation between aggregate investment in the new technology and the marginal utility of period two income. In the absence of complete markets, producers over-invest in the new technology as a self-insurance strategy. The dependence of marginal utility on the wealth of others, combined with incomplete markets, cause agents to act as if they are risk-seeking. They may even invest in the new technology to the point where expected returns are negative.

In the DeMarzo, Kaniel and Kremer (2007) model investment occurs only once – there is no increasing path of investment in response to market data – and only two periods, so there is no crash following the boom. Their focus is on providing an explanation for "predictably unprofitable" investment based on rational behaviour. However, the boom in their model only appears irrational if one assumes that all investors are risk-averse (as, of course, they are in the model). The investment bubble produced by our model is different. There is an increasing path of initial investment followed by a decline, and new technology earnings are first-order stochastically dominated by old technology earnings.

Barbarino and Jovanovic (2007) also model an investment bubble in which there is a run-up in investment over multiple periods, followed by a crash. Their model is based on a market whose demand function has a saturation point with unknown location. They explain the run-up in optimism prior to the crash as a type of 'Peso Problem' (Krasker, 1980). The distribution of the location of the saturation point exhibits a declining hazard rate, so the probability of reaching saturation at the next increment in supply is a declining function of the current level of supply. This produces a rising path of investment (and share price) despite a flat earnings profile, with a subsequent crash as saturation is reached.

Our story is complementary to that of Barbarino and Jovanovic. Theirs is a story of over-capacity: investments which make a good return initially are eventually stranded by excess capacity in the market. Ours is a story of an inferior technology (or an inferior business model) which nevertheless attracts increasing – though later declining – investment. In our model, new technology firms make lower profits (on average) *throughout the cycle.*

3 Ambiguity, inference and decision-making

Consider a real-valued random variable Y with associated density function $g(\cdot \mid \theta)$, where θ is an unknown parameter. Let the parameter (or state) space be $\Theta = \{\theta_1, ..., \theta_n\}$ and define $\Delta(\Theta)$ to be the set of probabilities on Θ. Given an observation y from Y, we define the likelihood function $l(\theta) = g(y \mid \theta)$. If we start with prior $\pi \in \Delta(\Theta)$, then we obtain the following posterior probability via Bayes' Rule:

$$\pi(\theta \mid y) \; = \; \frac{\pi(\theta)\, l(\theta)}{\sum_{\theta' \in \Theta} \pi(\theta')\, l(\theta')} \tag{1}$$

But what if we do not have a single prior? What if our prior information determines a *set* of probabilities, as in the famous experiments of Ellsberg (1961)? How should we form our posterior beliefs then?

These are the questions that we wish to address in the present section. We shall suppose that the prior information about θ comprises a (closed and convex) set $\Pi \subseteq \Delta(\Theta)$. The next section explains how such prior ambiguity may arise, and when Π may be summarised in the form of a *belief function* (Dempster, 1967; Shafer, 1976). Section 2.2 introduces an extension of Bayes' Rule (1) to the multiple-priors environment. This will be the basis of our model of learning in Section 3.

3.1 Ambiguity

Dempster (1967) describes a class of situations in which objective information about Θ leads naturally to a closed and convex set $\Pi \subseteq \Delta(\Theta)$ of priors. Dempster's framework is not the most general that one could imagine – it imposes additional structure on Π beyond closedness and convexity – but its explicitness about the source and nature of ambiguity makes it particularly useful for thinking about inference.[1] Moreover, there exist axiomatically grounded models of decision-making in environments of Dempsterian ambiguity – see Section 2.3.

Dempster imagines ambiguous information induced by a multi-valued mapping (correspondence) from a state space S, about which we have precise information, to Θ. Specifically, let

[1] See Shafer (1976) for theoretical arguments in support of Dempster's structure.

(S, Σ, μ) be a measure space and let $\Gamma : S \twoheadrightarrow \Theta$ be a non-empty-valued, measurable correspondence – the *information correspondence*. Our (objective) information about S is captured by the measure μ. Moreoever, if $s \in S$ is realised, then it is known that some $\theta \in \Gamma(s)$ also occurs, but nothing more than this – there is no information to pin down the relative likelihoods of states in $\Gamma(s)$.

Given (S, Σ, μ) and Γ, we define an associated *belief function* $\underline{v} : 2^{\Theta} \to [0,1]$ as follows:

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

We say that (S, Σ, μ, Γ) is a *source* for $\underline{v}$. The quantity $\underline{v}(E)$ is the lower probability of E: the smallest probability assigned to E according to the available information.[2] The probability of event E must lie in the interval $[\underline{v}(E), 1 - \underline{v}(E^c)]$, so[3]

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

is the set of probabilities consistent with the available information. Economists refer to Π as the *core* of the belief function $\underline{v}$.[4] Note that $\underline{v}$ may be recovered as the lower envelope of Π:

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

If Γ is singleton-valued (i.e., a function), then the source (S, Σ, μ, Γ) induces a probability on Θ and Π is a singleton. This is the case of *pure risk*. At the other extreme, if $\Gamma(s) = \Theta$ for every $s \in S$ – knowledge of the true state in S tells us nothing about θ – then $\Pi = \Delta(\Theta)$ and we have *pure uncertainty*. In between are situations such as Ellsberg's 3-colour experiment.[5]

The Ellsberg Experiment. A ball is to be drawn at random from 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). In this case, $\Theta = \{r, b, g\}$ but we can only assign objective probabilities to the elements of the partition $\{\{r\}, \{b, g\}\}$. To model this situation within Dempster's framework, let the points in S be the cells in the unambiguous partition, let $\Gamma(s)$ be the subset of Θ identified with s, and let $\mu(s)$ be the known probability of $\Gamma(s)$. Specifically: $S = \{s', s''\}$ with $\Sigma = 2^S$, $\Gamma(s') = \{r\}$, $\Gamma(s'') = \{b, g\}$, $\mu(s') = \frac{1}{3}$ and

[2] Observe that $\underline{v}(\emptyset) = 0$ (since Γ is non-empty-valued), $\underline{v}(\Theta) = 1$ and $\underline{v}$ is monotone: $E \subseteq F$ implies $\underline{v}(E) \leq \underline{v}(F)$.

[3] We define

$$\pi(E) := \sum_{\theta \in E} \pi(\theta).$$

[4] There is some potential for confusion, as Shafer (1976, p.40) uses the term "core" to refer to a different property of belief functions.

[5] Some decision-theorists refer to these intermediate cases as situations of *imprecise risk* – for example, Jaffray (1991) and Philippe (2002).

$\mu(s'') = \frac{2}{3}$. From this source we obtain the following belief function:

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

and associated set of probabilities:

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

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

Throughout the paper, we assume that information about Θ can be summarised in the form of a belief function; or equivalently, as a set of probabilities which coincide with the core of some belief function. It is important to note that this is a model of *objective* ambiguity. The set of probabilities that characterise the ambiguity about the true state is based on objective information – it is a property of the decision-making *environment* and can be determined by an outside observer; it does not describe the decision-maker's *subjective response* to the ambiguity.[6] In Section 2.3, we discuss individual decision-making in ambiguous environments of this sort.

3.2 Inference

Recall that y is an observation from the random variable Y, with associated likelihood function $l(\theta) = g(y \mid \theta)$. If our prior information determines a set $\Pi \subseteq \Delta(\Theta)$ of probabilities, how should we update Π on the basis of the new information contained in y? Unless Π is a singleton, this is a non-standard inference problem and there does not exist a consensus as to the most suitable approach. We shall here describe a method proposed by Shafer (1976, Chapter 11). For convenience, we shall refer to it as the Shafer Method (SM).

It should be noted that other inference procedures have also been studied in the literature. The best known alternative to SM inference is the robust Bayesian (RB) approach, in which each prior in Π is updated according to Bayes' Rule (1). Wasserman (1990) provides a detailed analysis and comparison of the RB and SM procedures. Both methods are equivalent to ordinary Bayesian inference when Π is a singleton. For our purposes, the most important distinction between them is that the DS approach "washes away" ambiguity more rapidly than the RB method. As we shall see, this will be important for our analysis of innovation. Indeed, a common criticism of RB

[6] Contrast Gilboa and Schmeidler's (1989) well-known *maxmim expected utility (MEU)* model. In MEU, decision-making takes place in an environment of pure uncertainty (so far as the outside observer is concerned) and decisions are guided by a set of *subjective* priors.

inference is that implausible degrees of ambiguity are retained in the posterior. New information may even *increase* the level of ambiguity under RB inference – in extreme cases resulting in the paradoxical phenomenon known as *dilation*. These matters are discussed in more detail in Appendix A. The interested reader should also consult Wasserman (1990).

The SM method can be applied only to sets Π that may be derived from a belief function. A *likelihood-based belief function* is constructed to encapsulate the new information from the sample data, then this is combined 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 from two independent sources, say (S, Σ, μ, Γ) and $(S', \Sigma', \mu', \Gamma')$. One first constructs a combined source

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

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

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

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 $\underline{v}$ is the belief function induced by (S, Σ, μ, Γ) and $\underline{v}'$ 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 $\underline{v} \oplus \underline{v}'$ and satisfies

$$\left(\underline{v} \oplus \underline{v}'\right)(E) \;=\; \frac{(\mu \times \mu')\left(\{(s, s') \in S \times S' \mid \Gamma^*(s) \subseteq E\}\right)}{(\mu \times \mu')(E^*)}$$

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

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

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

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

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

$$\underline{v}^l(E) \;=\; 1 \;-\; \frac{\sup_{\theta \in \Theta \setminus E} l(\theta)}{\sup_{\theta' \in \Theta} l(\theta')} \tag{2}$$

To help interpret the likelihood-based belief function (2), observe that $\underline{v}^l(E)$ is the contribution of E to achieving the maximum likelihood: $100\left[1 - \underline{v}^l(E)\right]\%$ of the maximum likelihood can be achieved within $\Theta \setminus E$. Wasserman (1990) discusses axiomatic foundations for (2).

To better understand the dynamics of SM inference, we consider a simple illustrative example.

A Bernoulli example. 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 $\underline{v}_1$ satisfying

$$\underline{v}_1(\{\theta_1\}) \ = \ \underline{v}_1(\{\theta_2\}) \ = \ 0.1.$$

The core of this belief function is

$$\Pi \ = \ \{\pi \in \Delta(\Theta) \ | \ \pi(\theta_1) \in [0.1, \ 0.9]\}.$$

We now observe a sequence $\{y_t\}_{t=1}^m$ of random draws from Y, where $y_t \in \{s, f\}$ with $y_t = s$ indicating a "success" in period t and $y_t = f$ a "failure". After each draw we apply the SM procedure to update the belief function. Let $\underline{v}_t$ denote the belief function prevailing at the start of period t. To simplify notation, we write $\underline{v}_{t,i}$ for $\underline{v}_t(\{\theta_i\})$.

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

$$\underline{v}_{t+1}(\{\theta_1\} \ | \ y_t = s) \ = \ \frac{\underline{v}_{t,1}\theta_1}{\underline{v}_{t,1}\theta_1 + \left(1 - \underline{v}_{t,1}\right)\theta_2} \ \le \ \underline{v}_{t,1} \tag{3}$$

$$\underline{v}_{t+1}(\{\theta_2\} \ | \ y_t = s) \ = \ \frac{\underline{v}_{t,2}\theta_2 + \left(1 - \underline{v}_{t,1} - \underline{v}_{t.1}\right)(\theta_2 - \theta_1)}{\underline{v}_{t,1}\theta_1 + \left(1 - \underline{v}_{t,1}\right)\theta_2} \ \ge \ \underline{v}_{t,2} \tag{4}$$

$$\underline{v}_{t+1}(\{\theta_1\} \ | \ y_t = f) \ = \ \frac{\underline{v}_{t,1}(1 - \theta_1) + \left(1 - \underline{v}_{t,1} - \underline{v}_{t,2}\right)(\theta_2 - \theta_1)}{\underline{v}_{t,2}(1 - \theta_2) + \left(1 - \underline{v}_{t,2}\right)(1 - \theta_1)} \ \ge \ \underline{v}_{t,1} \tag{5}$$

$$\underline{v}_{t+1}(\{\theta_2\} \ | \ y_t = f) \ = \ \frac{\underline{v}_{t,2}(1 - \theta_2)}{\underline{v}_{t,2}(1 - \theta_2) + \left(1 - \underline{v}_{t,2}\right)(1 - \theta_1)} \ \le \ \underline{v}_{t,2}. \tag{6}$$

These coincide with the standard Bayesian updating formulae when

$$\underline{v}_{t,1} \ + \ \underline{v}_{t,2} \ = \ 1$$

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

Suppose we need to make a decision whose payoff consequences depend on θ. We may therefore want to compute an expected value for this success probability. At the start of period t, this *expected success probability* lies in the interval

$$\left[\left(1 - \underline{v}_{t,2}\right)\theta_1 + \underline{v}_{t,2}\theta_2, \ \underline{v}_{t,1}\theta_1 + \left(1 - \underline{v}_{t,1}\right)\theta_2\right] \tag{7}$$

Observing a "success" in period t shifts both ends of this probability interval up; while "failure" in period t shifts both ends down. Hence, the mid-point of the interval also moves up or down depending on whether we observe a "success" or a "failure", respectively. This is what we call the *news effect*. It is familiar from Bayesian updating from a single prior. However, with multiple priors there is also an *ambiguity-reduction effect* – each piece of new data, whether a success or a failure, reduces the width of the interval. To see this, note that the DS update formulae imply

$$\underline{v}_{t+1}(\{\theta_1\} \mid y_t) + \underline{v}_{t+1}(\{\theta_2\} \mid y_t) \;\geq\; \underline{v}_{t,1} \;+\; \underline{v}_{t,2}$$

for any $y^t \in \{s, f\}$. This ensures that the width

$$\left(1 - \underline{v}_{t,1} - \underline{v}_{t,2}\right)(\theta_2 - \theta_1),$$

of the interval (7) will reduce conditional on any new piece of data. The ambiguity-reduction effect arises because the new data come from a perfectly reliable source, so ambiguity is washed away as new data accumulate.

Figure 1 plots a particular sequence $\{y_t\}_{t=1}^{9}$, and the associated sequence of intervals (7), for the case $\theta = \theta_1 = 0.2$ and $\theta_2 = 0.7$. The vertical bars represent the observations – a tall bar for a "success" and a short one for a "failure". The lines plot the upper and lower bounds for (7), as well as the mid-point.[7] There is also a horizontal line at the true θ value ($\theta_1 = 0.2$) for reference. Note that the interval is converging – as one would hope – to a single point at the true state.

Figure 1 illustrates how the news and ambiguity-reduction effects interact. When "good news" is received (i.e., $y_t = s$), these two effects are off-setting at the upper bound of the interval (7) and reinforcing at the lower bound. Conversely when "bad news" is received (i.e., $y_t = f$). Consider the path of the interval in Figure 1 from $t = 1$ to $t = 4$. The first 3 periods brought two pieces of bad news and one piece of good news. On balance, news has been bad, resulting in a lowering of the mid-point and of the upper bound from $t = 1$ to $t = 4$. However, the lower bound of the interval has *increased.* The accumulated ambiguity-reduction effect dominates the accumulated news effect at the lower bound. This phenomenon is central to the possibility of a boom-bust profile for inferior innovations, as we shall see.

3.3 Decision-making

Consider the following decision environment. Let $\mathcal{L}$ denote a convex set of *lotteries* – random variables with outcomes in $\mathcal{R}$. Objects of choice are *Anscombe-Aumann acts* of the form

$$f : \Theta \to \mathcal{L}.$$

There is objective information about Θ described by a belief function with core Π.

Jaffray (1989, 1991, 1994), Hendon *et al.* (1994) and Jaffray and Wakker (1994) axiomatise a wide class of decision models for this environment, including the famous the *Arrow-Hurwicz criterion* (Arrow and Hurwicz, 1972; Luce and Raiffa, 1957, p.282). Their axioms ensure that

[7] Recall that the period t interval is the one that prevails at the *start* of period t – before the period t observation is incorporated. Thus, the $t = 1$ interval is based on the prior.

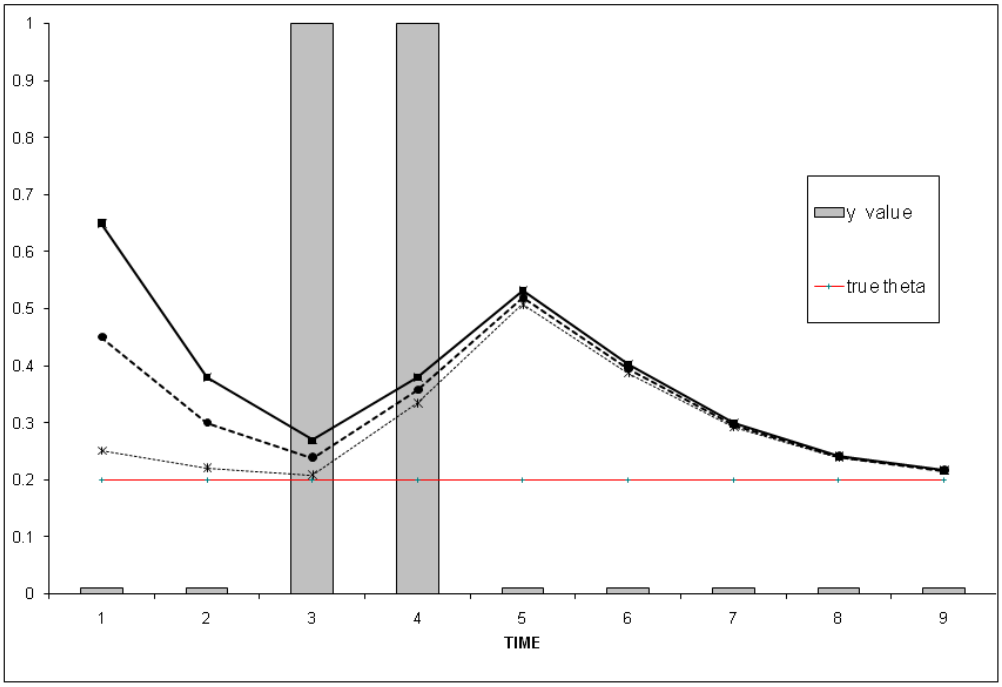

Figure 1: Updating with Dempster's rule

each decision-maker's risk preferences can be described using a von Neumann-Morgenstern utility function, $u : \mathcal{L} \rightarrow \mathcal{R}$. Act $f : \Theta \rightarrow \mathcal{L}$ induces a belief function on $\mathcal{R}$ with source

$$(S, \Sigma, \mu, u \circ f \circ \Gamma),$$

where (S, Σ, μ, Γ) is a source for the belief function on Θ and

$$(u \circ f \circ \Gamma)(s) \;=\; \{u(f(\theta)) \mid \theta \in \Gamma(s)\}$$

for each $s \in S$. Choosing amongst acts is therefore equivalent to choosing amongst belief functions.[8]

A preference ranking over acts is of the Arrow-Hurwicz form if there exists some $\lambda \in [0,1]$ such that the ranking can be represented by the following utility function:

$$U(f;\ \lambda) \;=\; \lambda \left[\max_{\pi \in \Pi} \sum_{\theta \in \Theta} \pi(\theta)\, u(f(\theta)) \right] \;+\; (1-\lambda) \left[\min_{\pi \in \Pi} \sum_{\theta \in \Theta} \pi(\theta)\, u(f(\theta)) \right] \tag{8}$$

[8] This generalises the idea of choosing amongst lotteries (pure risk) to environments with *imprecise risk*.

The parameter $\lambda \in [0,1]$ is the *ambiguity tolerance* (or *pessimism-optimism*) *index*. A decision-maker with $\lambda = 0$ (respectively, $\lambda = 1$) focusses all attention on the most pessimistic (respectively, optimistic) $\pi \in \Pi$.

Further discussion of the Arrow-Hurwicz choice criterion and related models of choice may be found in Rigotti, Ryan and Vaithianathan (2011, §2.3).

It will be useful to assume convexity of preferences in the sequel. That is, we will assume that the set

$$\{f \mid U(f;\lambda) \geq k\}$$

is convex for any $k \in \mathcal{R}$, where mixtures over Anscombe-Aumann acts are computed in the usual way. The following result shows that the preferences defined by (8) are convex provided ambiguity tolerance is sufficiently low.[9]

Theorem 3.1 *Let $\Theta = \{\theta_1, ..., \theta_n\}$ and $\Pi \subseteq \Delta(\Theta)$ be given, with Π the core of some belief function. There exists $\overline{\lambda} \in \left(0, \frac{1}{2}\right]$ such that the preferences described by (8) are convex for any $\lambda \leq \overline{\lambda}$. If $n = 2$, then preferences are convex iff $\lambda \leq \frac{1}{2}$.*

4 An entrepreneurial economy

In this section we build a simple model of firm formation, similar to that in Rigotti, Ryan and Vaithianathan (2011).

There is a continuum of agents who differ only in their tolerance for ambiguity. We let $H(z)$ denote the mass of agents with ambiguity tolerance $\lambda \leq z$, where H is a continuous distribution function on $[0,1]$ satisfying $H(0) = 0$ and $H(1) = 1$. Thus, the total mass of agents is unity. We further assume that H is differentiable and strictly increasing at any point x such that $H(x) < 1$. Since the distribution of ambiguity attitudes is therefore atomless, we exclude the scenario in Rigotti, Ryan and Vaithianathan (2011) where all agents have $\lambda \in \{0,1\}$. For convenience, we shall also make the following:

Assumption 1 *All agents are* ***risk neutral****. In particular, $u(f(\theta))$ is the expected value of the lottery $f(\theta)$, which we denote $\overline{f(\theta)}$.*

There is 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. We assume that the 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 decide to join forces to produce the consumption good using one of the available technologies.

A firm's output is a random variable whose distribution depends on the technology it employs plus an unknown parameter (state) $\theta \in \Theta = \{\theta_1, \theta_2\}$. There is ambiguous public information about the state, embodied in a closed and convex set of probabilities Π. It is further assumed that Π is the core of a belief function.

[9] Proofs of all results can be found in Appendix B.

Technology α is unambiguous, and generates expected output of $2K > 0$ units (i.e., K units *per capita*) independently of the state. The output of a β technology firm, on the other hand, is a random variable $R(\theta) \in \mathcal{L}$ that may depend non-trivially on θ. We shall assume that $R(\theta)$ is a Bernoulli random variable that takes value $M > 0$ with probability θ, and $m \in (0, M)$ with probability $1 - \theta$. The state θ represents a general, not a firm-specific shock. One may think of θ as the intrinsic productivity of technology β. The new technology is superior to the old if

$$\theta M + (1 - \theta) m \; > \; 2K$$

and inferior if the inequality is reversed. In fact, since agents are risk neutral, we can suppose that the return to the old technology is random, taking value M with probability $2K/M$ and zero otherwise.[10] In this case, the superior technology first-order stochastically dominates the inferior technology – investment in the inferior technology would be irrational for anyone who knew the true θ value and had strictly increasing utility of wealth, irrespective of their attitude to risk. The investment bubble that we obtain below therefore exhibits a very strong form of irrational investment behaviour, as perceived by an outside observer.

Agents in our economy are interested in the expected value of θ, so it will be convenient to define

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

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

and

$$\Delta p \; = \; \overline{p} - \underline{p}.$$

Thus, $\overline{p}$ is the upper bound, and $\underline{p}$ the lower bound, on the probability that a β technology firm realises output level M. The quantity Δp is the difference between these two extremes, and is a metric of level of ambiguity about the prospects of β firms.

Each agent must decide which sort of firm to join (an α firm or a β firm), and the partners in each firm must decide how to share the realised output. We make two assumptions about sharing rules. First, we impose the institutional constraint that sharing satisfy limited liability: agents commit no private funds to the firm – they cannot receive a negative payout in any state.[11] The limited liability clause implies that m is the highest sure level of consumption that can be guaranteed to any agent in a β technology firm.[12] Second, we require sharing rules to be Pareto efficient within each partnership.[13]

[10] We assume $2K \in (m, M)$ since otherwise it would be apparent *ex ante* that one technology dominates the other, making the problem uninteresting.

[11] For a setting with multiple priors but no limited liability restrictions, see Kelsey and Spanjers (2004).

[12] It would be straightforward to elaborate the model so that agents also have heterogeneous endowments of the consumption good, and must use these to meet obligations to their partners if necessary (as in Kihlstrom and Laffont, 1979). Wealthier owners may then be able to pay lower wages in equilibrium, because of the lower likelihood of default. Formally, this effect is similar to a credit constraint in a model where owners must invest capital in order to start their businesses, and credit markets are imperfect (Aghion and Bolton, 1997; Banerjee and Newman, 1993; Ghatak, Morelli and Sjöström, 2001).

[13] One may think of this as an additional equilibrium condition (cf, Definition 1 below).

Technology α produces $2K$ units (in expectation) in each state, so any sharing rule is Pareto efficient – recall that all agents are risk neutral. The Pareto optimal rules for β firms are not so obvious.

Consider two agents, one of type λ' and the other of type $\lambda'' > \lambda'$, operating the β technology. Let s'_M be the output granted to the type λ' agent when M is realised, and s'_m the output granted to the type λ' agent when m is realised. Define s''_M and s''_m similarly. Limited liability implies $s'_y, s''_y \in [0, y]$ and $s'_y + s''_y = y$ for each $y \in \{M, m\}$. The Pareto optimal rules depend on the ambiguity tolerance parameters λ' and λ''. If both partners have convex preferences, the Pareto optimal rules take a familiar form: the less ambiguity tolerant partner is offered a fixed wage $w = s'_M$, which is subject to default if $w > m$. The partner with the higher tolerance for ambiguity becomes the residual output claimant. Formally:

Theorem 4.1 *Suppose $\Delta p > 0$ and the agents have ambiguity tolerance parameters λ' and λ'' with $\lambda' < \lambda'' \leq \frac{1}{2}$. Then the sharing rule is Pareto efficient iff $s'_m = \min\{s'_M, m\}$.*

In view of Theorem 4.1, it is natural to describe the less ambiguity-tolerant partner in a firm as the "worker" and the more ambiguity-tolerant partner as the "owner". The payment s'_M will be called the "wage" paid (subject to limited liability) by the owner to the worker. We assume henceforth that output from β firms is shared using a fixed-wage contract of the sort described in Theorem 4.1, and we denote the wage by w.[14] We also assume that each partner in a firm using technology α receives expected payment K.[15]

4.1 Equilibrium

To establish equilibrium in our model, we need to assume that all agents have convex preferences. Therefore, recalling Theorem 3.1, we shall make:

Assumption 2 *The distribution of ambiguity tolerance satisfies $H\left(\frac{1}{2}\right) = 1$.*

Assumption 2 is analogous to excluding *risk-seeking* preferences in a model of behaviour in the face of pure risk.

We seek an equilibrium in occupational choices, taking the wage paid by β firms as the equilibrating parameter. In particular, we assume that all β firms offer the same wage. Since all β firms are identical – recall that θ represents a general, not a firm-specific, stochastic factor – if wages differed across β 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 the market wage, each individual chooses an occupation. They can either work in the α sector (and earn expected utility K), become the owner of a β firm or a worker in a β firm. An

[14] If $\Delta p = 0$, so there is no ambiguity about technology β, there are many other sharing rules which are Pareto efficient, but it is without loss of generality to focus on fixed-wage contracts.

[15] If there were "unequal treatment" in a non-zero mass of such firms, then the disadvantaged partners could leave and form new α firms that split returns equally. Each would be better off by doing so.

equilibrium (defined formally below) requires that "every" agent[16] makes a utility maximising occupation choice and the β labour market clears.[17]

Let $\mathcal{O} = \{a, b_w, b_o\}$ denote the set of occupations: a denotes an occupation in the α sector, b_w denotes working in a β firm and b_o denotes owning a β firm. Define $BR(w;\lambda) \subseteq \mathcal{O}$ to be the set of utility-maximising occupations for type λ given w.

An *allocation function* for the economy assigns individuals to occupations in a suitably measurable way. We economise on notation by supposing that all agents of a given type are allocated to the same occupation (*type-constant* allocation functions). We therefore define an allocation function to be a (Borel-measurable) mapping $\phi : \left[0, \frac{1}{2}\right] \to \mathcal{O}$, where $\phi(\lambda)$ is the occupation to which type λ agents are assigned. The restriction to type-constant allocations is without loss of essential generality because of the following result (and the fact that the type distribution is atomless):

Lemma 4.1 *If $\Delta p > 0$ and $w < \frac{1}{2}(M+m)$, 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

$$BR(w;\lambda) = \begin{cases} \{b_o\} & \text{if } \lambda > \overline{\lambda}(w) \\ \{b_w\} & \text{if } \lambda \in \left(\underline{\lambda}(w), \overline{\lambda}(w)\right) \text{ and } (w > m \text{ or } w \neq K) \\ \{a, b_w\} & \text{if } \lambda \in \left(\underline{\lambda}(w), \overline{\lambda}(w)\right) \text{ and } w = K \leq m \\ \{a\} & \text{if } \lambda < \tilde{\lambda}(w) \end{cases}$$

Under the conditions of Lemma 4.1, $BR(w;\lambda)$ is a singleton for all but a measure zero set of agents, unless $w = K \leq m$. In the latter case, types in $\left(\underline{\lambda}(w), \overline{\lambda}(w)\right)$ are indifferent between an occupation in an α firm and working in a β firm. Equilibrium at such a wage may require that the mass

$$H\left(\overline{\lambda}(w)\right) - H\left(\underline{\lambda}(w)\right)$$

of such agents be allocated between these two occupations in a particular manner. However, since H is continuous, any allocation of this mass can be achieved using a type-constant allocation.

Lemma 4.1 also exlcudes two trivial cases. First, economies in which there is no ambiguity about technology β. In this case, $BR(w;\lambda)$ is constant in λ and it is easy to observe – using the continuity of H – that nothing is lost by restricting attention to type-constant allocation

[16] Statements about "every agent", "all agents", "no agent" and so forth should be understood with the usual "except possibly for a set of measure 0" qualification.

[17] It is easy to check that all of our equilibria satisfy the *innovation-proofness* condition in Rigotti, Ryan and Vaithianathan (2011). Indeed, they also satisfy the following related condition: no pair of agents could leave their current occupations, form a new firm and both be better off.

functions. The second case is when $w \geq \frac{1}{2}(M+m)$. But there can never be an equilibrium with such a wage; at least, not one in which there is a non-zero mass of β firms. Since $M > m > 0$ by assumption, if $w \geq \frac{1}{2}(M+m)$, then workers in β firms receive *all* output when m is realised and *more than half* of the output when M is realised, so no agent would rationally choose to own a β firm.

The other important message of Lemma 4.1, which recalls a similar result in Rigotti, Ryan and Vaithianathan (2011), is that occupations are strictly ambiguity-ordered (unless $w = K \leq m$, in which case they are weakly ordered): only the most ambiguity-tolerant own β firms, the least ambiguity-tolerant are in α firms, and a middle group supply labour to β firms.

An equilibrium of our economy consists of a pair (w, ϕ) that satisfies two conditions. In the following, μ is the measure (on the Borel subsets of $\left[0, \frac{1}{2}\right]$) associated with the distribution function H.

Definition 1 *The couple* (w, ϕ) *is an* ***equilibrium*** *if*

(i) $\phi(\lambda) \in BR(w; \lambda)$ *for all* $\lambda \in \left[0, \frac{1}{2}\right]$*, and*

(ii) $\mu\left[\phi^{-1}(b_o)\right] = \mu\left[\phi^{-1}(b_w)\right]$.

Condition (i) says that each type is assigned to a utility-maximising occupation. Condition (ii) says that the β labour market clears.

The following is the main result of this section:

Theorem 4.2 *If* $\Delta p > 0$*, then an equilibrium* (w, ϕ) *exists and is essentially unique:* $\mu\left[\phi^{-1}(j)\right]$ *is unique for each* $j \in \mathcal{O}$ *and* w *is unique if* $\mu\left[\phi^{-1}(a)\right] < 1$.

If $\Delta p = 0$, then matters are even more straightforward. Since there is no ambiguity, all agents seek occupations in whichever industry offers the highest expected output, and output is shared equally (in expectation) within each firm. Therefore, equilibrium exists in this case also, and is unique in all "macro" aspects unless technologies α and β generate identical expected output.

5 Ambiguity and innovation

In this section, we add learning dynamics to the model. Over time, realised output from β firms provides information about θ. This information is public and agents incorporate new information according to the SM rule. Since information evolves over time, it is appropriate to introduce an index $t \in \{1, 2, ...\}$. In period t, agents will base their decisions on a common set Π_t of probabilities, which is just the prior $\Pi \equiv \Pi_1$ updated to incorporate all public information up to the start of period t. The belief function associated with Π_t is denoted $\underline{v}_t$. All β firms receive the same realisation $y_t \in \{M, m\}$ of the random variable $R(\theta)$ in period t.[18]

[18] In other words, there are no firm-specific shocks – each period generates a single observation from the random variable $R(\theta)$. If each firm receives an independent draw from $R(\theta)$, then a continuum of observations would be generated in each period and learning would cease after one round of updating.

To be clear, our indexing convention is that of Figure 1: y_t denotes the output level realised during period t and all other variables indexed by t indicate their values at the *start* of period t (i.e., *prior* to the observation of y_t).

New information may alter decision-making, so the economic equilibrium will also change. We summarise the essential features of the period t equilibrium (w_t, ϕ_t) using the triple

$$\left(w_t,\ \lambda_t^*,\ \overline{\lambda}_t = \overline{\lambda}_t\left(w_t\right)\right),$$

where $\overline{\lambda}\left(w\right)$ is defined in Lemma 1 and $\lambda = \lambda_t^*$ is the (unique) solution to

$$H\left(\lambda\right) \ = \ \mu\left[\phi_t^{-1}\left(a\right)\right].$$

Thus, agents with $\lambda > \overline{\lambda}_t$ own β firms, agents with $\lambda \in \left(\lambda_t^*,\ \overline{\lambda}_t\right)$ work in them, and agents with $\lambda < \lambda_t^*$ have occupations in the α technology sector. In particular,

$$\delta_t \ = \ 1 - H\left(\overline{\lambda}_t\right) \tag{9}$$

is the mass of β firms in the period t equilibrium. The time path of δ_t traces out the diffusion profile for the innovative technology. We shall show that it is possible for δ_t to rise substantially and then collapse when the new technology is inferior to the old.

In our model there are no entry costs, technology choices are freely reversible, all information is publicly available, and individuals cannot save – the consumption good is assumed to be perishable and there is no physical capital. We therefore have no dynamic trade-offs of the sort considered in Jovanovic (1982), Mookehrjee and Ray (1991), Sjöstrom (1991), Vettas (1998) or Banerjee and Newman (1993). There is only one respect in which decisions may be temporally dependent in our model. If $\delta_t = 0$, then agents learn nothing about β in period t. In this case, some types may have an incentive to experiment with technology β even though its expected return is less than K, in order to produce information that will alter the future values of equilibrium variables. We shall sidestep this potential source of temporal dependence also, by supposing the existence of a laboratory that publishes the results of tests on the new technology in each period (as in Jensen, 1982).

Suppose that $0 < \theta_1 < \theta_2 < 1$ and

$$\theta_2 M + \left(1 - \theta_2\right) m \ > \ 2K \ > \ \theta_1 M + \left(1 - \theta_1\right) m \tag{10}$$

Thus, β is a superior technology to α in state θ_2 and inferior in state θ_1. We will study the diffusion profile $\{\delta_t\}_{t=1}^{\infty}$ when $\theta = \theta_1$.

Figure 2 presents the results of two simple simulation exercises. We chose the following parameter values for the model: $M = 70$, $m = 10$, $\theta_1 = 0.45$, $\theta_2 = 0.5$, $K = 19.15$ and $\underline{v}_{1,1} = \underline{v}_{1,2} = 0.1$.[19] We also chose two different H distributions, one left-skewed and one right-skewed, and conducted separate simulation exercises for each.[20] Each simulation exercise proceeds

[19] In particular, (10) is satisfied.

[20] 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 $\left[0, \frac{1}{2}\right]$.

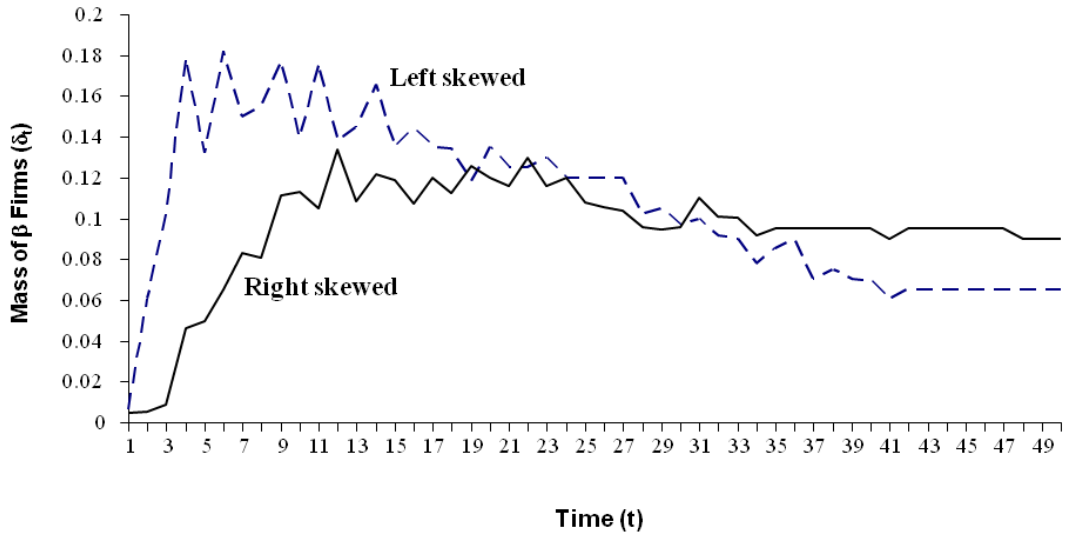

Figure 2: Diffusion profile for $\theta = \theta_1$ and λ distribution with right (solid) vs left (dashed) skew

as follows. First, we generate data $\{y_t\}_{t=1}^{50}$ by taking independent draws from the random variable $R(\theta_1)$. Using these data, we compute $\left\{\left(\underline{v}_{t,1},\ \underline{v}_{t,2}\right)\right\}_{t=1}^{50}$ according to the SM rule. We then solve the model at each t to find the equilibrium values $\left\{\overline{\lambda}_t\right\}_{t=1}^{50}$, and hence compute the diffusion path $\{\delta_t\}_{t=1}^{50}$ using (9). For each H distribution, we perform 100 such simulations and report the *average* diffusion path in Figure 2.

For each H distribution, we observe that diffusion trends upwards over the first few periods, before heading down. It is important to recall that Figure 2 reports *averaged* diffusion paths over 100 trials, so these shapes are not the artefacts of a particular sample path for y_t. On average, the data favour θ_1 – the average posterior probability on θ_1 would therefore increase monotonically in a standard Bayesian (single prior) model, giving a monotonically declining path for δ_t. The non-monotonicity evident in Figure 2 reveals that the ambiguity-reduction effect is at work during the "boom" phase.

To see what is going on, consider Figure 3, which plots the averaged values for $\underline{p}_t$ and

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

These are the lower bound and the mid-point, respectively, for the interval describing the expected "success" probability based on information up to the start of period t (recall the Bernoulli example from Section 3.2). A type λ agent makes decisions as if the probability of high output (M) from the β technology is

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

From Figure 3, we see that the path of $\underline{p}_t$ rises, even though the news is bad on average, since the ambiguity-reduction effect dominates the news effect. On the other hand, the path of

$$\frac{1}{2}\left(\overline{p}_t + \underline{p}_t\right)$$

declines, because the mid-point of the probability interval exhibits only the news effect. During the early phases of the diffusion, the least ambiguity tolerant agents in the β sector – those at the $\underline{\lambda}_t$ margin – feel more confident about the security of their income as time passes, so demand for jobs in the β sector rises at the current wage. This drives the "boom". Eventually, however, the marginal β owner – type $\overline{\lambda}_t$ – will be discouraged by the accumulating bad news, and the new technology sector begins to shed firms.

A left-skewed H distribution corresponds to a more ambiguity tolerant – more "entrepreneurial" – economy. Compared with the right-skewed distribution, it produces a sharper and more rapid "boom", with a correspondingly sharper decline that commences somewhat earlier (Figure 2). The average per period, *per capita* output for the more entrepreneurial (left-skewed) economy is 18.86, with a variance of 0.41. For the less entrepreneurial (right-skewed) economy, the average is 19.03 with a variance of 0.29. In the steady state, where all firms use the old technology, average *per capita* output is $K = 19.15$. Herein lies a cautionary tale: innovators embrace ambiguity, but they do not always back the right horses. The optimal degree of innovation in an economy may be a complex matter to judge, but there is no reason to expect that more entrepreneurs is necessarily better.

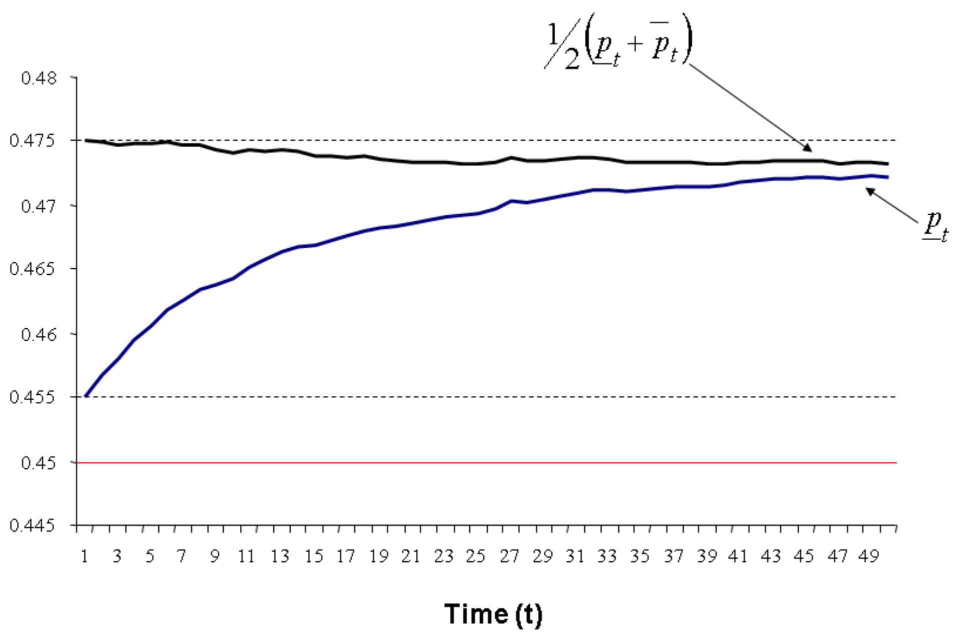

Figure 3: Evolution of the probability interval for the expected value of θ

6 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.[21] In particular, risks are technology-specific 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 precise 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

Let $y = (y_1, ..., y_m)$ be a random sample of observations from the random variable Y, with associated likelihood function

$$l(\theta) = \prod_{i=1}^{m} g(y_i \mid \theta) \tag{11}$$

Robust Bayesian (RB) inference applies Bayes' Rule to every prior in Π to construct a set of posteriors:

$$\{\pi(\cdot \mid y) \in \Delta(\Theta) \mid \pi \in \Pi\},$$

where $\pi(\cdot \mid y)$ is determined according to (1).[22] This procedure can be applied to any Π, whether or not it is the core of a belief function. The DS inference procedure described in Section 2.2 applies only to sets Π that may be derived from a belief function.

Both inference procedures share the following desirable features:

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

[22] To keep things simple, we assume that no prior can be logically excluded on the basis of sample y.

(i) When Π is a singleton, both are equivalent to conventional Bayesian inference.

(ii) The posterior is invariant to the order in which data are received – permuting y does not affect inference.[23]

Two important differences are:

(iii) The DS posterior probability interval for any event is always contained in the RB posterior interval (Wasserman, 1990, Theorem 7). Roughly speaking, DS inference resolves ambiguity more rapidly than RB inference.

(iv) With RB updating, the set of posteriors is the same, whether one updates sequentially (i.e., one data point at a time) or combines all data into a single likelihood function. With the DS method, these two calculations may yield different posterior sets.

Phenomenon (iii) is well-known. Phenomenon (iv) was observed in reviews of Shafer (1976) – see Shafer (1982, p.338) – but has received less attention.

A striking example of (iii) is Gelman (2006).

The Gelman Example. Imagine two urns. There are 200 balls and these are distributed across the two urns according to some unspecified process. It is known that each ball is either red (r) or black (b), and that by the end of the allocation process Urn I contains exactly 50 red and 50 black balls. One ball is drawn at random from each urn. Define the following random variables:

$$X = \begin{cases} 1 & \text{if the ball from Urn I is black} \\ 0 & \text{if the ball from Urn I is red} \end{cases}$$

$$Y = \begin{cases} 1 & \text{if the ball from Urn II is black} \\ 0 & \text{if the ball from Urn II is red} \end{cases}$$

For this problem, the state space is $\Theta = \{bb, br, rb, rr\}$ and the natural set of priors is[24]

$$\Pi = \left\{ \pi \in \Delta(\Theta) \;\middle|\; \pi(\{bb, br\}) = \frac{1}{2} \right\}$$

[23] In particular, Dempster's rule of combination is associative and commutative – Wasserman (1990, p.184).

[24] In Gelman's presentation, X is the outcome of a coin flip and Y is the outcome of an exhibition fight between a boxer and a wrestler. This context makes it reasonable to assume that X and Y are independent, so Gelman specifies a smaller set of priors:

$$\Pi' = \{\pi \in \Pi \mid \pi(\{bb\}) = \pi(\{rb\})\}.$$

However, what Gelman fails to observe is that this set is not the core of any belief function – as one can easily confirm by computing the Möbius inverse of its lower envelope (Chateauneuf and Jaffray, 1989) – so Gelman's application of DS inference is not strictly valid. Applying the RB method gives the same set of posteriors, whether one begins from Π or Π'. In the DS world, the independence of X and Y is already embodied in Π. This is the core of the belief function over (the product space) Θ which captures our information about X. Since our information about Y is described by the vacuous belief function (over Θ), the orthogonal sum of these two belief functions – a process which expresses the idea that the two sources of information are independent – yields a belief function whose core is also Π. Indeed, the treatment of independence in Dempster-Shafer theory has been the subject of criticism, most notably by Walley (1987).

Suppose we are informed that $X = Y$. That is, we receive reliable information that

$$\theta \in A = \{bb, rr\}.$$

How should we form our posterior beliefs about the colours of the two balls?

Applying the RB method gives the posterior set $\Delta(A)$ – the set of all probabilities on A – as they reader may easily verify. Thus, the posterior probability interval for any singleton subset of A is $[0, 1]$. **No ambiguity** about the relative likelihood of the states in A has been resolved by the new information.

Now consider DS updating. The set Π has the following lower envelope:

$$\underline{v}(E) = \begin{cases} 1 & \text{if } E = \Theta \\ \frac{1}{2} & \text{if } \{bb, br\} \subseteq E \neq \Theta \text{ or } \{rb, rr\} \subseteq E \neq \Theta \\ 0 & \text{otherwise} \end{cases}$$

It is easily verified that $\underline{v}$ is a belief function: a source for $\underline{v}$ is

$$\left(\{0, 1\},\ 2^{\{0,1\}},\ p,\ \Gamma\right) \tag{12}$$

where $p(0) = p(1) = \frac{1}{2}$, $\Gamma(0) = \{bb, br\}$ and $\Gamma(1) = \{rb, rr\}$. The inference problem is a straightforward conditioning exercise, so it is intuitively obvious how to encode the new information in the form of a belief function.[25] A natural source for this information is

$$\left(\{x\},\ 2^{\{x\}},\ q,\ \hat{\Gamma}\right),$$

where $q(x) = 1$ and $\hat{\Gamma}(x) = E$. That is, the additional source reliably informs us that $\theta \in E$, but nothing more than this. Applying Dempster's Rule of Combination gives a posterior belief function which satisfies

$$\underline{v}(\{bb\} | E) = \underline{v}(\{rr\} | E) = \frac{1}{2}.$$

This is a probability: the posterior probability interval for each singleton subset of E is $\left\{\frac{1}{2}\right\}$. **All ambiguity** about the relative likelihood of the states in E has been resolved by the new information.

The Gelman example not only illustrates an extreme case of phenomenon (iii), it also highlights the difficulty of choosing between the two procedures. Gelman regards each set of posteriors as implausibly extreme: he remarks that the RB posterior "seems wrong in that it has completely degraded our information about the coin flip" (*ibid.*, p.147) but opines that the DS posterior "does not seem right at all: coupling the fight outcome Y with the purely random X has caused the

[25] The following coincides with the *likelihood-based belief function* (2).

belief function for Y to collapse from pure ignorance to a simple 50/50 probability distribution" (*ibid.,* p.148). When Arthur Dempster was confronted with Gelman's example, however, he was unperturbed and felt that it supported the reasonableness of the DS method (*ibid.,* p.149). 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 seemingly paradoxical possibility of an event $A \subseteq \Theta$ and a partition $\mathcal{B}$ of Θ such that the posterior interval for A, *conditional on any* $B \in \mathcal{B}$, strictly contains the prior interval (Seidenfeld and Wasserman, 1993).[26]

To understand (iv), one must appreciate that the DS method is based on the idea of combining different sources of information with different degrees of precision or reliability. When accumulating information, the DS approach factors in the reliability of the source of each new datum. Suppose we begin with an ambiguous prior from a relatively unreliable source, and then receive a sequence of random draws from a precise (reliable) source. If all data are presented in the form of a single likelihood(-based belief) function, the DS method treats this as an inference problem with two sources – one vague and one precise. If updating occurs one data point at a time, then the DS method combines one vague source with as many precise sources as there are data points. The latter approach will, naturally enough, produce a less ambiguous posterior.

In the DS world, the ambiguity in the prior is progressively washed away by the accumulation of hard data from a reliable source. With robust Bayesian updating this may not be so. If the data happens to be agnostic about the true parameter value, then ambiguity persists under RB updating no matter how many observations have been made. For example, suppose one is running a binomial experiment with unknown success probability $\theta \in \left\{\hat{\theta}, 1-\hat{\theta}\right\}$, where $0 < \hat{\theta} < \frac{1}{2}$. After 2 million draws, exactly 1 million successes have been observed. The RB posterior set will be identical to the prior set, Π, as will the DS posterior if all data are incorporated via a single likelihood-based belief function.[27] However, if one applies the DS approach with sequential updating, one obtains a posterior set, all of whose members assign probability close to $\frac{1}{2}$ to each state. It seems rather natural that precise information should wash away ambiguity in this fashion. Shafer (1982, Examples 1 and 4) discusses inference from a binomial experiment with a completely vacuous prior, and likewise finds it reasonable that the posterior odds should converge to the Bayesian posterior odds derived from equal prior probabilities after large numbers of successes and failures have been observed.

We do not wish to claim via these observations that the DS method is preferable to the RB approach. Experts continue to dispute their relative merits. However, we *do* claim that there is no compelling reason to favour the robust Bayesian approach over DS inference in general,

[26] 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

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

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 $[0, 1]$.

[27] The latter is implied by Theorem 2 in Wasserman (1990).

and we further assert that the logic of the DS approach is consistent with the use of sequential updating.

Let us also briefly touch upon alternative methods of inference which use belief functions. Shafer (1982, Section 4) himself appears to have rejected the DS method, largely out of discomfort with phenomenon (iv).[28] Shafer does not propose a single alternative, but instead proposes that inference should be based on a description of the model $\{g(\cdot \mid \theta)\}_{\theta\in\Theta}$ through a belief function on $\Theta\times\mathcal{Y}$, where on $\mathcal{Y}$ is the sample space. The construction of this belief function will depend on the specific context that gives rise to the model. Shafer (1982) analyses three specific contexts, and proposes a different mode of inference for each. None coincides with the DS method, and none exhibits the controversial phenomenon (iv). Our context seems best captured by what Shafer calls "Models Composed of Independent Frequency Distributions" (Shafer, 1982, Section 3.1), so let us compare the DS method to Shafer's proposed method of inference for such models.[29]

The Bernoulli example revisited. Recall that Y is a Bernoulli random variable with success probability

$$\theta \in \Theta = \{\theta_1, \theta_2\},$$

where $\theta_1 < \theta_2$. The prior is described by a belief function $\underline{v}$. We define $\underline{v}_i = \underline{v}(\{\theta_i\})$ for convenience. The sample space (for a single observation) is $\mathcal{Y} = \{s, f\}$.

We first encode the model into a suitable belief function Bel on $\Theta\times\mathcal{Y}$. We require that $g(\cdot \mid \theta)$ coincides with Bel's conditional belief function on $\{\theta\}\times\mathcal{Y}$ for each $\theta\in\Theta$, so we construct what Shafer (1982, p.326) calls the *conditional embedding* of $g(\cdot \mid \theta)$ in $\Theta\times\mathcal{Y}$. Denote this by Bel_θ. The model $\{g(\cdot \mid \theta)\}_{\theta\in\Theta}$ is encoded as the orthogonal sum $\mathrm{Bel} = \mathrm{Bel}_{\theta_1} \oplus \mathrm{Bel}_{\theta_2}$. Given an observation $y \in \mathcal{Y}$, we may condition Bel on the event $[y] = \{(\theta_1, y), (\theta_2, y)\} \subseteq \Theta\times\mathcal{Y}$ by constructing the orthogonal sum $\mathrm{Bel}(\cdot \mid y) = \mathrm{Bel} \oplus \mathrm{Bel}_{[y]}$, where

$$\mathrm{Bel}_{[y]}(E) = \begin{cases} 1 & \text{if } [y] \subseteq E \\ 0 & \text{otherwise} \end{cases}$$

Given data on m independent observations $\{y_1, y_2, ..., y_m\}$, we may summarise this information in the form of the belief function[30]

$$\mathrm{Bel}(\cdot \mid \{y_1, y_2, ..., y_m\}) = \mathrm{Bel}(\cdot \mid y_1) \oplus \mathrm{Bel}(\cdot \mid y_2) \oplus \cdots \oplus \mathrm{Bel}(\cdot \mid y_m).$$

Next, we encode the prior into a belief function on $\Theta\times\mathcal{Y}$ through its *minimal extension* from Θ to $\Theta\times\mathcal{Y}$ (Shafer, 1982, p.325). Let Bel_1 denote the minimal extension of the prior.

[28] Consider our previous coin flipping example. Suppose we make observations after every second flip and update sequentially using the DS method, based on these observed pairs. That is, each update incorporates the newly observed *pair* in one step. If each observation contains one success and one failure, then this procedure leads to a vacuous posterior even after 1 million observations have been taken (Wasserman, 1990, Theorem 2).

[29] The latter method (as Shafer observes) is due to Smets (1978).

[30] The conditional belief function $\mathrm{Bel}(\cdot \mid \{y_1, y_2, ..., y_m\})$ could also be formed by conditionally embedding the model for m observations directly in $\Theta\times\mathcal{Y}^m$ and then conditioning on $\Theta\times\{(y_1, y_2, ..., y_m)\}$. This yields the same answer. Inference is the same, whether we update one data point at a time or all in one step – phenomenon (iv) does not arise under Smets' (1978) approach. See Shafer (1982, p.330).

Finally, we combine the information in the prior and the sample data to give

$$\mathrm{Bel}_{m+1}\left(\cdot \mid \{y_1, y_2, ..., y_m\}\right) = \mathrm{Bel}_1 \oplus \mathrm{Bel}\left(\cdot \mid \{y_1, y_2, ..., y_m\}\right).$$

The Θ-marginal of Bel_{m+1} is the posterior belief function.

Let us construct Bel_2 for our Bernoulli example. It is easiest to describe the various belief functions through their Möbius inverses.[31] Shafer (1982, p.329) gives that Bel has Möbius inverse m satisfying

$$\begin{aligned} m\left(\{(\theta_1, s), (\theta_2, s)\}\right) &= \theta_1\theta_2 \\ m\left(\{(\theta_1, s), (\theta_2, f)\}\right) &= \theta_1(1-\theta_2) \\ m\left(\{(\theta_1, f), (\theta_2, s)\}\right) &= (1-\theta_1)\theta_2 \\ m\left(\{(\theta_1, f), (\theta_2, f)\}\right) &= (1-\theta_1)(1-\theta_2) \end{aligned}$$

and therefore

$$\begin{aligned} \mathrm{Bel}\left(\{(\theta_1, s)\} \mid s\right) &= \frac{\theta_1(1-\theta_2)}{1-(1-\theta_1)(1-\theta_2)}; \\ \mathrm{Bel}\left(\{(\theta_2, s)\} \mid s\right) &= \frac{(1-\theta_1)\theta_2}{1-(1-\theta_1)(1-\theta_2)}; \\ \mathrm{Bel}\left(\{(\theta_1, f)\} \mid f\right) &= \frac{(1-\theta_1)\theta_2}{1-\theta_1\theta_2}; \\ \mathrm{Bel}\left(\{(\theta_2, f)\} \mid f\right) &= \frac{\theta_1(1-\theta_2)}{1-\theta_1\theta_2}. \end{aligned}$$

The extended prior belief function Bel_1 has associated Möbius inverse m_1 satisfying:

$$m_1\left(\{(\theta_i, s), (\theta_i, f)\}\right) = \underline{v}_i \qquad \text{for } i \in \{1, 2\}; \text{ and}$$

$$m_1(\Theta \times \mathcal{Y}) = 1 - \underline{v}_1 - \underline{v}_2.$$

Finally, we obtain the (Möbius inverse m_2 of the) posterior belief function $\mathrm{Bel}_2(\cdot \mid y)$ as follows:

$$\begin{aligned} m_2\left(\{(\theta, y)\} \mid y\right) &= \mathrm{Bel}_2\left(\{(\theta, y)\} \mid y\right) \\ m_2\left(\Theta \times \{y\} \mid y\right) &= 1 - \mathrm{Bel}_2\left(\{(\theta_1, y)\} \mid y\right) - \mathrm{Bel}_2\left(\{(\theta_2, y)\} \mid y\right) \end{aligned}$$

[31] If (S, Σ, μ, Γ) is a source for the belief function $\underline{v}$, then

$$m(E) = \mu\left(\{s \mid \Gamma(s) = E\}\right).$$

Thus,

$$\underline{v}(E) = \sum_{A \subseteq E} m(A).$$

See Chateauneuf and Jaffray (1989) for a thorough discussion of Möbius inversion.

so we have

$$m_2\left(\{(\theta_1,s)\}\mid s\right) = \frac{\theta_1\theta_2\underline{v}_1+\theta_1(1-\theta_2)(1-\underline{v}_2)}{\theta_2-(1-\theta_1)\theta_2\underline{v}_1+\theta_1(1-\theta_2)(1-\underline{v}_2)}$$

$$m_2\left(\{(\theta_2,s)\}\mid s\right) = \frac{(1-\theta_1)\theta_2(1-\underline{v}_1)+\theta_1\theta_2\underline{v}_2}{\theta_2-(1-\theta_1)\theta_2\underline{v}_1+\theta_1(1-\theta_2)(1-\underline{v}_2)}$$

$$m_2\left(\Theta\times\{s\}\mid s\right) = \frac{\theta_1\theta_2(1-\underline{v}_1-\underline{v}_2)}{\theta_2-(1-\theta_1)\theta_2\underline{v}_1+\theta_1(1-\theta_2)(1-\underline{v}_2)}$$

and

$$m_2\left(\{(\theta_1,f)\}\mid f\right) = \frac{(1-\theta_1)(1-\theta_2)\underline{v}_1+(1-\theta_1)\theta_2(1-\underline{v}_2)}{(1-\theta_2)-(1-\theta_2)\theta_1\underline{v}_1+\theta_2(1-\theta_1)(1-\underline{v}_2)}$$

$$m_2\left(\{(\theta_2,f)\}\mid f\right) = \frac{(1-\theta_1)(1-\theta_2)\underline{v}_2+\theta_1(1-\theta_2)(1-\underline{v}_1)}{(1-\theta_2)-(1-\theta_2)\theta_1\underline{v}_1+\theta_2(1-\theta_1)(1-\underline{v}_2)}$$

$$m_2\left(\Theta\times\{f\}\mid f\right) = \frac{(1-\theta_1)(1-\theta_2)(1-\underline{v}_1-\underline{v}_2)}{(1-\theta_2)-(1-\theta_2)\theta_1\underline{v}_1+\theta_2(1-\theta_1)(1-\underline{v}_2)}$$

These correspond to $\mathrm{Bel}(\cdot\mid y)$ when the prior is vacuous (i.e., $\underline{v}_1=\underline{v}_2=0$), and to the usual Bayesian posterior when $\underline{v}_1+\underline{v}_2=1$ (i.e., when the prior is a probability).

Note that this updating process satisfies the ambiguity-reduction property: any news reduces ambiguity. This follows from the fact that

$$m_2\left(\Theta\times\{y\}\mid y\right) \le 1-\underline{v}_1-\underline{v}_2$$

for each $y\in\{s,f\}$. See this as follows:

$$\begin{aligned}\frac{\theta_1\theta_2}{\theta_2-(1-\theta_1)\theta_2\underline{v}_1+\theta_1(1-\theta_2)(1-\underline{v}_2)} &= \frac{\theta_1\theta_2}{\theta_2+\theta_1-[\theta_2\underline{v}_1+\theta_1\underline{v}_2+\theta_1\theta_2(1-\underline{v}_1-\underline{v}_2)]}\\ &\le \frac{\theta_1\theta_2}{\theta_2+\theta_1-\theta_2} = \theta_2 \le 1.\end{aligned}$$

and

$$\frac{(1-\theta_1)(1-\theta_2)}{(1-\theta_2)-(1-\theta_2)\theta_1\underline{v}_1+\theta_2(1-\theta_1)(1-\underline{v}_2)} \le 1-\theta_1 \le 1.$$

Appendix B

Proofs

Proof of Theorem 3.1. Let $\Pi \subseteq \Delta(\Theta)$ denote the core of a belief function over the set $\Theta = \{\theta_1, ..., \theta_n\}$. The following facts are well-known (see, for example, Shapley, 1971): there exist $\{\pi_\rho \mid \rho \in P\}$, where P is the set of permutations of Θ (i.e., the set of one-to-one mappings $\rho : \{1, ..., n\} \to \Theta$), such that

$$\Pi \;=\; \mathrm{co}(\{\pi_\rho \mid \rho \in P\}),$$

where $\mathrm{co}(A)$ denotes the convex hull of A. Furthermore, defining

$$F^\rho \;=\; \{g : \Theta \to \mathcal{R} \mid g(\rho(1)) \leq \cdots \leq g(\rho(n))\}$$

for each $\rho \in P$,

$$\min_{\pi \in \Pi} \pi \cdot g \;=\; \pi_\rho \cdot g$$

for every $g \in F^\rho$, where $\pi \cdot g = \sum_{\theta \in \Theta} \pi_\rho(\theta)\, g(\theta)$. Given $\rho \in P$, we let $\rho^* \in P$ denote the "reverse" permutation: $\rho^*(i) = \rho(n - i + 1)$ for each $i \in \{1, ..., n\}$. It follows that, for any $f \in F^\rho$,

$$U(f;\ \lambda) \;= [\lambda \pi_{\rho^*} + (1-\lambda)\, \pi_\rho] \cdot (u \circ f) \tag{13}$$

Proposition 6.1 *Let*

$$\Pi^\lambda \;=\; \mathrm{co}(\{\lambda \pi_{\rho^*} + (1-\lambda)\, \pi_\rho \mid \rho \in P\}).$$

There exists $\overline{\lambda} > 0$ *such that, for any* $\lambda \leq \overline{\lambda}$,

$$U(f;\ \lambda) \;=\; \min_{\pi \in \Pi^\lambda} \pi \cdot (u \circ f)$$

for any f.

Proof. We need to show that

$$[\lambda \pi_{\rho^*} + (1-\lambda)\, \pi_\rho] \cdot g \;\leq\; \left[\lambda \pi_{\hat{\rho}^*} + (1-\lambda)\, \pi_{\hat{\rho}}\right] \cdot g \tag{14}$$

for every $\rho \in P$, every $g \in F^\rho$ and every $\hat{\rho} \in P$. Since (14) is trivial when $g \equiv 0$, and since scaling g by some strictly positive constant will not affect its membership of F^ρ or inequality (14), it suffices to restrict attention to functions in

$$D \;=\; \left\{ g : \Theta \to \mathcal{R} \;\middle|\; \sum_{\theta \in \Theta} |g(\theta)| = 1 \right\}.$$

Indeed, as D is polyhedral, it is enough to consider the extreme points of D, $\mathrm{ext}(D)$. That is, we must verify (14) for every $\rho \in P$, every $g \in F^\rho \cap \mathrm{ext}(D)$ and every $\hat{\rho} \in P$. Thus, we have finitely many inequalities of the form (14) to verify.

Note that (14) may be written

$$A(\rho, \hat{\rho}, g) \;+\; B(\rho, \hat{\rho}, g)\, \lambda \;\leq\; 0$$

with

$$A(\rho, \hat{\rho}, g) \;=\; (\pi_\rho - \pi_{\hat{\rho}}) \cdot g$$

and

$$B(\rho, \hat{\rho}, g) \;=\; (\pi_{\rho^*} - \pi_\rho) \cdot g \;-\; \left(\pi_{\hat{\rho}^*} - \pi_{\hat{\rho}}\right) \cdot g.$$

When $g \in F^\rho$, $A(\rho, \hat{\rho}, g) \leq 0$ and $B(\rho, \hat{\rho}, g) \geq 0$, with $B(\rho, \hat{\rho}, g) = 0$ iff $A(\rho, \hat{\rho}, g) = 0$. Thus,

$$\overline{\lambda} \;=\; \min_{\substack{\rho, \hat{\rho} \in P \\ g \in F^\rho \cap \text{ext}(D)}} \left\{ \frac{|A(\rho, \hat{\rho}, g)|}{B(\rho, \hat{\rho}, g)} \,\middle|\, B(\rho, \hat{\rho}, g) > 0 \right\} \;>\; 0$$

does the needful. □

Proposition 6.1 implies that an Arrow-Hurwicz decision-maker with sufficiently low ambiguity tolerance acts as a maxmin expected utility decision-maker with respect to the set Π^λ of probabilities (Gilboa and Schmeidler, 1989).[32] Such preferences are convex (*ibid.*, Axiom A.5). If we restrict the range of λ values in the economy to $\left[0, \overline{\lambda}\right]$ this will guarantee that all agents have convex preferences. The claim for $n = 2$ is easily demonstrated using (13) – we leave the details to hte reader. This completes the proof of Theorem 3.1. □

Proof of Theorem 4.1. Consider an agent of type λ who receives $s_M \in [0, M]$ when revenue M is realised, and $s_m \in [0, m]$ otherwise. Denote by s the state-contingent lottery corresponding to this sharing rule: $s(\theta)$ is a lottery delivering s_M with probability θ and s_m with probability $1 - \theta$. If $\lambda \leq \frac{1}{2}$, then one easily verifies that[33]

$$U(s;\, \lambda) \;=\; \min_{p \in \left[p^\lambda, q^\lambda\right]} p s_M \;+\; (1 - p)\, s_m \tag{15}$$

where

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

and

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

Observe that (15) is quasi-concave in (s_M, s_m). Since p^λ (respectively, q^λ) is increasing (respectively, decreasing) in λ, we also note that $p^{\lambda''} > p^{\lambda'}$ and $q^{\lambda''} < q^{\lambda'}$ when $\overline{p} > \underline{p}$ and $\lambda' < \lambda'' \leq \frac{1}{2}$.

We first prove necessity. Let $U\left(s';\, \lambda'\right) = u'$, so u' is the certainty equivalent of s' for type λ'. Suppose, contrary to what we wish to show, it is *not* the case that $s'_m = \min\{s'_M,\, m\}$. Then *either* $s'_m > s'_M$ *or* $s'_m < \min\{m, s'_M\}$. Hence, $u' < \max\{s'_m, s'_M\}$ and

$$\alpha\left(u',\, u'\right) \;+\; (1 - \alpha)\left(s'_M,\, s'_m\right) \;\leq\; (M,\, m) \tag{16}$$

[32] In fact, it is straightforward to show that Π^λ is *m-closed* – it contains all the probabilities that dominate the *lower probability* defined for all $E \subseteq \Pi^\lambda$ as follows: $\underline{p}(E) = \inf_{\pi \in \Pi^\lambda} \pi(E)$ – when $\lambda \leq \overline{\lambda}$, so these preferences also conform to the *Choquet expected utility* model (Schmeidler, 1989).

[33] If $\lambda > \frac{1}{2}$, then $p^\lambda \geq q^\lambda$ (with equality iff $\overline{p} = \underline{p}$) and

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

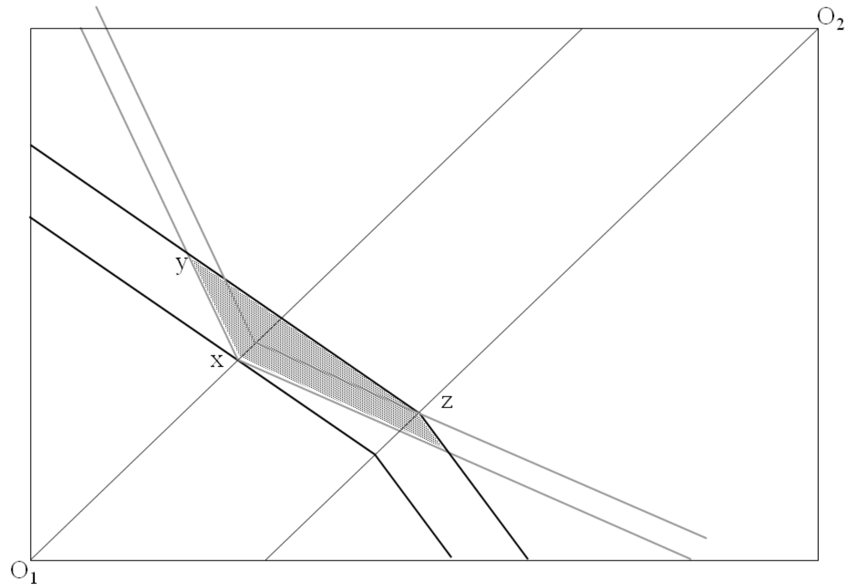

Figure 4: Pareto efficiency in an Edgeworth Box

for $\alpha \in (0,1)$ sufficiently close to zero. In other words, it is possible to move the λ'-type's contract marginally in the direction of her certainty equivalent without violating feasibility. We shall show that this yields a Pareto improvement.

Choose $\alpha \in (0,1)$ small enough to satisfy (16), and define

$$\left(s_M^{\alpha'},\, s_m^{\alpha'}\right) \;=\; \alpha\left(u',\, u'\right) \;+\; (1-\alpha)\left(s'_M,\, s'_m\right)$$

$$\left(s_M^{\alpha''},\, s_m^{\alpha''}\right) \;=\; (M,\, m) \;-\; \left(s_M^{\alpha'},\, s_m^{\alpha'}\right).$$

Recalling that that $p^{\lambda''} > p^{\lambda'}$ and $q^{\lambda''} < q^{\lambda'}$, we have

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

and (Aubin, 1998, Proposition 4.4)

$$\begin{aligned}
\lim_{\alpha\downarrow 0} \frac{U\left(s^{\alpha''};\; \lambda''\right) \;-\; U\left(s'';\; \lambda''\right)}{\alpha} \quad &= \quad \min_{p\in\left[p^{\lambda''},q^{\lambda''}\right]} p s_M^{\alpha'} \;+\; (1-p)\, s_m^{\alpha'} \;-\; u' \\
&> \quad \min_{p\in\left[p^{\lambda'},q^{\lambda'}\right]} p s_M^{\alpha'} \;+\; (1-p)\, s_m^{\alpha'} \;-\; u' \\
&= \quad 0.
\end{aligned}$$

Therefore, when $\alpha \in (0,1)$ is small enough, $\left(s_M^{\alpha'}, s_m^{\alpha'}\right)$ and $\left(s_M^{\alpha''}, s_m^{\alpha''}\right)$ Pareto improve on the original revenue sharing contracts. This proves the necessity part of the Proposition.

To see the sufficiency part, consider Figure 4. Here we have two agents with differing degrees of ambiguity tolerance. The grey indifference curves represent the relatively more ambiguity averse individual; the black indifference curves represent the relatively more ambiguity tolerant. Since no two individuals can have the same indifference curves, the picture represents a generic situation. One can easily see that a point like y cannot be Pareto optimal, since there is a large area (shaded) of allocations which are preferred by both individuals. This reasoning also applies to a point like z, which is situated on the certainty line of individual 2. Therefore, all that is left are points on the certainty line of individual 1. At a point like x, in fact, there are no mutually profitable trades available. □

Proof of Lemma 4.1. An agent who owns a β firm paying wage w obtains utility $U(\max\{0, R-w\};\ \lambda)$, while her worker gets $U(\min\{w, R\};\ \lambda)$.[34] Observe that expected output $\overline{R(\theta)}$ may be decomposed into the owner's and worker's expected shares as follows:

$$\overline{R(\theta)} \;=\; \overline{\max\{R(\theta)-w,\,0\}} \;+\; \overline{\min\{R(\theta),\,w\}}.$$

Each expected share is *comonotone* with expected output,[35] so we have (Schmeidler, 1986):

$$U(R;\,\lambda) \;=\; U(\max\{R-w,\,0\};\,\lambda) + U(\min\{R,\,w\};\,\lambda). \tag{17}$$

The right-hand side of (17) is the sum, for an agent of type λ, of the utility from owning a β firm and the utility from working in one. Defining $p^\lambda = \lambda\overline{p} + (1-\lambda)\underline{p}$, these utilities are, respectively:

$$U(\max\{R-w,\,0\};\,\lambda) = \begin{cases} p^\lambda M + \left(1-p^\lambda\right)m - w & \text{if } w \le m \\ p^\lambda(M-w) & \text{if } w > m \end{cases} \tag{18}$$

and

$$U(\min\{R,\,w\};\,\lambda) = \begin{cases} w & \text{if } w \le m \\ p^\lambda w + \left(1-p^\lambda\right)m & \text{if } w > m \end{cases} \tag{19}$$

Think of these as functions of (w, λ). 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 (17), the *average* of these two functions is equal to $\frac{1}{2}U(R;\,\lambda)$, so $\frac{1}{2}U(R;\,\lambda)$ is strictly increasing in λ.

We next observe that the difference

$$U(\max\{R-w,\,0\};\,\lambda) - U(\min\{R,\,w\};\,\lambda) \;=\; \begin{cases} m \,-\, 2w \,+\, p^\lambda(M-m) & \text{if } w \le m \\ \\ -m \,+\, p^\lambda\left[(M+m)-2w\right] & \text{if } w > m \end{cases} \tag{20}$$

[34] By $\min\{R, w\}$ we refer to the act that delivers, in state θ, a lottery that takes value $\min\{M, w\}$ with probability θ and $\min\{m, w\}$ with probability $1-\theta$. The act $\max\{R-w,\,0\}$ should be interpreted similarly.

[35] Functions $f : \Theta \to \mathcal{R}$ and $g : \Theta \to \mathcal{R}$ are *comonotone* provided

$$\left[f(\theta) - f(\theta')\right]\left[g(\theta) - g(\theta')\right] \;\ge\; 0$$

for all $\theta, \theta' \in \Theta$. Observe that the expected return of each participant in a β firm is weakly increasing in expected total output.

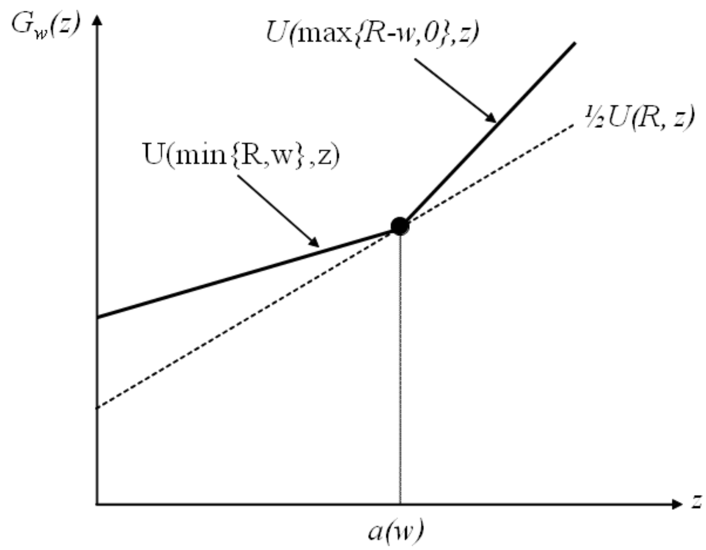

Figure 5: The piecewise linear function $G_w(z)$

is strictly increasing in λ, since p^λ is strictly increasing in λ ($\Delta p > 0$) and

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

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

$$U(\max\{R-w, 0\}; a(w)) = U(\min\{R, w\}; a(w)) = \frac{1}{2}U(R; a(w)).$$

For any $z > a(w)$,

$$U(\max\{R-w, 0\}; z) > U(\min\{R, w\}; z)$$

and for any $z < a(w)$,

$$U(\max\{R-w, 0\}; z) < U(\min\{R, w\}; z).$$

Finally, consider the piecewise linear function

$$G_w(z) = \max\{U(\max\{R-w, 0\}; z), U(\min\{R, w\}; z)\}.$$

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

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

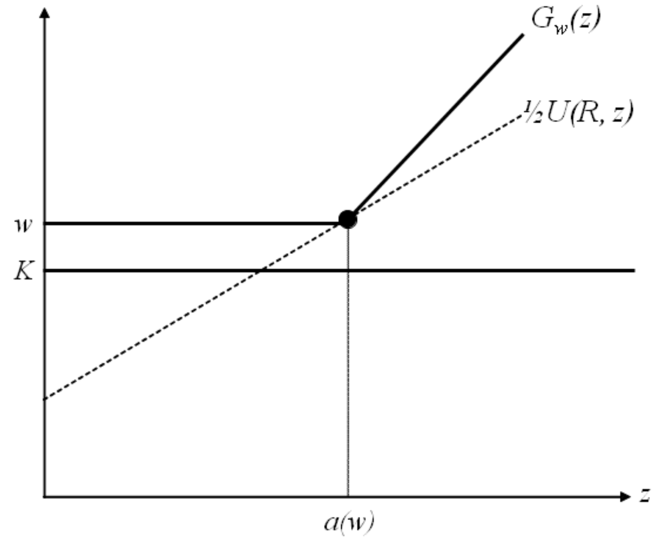

Figure 6: Case I with $w > K$

Case I: $K \le w \le m$**.** In this case, $G_w(z) = w \ge K$ when $z \le a(w)$, and $G_w(z) > w \ge K$ 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 $b(w) \le a(w)$ 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 4.1. □

Proof of Theorem 4.2. The essence of the result is easily grasped using Lemma 4.1. From (18) and (19) we observe that

$$w > w' \quad \Rightarrow \quad U(\min\{R, w\}; z) > U\left(\min\left\{R, w'\right\}; z\right) \text{ for all } z \tag{21}$$

and

$$w > w' \quad \Rightarrow \quad a(w) > a\left(w'\right) \tag{22}$$

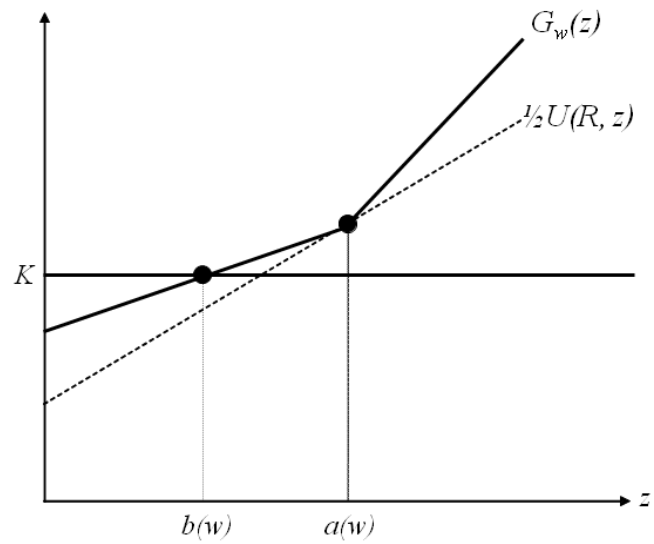

Figure 7: Case II with $w > m$

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.[36] 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 pair (w, ϕ) can be found that satifies (i) and (ii). The uniqueness properties of (w, ϕ) follow directly from (21) and (22). Figure 8 illustrates an equilibrium.

We now present a more formal proof of the existence and essential uniqueness of a pair (w, ϕ) satisfying (i) and (ii). To do so, we shall construct an excess labour demand correspondence $\Lambda : \left[0, \frac{1}{2}(M+m)\right] \twoheadrightarrow [-1, 1]$ for the β labor market, and confirm that $0 \in \Lambda(w^*)$ for some

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

Define Λ as follows:

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

Standard arguments confirm that Λ has closed graph, and that

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

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

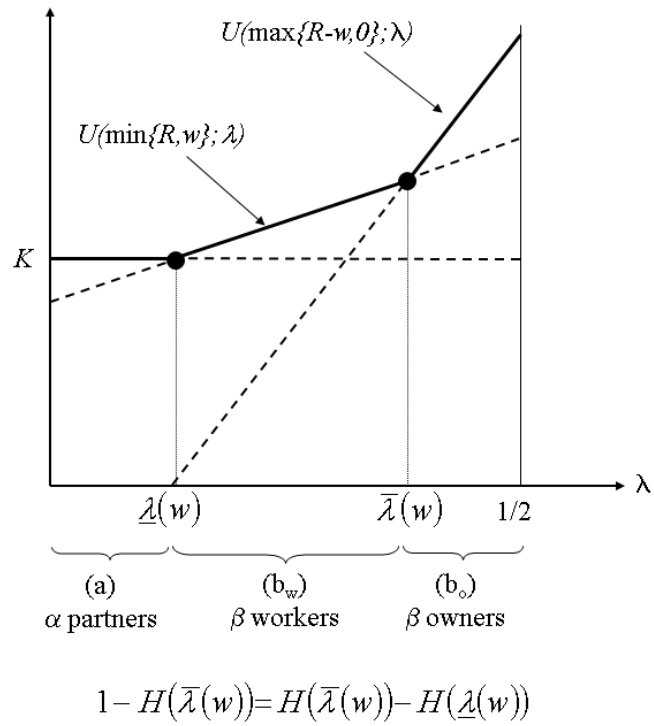

Figure 8: Equilibrium

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 also exists a unique $\lambda^* \in \left[\underline{\lambda}(w^*), \overline{\lambda}(w^*)\right]$ such that there is an equilibrium (w^*, ϕ^*) with $\phi^*(\lambda) = \{b_o\}$ for all $\lambda \in \left[\overline{\lambda}(w^*), \frac{1}{2}\right]$, $\phi^*(\lambda) = \{b_w\}$ for all $\lambda \in \left(\lambda^*, \overline{\lambda}(w^*)\right)$, and $\phi^*(\lambda) = \{a\}$ for all other λ.

Finally, let $\left(w^*, \overline{\lambda} = \overline{\lambda}(w^*), \lambda^*\right)$ describe an equilibrium as above, and assume it is *not* the case that $\overline{\lambda} = \lambda^* = \frac{1}{2}$ (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 $\left(w^*, \overline{\lambda}, \lambda^*\right)$, we have $\lambda^* < \overline{\lambda} < \frac{1}{2}$. But the utility difference (20) 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. □

References

Aghion, P and P. Bolton (1997). "A Theory of Trickle-Down Growth and Development", *Review of Economic Studies* **64:**151 -172**.**

Aghion, P and P. Howitt (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, **60**, 323-51.

Anscombe, F.J. and R.J. Aumann (1963). "A Definition of Subjective Probability", *Annals of Mathematical Statistics* **34**, 199-205.

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.

Balcer, Y. and S. A. Lippman (1984) "Technological Expectations and Adoption of Improved Technology" *Journal of Economic Theory* **34**, 292-318.

Banerjee, A. V and A.F. Newman (1993). "Occupational Choice and the Process of Development", *Journal of Political Economy* **101(2)**, 274 - 298.

Barbarino, A. and B. Jovanovic (2007): "Shakeouts and Market Crashes," *International Economic Review* **48(2)**, 385-420.

Berger, J.O. (1996): *Bayesian Robustness*, Institute of Mathematical Statistics Hayward, California.

Bernardo, A.E. and I. Welch (2001). "On the Evolution of Overconfidence and Entrepreneurs" *Journal of Economics and Management Strategy* **10,** 301–330**.**

Blanchflower, D. and A. Oswald (1998): "What Makes an Entrepreneur?", *Journal of Labor Economics* **16**, 26-60.

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

Bose, S., Ozdenoren, E. and Pape (2006): "Optimal Auctions with Ambiguity", *Theoretical Economics* **1**, 411-438

Calvo, G.A. and S. Wellisz (1980). "Technology, Entrepreneurs, and Firm Size," *Quarterly Journal of Economics* **95**, 663–677.

Camerer, C. (1995). "Individual Decision Making," in *The Handbook of Experimental Economics*, J.H. Kagel and A.E. Roth (eds.). Princeton: Princeton University Press.

Chateauneuf, A., Dana R.A., and J.M. Tallon (2000): "Optimal Risk-Sharing Rules and Equilibria with Choquet-Expected-Utility", *Journal of Mathematical Economics* **34**, 191–214.

Chateauneuf, A. and J.-Y. Jaffray (1989): "Some Characterizations of Lower Probabilities and Other Monotone Capacities through the use of Möbius Inversion," *Mathematical Social Sciences* **17(3)**, 263-283.

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

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

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

Evans, D.S. and B. Jovanovic (1989). "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," *Journal of Political Economy* **97**, 808–827.

Evans, D.S., Jovanovic, B. and L.S. Leighton (1989) "Some Empirical Aspects of Entrepreneurship" *American Economic Review* **79**, 519–535.

Fox, C.R., and A. Tversky (1995): "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics* **110**, 585-603.

Geanakplos, J (2003). "Liquidity, Default and Crashes: Endogenous Contracts in General Equilibrium" in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume II*, M. Dewatripont, L.P. Hansen and S.J. Turnovsky (eds.). Cambridge: Cambridge University Press.

Gelman, A. (2006): "The Boxer, the Wrestler, and the Coin Flip: A Paradox of Robust Bayesian Inference and Belief Functions," *The American Statistician* **60(2)**, 146-150.

Ghatak, M., Morelli, M. and T. Sjostroem (2001): "Occupational Choice and Dynamic Incentives," *Review of Economic Studies* **68**, 781-810.

Ghirardato, P. (2001). "Coping With Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty," *Economic Theory* **17**, 247-276.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility With Non-Unique Prior," *Journal of Mathematical Economics* 18, 141-153.

Gort, M. and S. Klepper (1982): "Time Paths in the Diffusion of Product Innovation", *Economic Journal* **92**, 630-653.

Griliches, Z. (1957). "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* **25**, 501-522.

Gruner, H.P (2003). "Redistribution as a Selection Device", *Journal of Economic Theory* **108**, 194–216.

Guiso, L., Sapienza, P. and L. Zingales (2004): "The Role of Social Capital in Financial Development," *American Economic Review* **94**, 526-556.

Heath, C. and A. Tversky (1991). "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty", *Journal of Risk and Uncertainty* **4**, 5–29.

Holtz-Eakin, D., Joulfaian, D., and H. Rosen (1994a): "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints", *Journal of Political Economy* **102**, 53-75.

Holtz-Eakin, D., Joulfaian, D., and H. Rosen (1994b): "Entrepreneurial Decisions and Liquidity Constraints," *RAND Journal of Economics* **25**, 334-347.

Huang, R. (2008): "Tolerance for Uncertainty and the Growth of Informationally Opaque Industries," *Journal of Development Economics* **87(2)**, 333-353.

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 F. Phillipe. (1997). "On the Existence of Subjective Upper and Lower Probabilities," *Mathematics of Operations Research* **22**, 165-185.

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.

Jensen, R. (1983). "Innovation Adoption and Diffusion When There Are Competing Innovations," *Journal of Economic Theory* **29**, 161-171.

Jovanovic, B. and S. Lach (1989): "Entry, Exit, and Diffusion with Learning by Doing," *American Economic Review*, **79**, 690-699.

Jovanovic, B. and G. M. MacDonald (1994): "The Life Cycle of a Competitive Industry," *Journal of Political Economy* **102**, 322-347.

KELSEY, D. AND W. SPANJERS. (1997). "Uncertainty in Partnerships," *Economic Journal*, **114**, 528-54.

KIHLSTROM, R.E. AND J.-J. LAFFONT. (1979). "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy* **87**, 719–748.

KING, R. G. AND R. LEVINE (1993): "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics* **108**, 717-37.

KNIGHT, F. (1921). *Risk, Uncertainty and Profit.* Cambridge (MA): Houghton Mifflin.

LAZEAR, E. (2002). "Entrepreneurship", NBER Working Paper 9109.

LLOYD-ELLIS, H. AND D. BERNHARDT (2000): "Enterprise, Inequality and Economic Development", *Review of Economic Studies* **67**, 147-168.

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

MANSFIELD, E. (1968). *Economics of Technological Change.* Norton.

MOSKOWITZ, T.J. AND A. VISSING-JØRGENSEN (2002). "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?", *American Economic Review* **92**, 745–778.

MUKERJI, S. (1997). "Understanding the Non-Additive Probability Model," *Economic Theory* **9**, 23–46.

MUKERJI, S. (1998). "Ambiguity Aversion and Incompleteness of Contractual Form," *American Economic Review* **88**, 1207–1231.

MUKERJI, S. AND J.-M. TALLON (2001). "Ambiguity Aversion and Incompleteness of Financial Markets," *Review of Economic Studies* **68**, 883–904.

PASTOR, L. AND P. VERONESI (2006): "Was There a Nasdaq Bubble in the Late 1990's?", *Journal of Financial Economics* **81**, 61–100.

PHILIPPE, F. (2002). "Tools for Decision Making Under Imprecise Risk," *Statistical Papers* **43**, 95–110.

PHILIPPE, F., G. DEBS AND J.-Y. JAFFRAY (1999). "Decision Making with Monotone Lower Probabilities of Infinite Order," *Mathematics of Operations Research* **24**, 767–784.

PIKKETY, T. (1997): "The Dynamics of the Wealth Distribution and the Interest Rate with Credit-Rationing," *Review of Economic Studies* **64**, 173-189.

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.

SCHMEIDLER D. (1989): "Subjective Probability and Expected Utility without Additivity", *Econometrica*, **57**: 571–587.

SCHUMPETER, J.A. (1947): "The Creative Response in Economic History," *Journal of Economic History* **7**, 149-159.

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.

SHAPLEY, L.S. (1971): "Cores of convex games", *International Journal of Game Theory*, **1**: 11–26.

SRIVASTAVA, R.P. AND T.J. MOCK (2002). *Belief Functions in Business Decisions.* New York: Physica-Verlag.

VETTAS, N. (1998): "Demand and Supply in New Markets: Diffusion and Bilateral Learning," *RAND Journal of Economics* **29**, 215–233.

VIERO, M.-L. (2006): "Exactly What Happens After the Anscombe-Aumann Race? Representing Preferences in Vague Environments", QED Working Paper No. 1094

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

WASSERMAN, L.A. AND KADANE, J.B. (1990): "Prior Envelopes Based on Belief Functions", *The Annals of Statistics*, **18**: 454–464.

WASSERMAN, L.A. AND KADANE, J.B. (1992): "Symmetric Upper Probabilities", *The Annals of Statistics*, **20**: 1720–1736.

WASSERMAN, L.A. AND KADANE, J.B. (1996): "Symmetric, Coherent, Choquet Capacities", *The Annals of Statistics*, **24**: 1250–1264.