Asymmetric Information and Adverse Selection

Heski Bar-Isaac$ \quad \quad $ Ian Jewitt$^\dagger \quad \quad $ Clare Leaver$^\$ 

January 17, 2014

Abstract

This paper develops a framework for the analysis of how asymmetric information impacts on adverse selection and market efficiency. We adopt Akerlof’s (1970) unit-demand model extended to a setting with multidimensional public and private information. Adverse selection and efficiency are defined quantitatively as real valued random variables. We characterize how public information disclosure and private information acquisition affect the relationship between adverse selection and efficiency. These results are applied to inform welfare and empirical analysis and, in an employer learning setting, to study the endogenous choice of information structures. Equilibrium information structures impose adverse selection efficiently. We show that this makes adverse selection hard to detect using standard positive correlation tests.

**Keywords:** asymmetric information, adverse selection, information structures, information acquisition, information disclosure, employer learning.

**JEL Classification:** D82, J30.

---

*We are grateful to very many people for helpful comments, in particular Ignacio Esponda, Paul Klemperer, Alessandro Lizzeri and Margaret Meyer. We also thank participants at the numerous conferences and seminars where versions of this paper, or an earlier overlapping paper Bar-Isaac, Jewitt and Leaver (2007), have been presented.

$^\dagger$Nuffield College, Oxford. Email: ian.jewitt@nuffield.ox.ac.uk. $^\$Univeristy of Oxford. Email: clare.leaver@economics.ox.ac.uk. $^\$University of Toronto. Email: heski.bar-isaac@rotman.utoronto.ca.
## Contents

1 Introduction .......................................................... 3

2 The Model ..................................................................... 9
   2.1 Description .................................................................. 9
   2.2 Statistical Assumptions and Definitions .......................... 11
   2.3 The Gaussian Case ..................................................... 12

3 Equilibrium ................................................................... 13
   3.1 Regression to the Mean and Willingness to Pay ................. 13
   3.2 Characterization of Equilibrium Market Outcomes ............... 14
   3.3 Insurance Interpretation ............................................. 15

4 Adverse Selection and Efficiency: Representation and Applications 16
   4.1 Definitions ............................................................... 16
   4.2 Representation Results ............................................... 17
   4.3 Applications ............................................................ 20
      4.3.1 Released workers may be paid more than retained workers .. 20
      4.3.2 Finding adverse selection when public information is not observed by the econometrician .............. 21
      4.3.3 Adverse selection falls mainly on trades that do not take place ... 22
      4.3.4 It is optimal to mandate when adverse selection is low rather than high 23
      4.3.5 There are three dimensions of information (Gaussian case) .... 24

5 Adverse Selection and Overall (Expected) Efficiency .................. 25
   5.1 General Results ...................................................... 26
   5.2 A Simple Case: Public Information Orthogonal to Match Quality ... 28

6 Imposing Adverse Selection Efficiently .................................. 29
   6.1 Preliminaries ......................................................... 30
   6.2 Disclosure with Fixed Private Information ........................ 31
   6.3 Information Acquisition ............................................ 33
   6.4 Endogenous Information and its Consequences for Wages and Tests for Adverse Selection ....................... 35

7 Conclusion .................................................................... 37
1 Introduction

This paper develops a framework for the analysis of how asymmetric information impacts on adverse selection and market efficiency. In doing so, we distinguish sharply between asymmetric information and adverse selection. Asymmetric information is the main parameter of the model and is represented in a very flexible way. Adverse selection and market efficiency are defined as equilibrium outputs of the model.

The structure of asymmetric information in the classical literature on adverse selection is restrictive in two respects. First, the valuations of market participants are all increasing functions of some underlying scalar type. The quality of the car in Akerlof (1970), the probability of accident in Rothschild and Stiglitz (1976) and the productivity of the worker in Greenwald (1986) are all unidimensional. Second, there is no public information. The informed party is assumed to know the type and the uninformed party to know nothing.

These assumptions, although simplifying, are clearly unrealistic.1 More importantly, they create difficulties for application of the theory. Finkelstein and McGarry (2006) provide convincing evidence of the inadequacy of the scalar type assumption to account for observed behavior in insurance markets. Chiappori and Salanie (2000), in response to the no public information assumption, note the importance of conditioning on all publicly available information and treating each realization as an entirely separate market. This is a problem if the econometrician is less well informed than the market participants were at the time of the transaction or if there are many dimensions of public information which are known to the econometrician and conditioning eats up many degrees of freedom.

We adopt Akerlof’s (1970) unit-demand model extended to a setting in which valuations of market participants are not all increasing functions of some underlying scalar type and there is multidimensional public and private information. A key feature of our analysis is that we define adverse selection quantitatively—it is the difference between the average quality of the population at large and the average quality of those who select to trade conditional on public information. Since adverse selection is defined conditionally on public information, it is a function of public information and therefore from an ex ante perspective is a random variable. Similarly, the expected value of the market allocation relative to the no trade status quo conditional on public information—the efficiency contribution—is also a random variable. Understanding how asymmetric information impacts on adverse selection and market efficiency therefore amounts to understanding how the joint distribution of adverse selection and efficiency contribution depends on the structure of asymmetric information.

---

1Cars differ in several dimensions as do consumer tastes, workers have a variety of skills, and those seeking insurance differ not only by probability of accident but by other factors affecting their demand for insurance such as attitude to risk. The assumption that uninformed agents know nothing is clearly an oversimplification. Second hand car prices depend on mileage and condition, insurance contracts depend on observable characteristics of the insured and workers are paid differently depending on their publicly observed employment history.
We couch most of the analysis in the language of a two-period employer learning model (Greenwald 1986, Gibbons and Katz 1991) but make reference to the insurance application at appropriate points in the paper. Sections 2 and 3 set out the game and justify the equilibrium at the heart of the analysis. In this introduction we present a special case of the model (introducing notation), give a preview of key results, and discuss related literature.

There are a finite number of firms indexed $I, J \in F$ and one worker. There are two periods, \textit{training} followed by \textit{employment}. The value of the worker to firm $I \in F$ during training is denoted by the random variable $Y_I$. Because the worker may develop skills during training, the value of the worker during (second period) employment depends on where the worker trained. The random variable $Y_{IJ} = X_I + Z_J$ denotes the value of the worker to firm $J$ during employment when the worker trained at firm $I$, where $X_I$ is general ability after training and $Z_J$ is a match-specific component assumed identically and independently distributed across firms. For simplicity, we assume $Y_I = Y_{II}$.

No firms have any information when hiring to training contracts. However, during training, firms acquire information about the value of the worker in an employment contract. An \textit{information structure} for firm $I$, indexed $\gamma_I \in \Gamma_I$, specifies a vector $(Q_{\gamma_I}, T_{\gamma_I})$ of random variables. $Q_{\gamma_I}$ is to be read as information private to the training firm, $T_{\gamma_I}$ represents information made public to all firms. Realizations of public information therefore define separate 	extit{submarkets} in the sense that equilibrium wages will generally depend on which value of $T_{\gamma_I}$ is realized. For example, private information could consist of the vector $(Y_I + \epsilon, X_I + \nu_1, X_I + \nu_2)$ and public information $X_I + \nu_1$, where $\epsilon$, $\nu_1$ and $\nu_2$ are noise terms. Here, private information consists of samples both of the worker’s productivity during training and of her general ability but public information consists only of a sample of general ability.

Firms compete to hire the worker in the first period by offering \textit{training contracts} consisting of a (non-negative) training wage and an information structure $\gamma_I \in \Gamma_I$. Then, after the realization of information $(Q_{\gamma_I}, T_{\gamma_I})$ they offer \textit{employment contracts} specifying a second period wage. We follow the literature in assuming that the training firm can make a counter-offer to the worker after other firms have simultaneously posted their offers.

In the above example, the training firm $I$ is better informed about both the value of the worker if retained and the value of the worker to outside firms. The training firm’s estimates are respectively denoted $R_{\gamma_I} = \mathbb{E}[Y_{II} | Q_{\gamma_I}] = \mathbb{E}[X_I + Z_I | Q_{\gamma_I}]$ and $G_{\gamma_I} = \mathbb{E}[Y_{IJ} | Q_{\gamma_I}] = \mathbb{E}[X_I | Q_{\gamma_I}]$, $I, J \in F, I \neq J$. The random variable $S_{\gamma_I} = \mathbb{E}[Z_I | Q_{\gamma_I}]$ is the training firm’s best estimate of the difference in values. It follows from the law of iterated expectations that the outside firms’ estimates of $Y_{II}$, $Y_{IJ}$ and $Y_{II} - Y_{IJ}$ may be written respectively as $\mathbb{E}[R_{\gamma_I} | T_{\gamma_I}]$, $\mathbb{E}[G_{\gamma_I} | T_{\gamma_I}]$ and $\mathbb{E}[S_{\gamma_I} | T_{\gamma_I}]$. When $\mathbb{E}[S_{\gamma_I} | T_{\gamma_I}]$ is negative, from the view of outside firms, the worker appears to be a bad match with the current employer. This creates an apparent ‘genuine reason for trade’. Therefore, we refer to $\mathbb{E}[S_{\gamma_I} | T_{\gamma_I}]$ as the \textit{apparent match quality}. It will play a central
role in what follows.

In Section 3 we establish the following second period wage equation as part of a perfect Bayesian equilibrium

$$w_{\gamma_1}(t) = \mathbb{E}[G_{\gamma_1} | T_{\gamma_1} = t, R_{\gamma_1} < w_{\gamma_1}(t)].$$

Trade occurs in the event that the training firm’s estimate of the value of the worker if retained is less than the equilibrium employment wage, $R_{\gamma_1} < w_{\gamma_1}(T_{\gamma_1})$. Hence the employment wage is the outside firms’ common estimate of the value of the worker in an outside employment conditional on public information and the event that trade occurs. This is essentially already familiar from Akerlof (1970), the main departure being the explicit conditioning on public information $T_{\gamma_1} = t$.

In Section 4.1 we define adverse selection $AS_{\gamma_1}$ quantitatively as a function of public information

$$AS_{\gamma_1} = as_{\gamma_1}(T_{\gamma_1}) = \mathbb{E}[G_{\gamma_1} | T_{\gamma_1}] - \mathbb{E}[G_{\gamma_1} | T_{\gamma_1}, \text{trade occurs}].$$

We seek to characterize features of the joint distribution of adverse selection with information and other equilibrium market outcomes. To this end, we make two simple observations. The first is to note that the trading event $R_{\gamma_1} < w_{\gamma_1}(T_{\gamma_1})$ can be rewritten using the wage equation as $AS_{\gamma_1} + \mathbb{E}[S_{\gamma_1} | T_{\gamma_1}] < R_{\gamma_1}$, where $R_{\gamma_1} = \mathbb{E}[R_{\gamma_1} | T_{\gamma_1}] - R_{\gamma_1}$ is the difference between the estimate of $Y_{II}$ based on public information and the estimate based on the superior private information. This makes sense, both positive adverse selection and high apparent match quality deter trade which therefore only takes place if the valuation error is sufficiently large. Using this in the definition of adverse selection establishes that adverse selection satisfies the recursive relationship $AS_{\gamma_1} = \mathbb{E} \left[ G_{\gamma_1} | T_{\gamma_1}, AS_{\gamma_1} + \mathbb{E}[S_{\gamma_1} | T_{\gamma_1}] < R_{\gamma_1} \right]$, where $G_{\gamma_1} = \mathbb{E}[G_{\gamma_1} | T_{\gamma_1}] - G_{\gamma_1}$ is the valuation error on $Y_{II}$, $J \neq I$. The second observation is that the valuation errors $R_{\gamma_1}, G_{\gamma_1}$ are, by construction, uncorrelated with $T_{\gamma_1}$, so under the stronger condition of statistical independence, orthogonality with $T_{\gamma_1}$, we may simply write

$$as_{\gamma_1}(t) = \mathbb{E} \left[ G | as_{\gamma_1}(t) + \mathbb{E}[S_{\gamma_1} | T_{\gamma_1} = t] < R_{\gamma_1} \right].$$

The consequence of these observations, noted in Section 4.2, is that, for an information structure $\gamma_I$ which satisfies orthogonality, adverse selection is an increasing function of the apparent match quality. In other words, adverse selection depends on multidimensional public information only via the scalar statistic $\mathbb{E}[S_{\gamma_1} | T_{\gamma_1}]$. Similarly, both the probability of trade and the efficiency contribution are shown to be decreasing functions of the apparent match quality.

Section 4.3 presents applications of these results. First, our representation speaks to existing approaches to testing for adverse selection in the labour market (Gibbons and Katz

---

2In Section 3.3 we show explicitly how the model and (second period) equilibrium translate to the study of insurance. In the insurance context the apparent match quality $\mathbb{E}[S_{\gamma_1} | T_{\gamma_1}]$ is the average risk premium for the submarket defined by public information. Much of the analysis in Section 4 also applies to insurance markets.
1991). The familiar intuition from models with scalar types is that workers who are retained by an employer possessing private information will be paid more than those who are released. In our model, no such general conclusion obtains. Using the results in Section 4.2, we can establish a simple sufficient condition under which released workers are paid more than those who are retained. In this event, we say that the employer is a *low wage firm*.

It is well known that adverse selection can be hard to find.\(^3\) Our representation sheds light on this. In our unit-demand model, the standard Chiappori and Salanie (2000) positive correlation test for adverse selection becomes a simple test of whether adverse selection is positive. It was stressed by Chiappori and Salanie that since theory implies different observable types should be offered different contracts, one should control for public information and hence test whether \(AS_{\gamma_I} = \mathbb{E}[G_{\gamma_I}|T_{\gamma_I}] - \mathbb{E}[G_{\gamma_I}|T_{\gamma_I}, \text{trade occurs}] > 0\). However, if there is unobserved heterogeneity, i.e. if \(T_{\gamma_I}\) is poorly observed by the econometrician, this control is difficult. A natural surrogate is to estimate the quantity of adverse selection in the aggregated market constructed by ignoring \(T_{\gamma_I}\), specifically \(\mathbb{E}[G_{\gamma_I}] - \mathbb{E}[G_{\gamma_I}|\text{trade occurs}]\). Using the results in Section 4.2, we can establish sufficient conditions for this quantity to be an overestimate or underestimate of the average quantity of adverse selection across all markets \(\mathbb{E}[AS_{\gamma_I}]\). When it is an underestimate, the employer is a low wage firm and there is a possibility of an effect reversal, where the quantity of adverse selection is positive in every submarket but the quantity of adverse selection in the aggregated market constructed by ignoring \(T_{\gamma_I}\) is negative. Clearly, in such cases, the surrogate positive correlation test is not a valid test for the presence of adverse selection.

A related point is that failure of the positive correlation test among market participants that trade does not necessarily mean there is no adverse selection in the market overall. It follows from the decreasing functional relationship between adverse selection and probability of trade that expected adverse selection conditional on trade taking place, \(\mathbb{E}[AS_{\gamma_I}|\text{trade occurs}]\), is less than expected adverse selection conditional on trade not taking place, \(\mathbb{E}[AS_{\gamma_I}|\text{no trade}]\). Hence adverse selection concentrates disproportionately where trades do not take place and depending on data availability this might impede its detection.

Our representation also speaks to policy. The canonical solution (Einav and Finkelstein 2011) to the inefficiencies created by adverse selection is to have mandatory trades. A naive view, supported by the negative relationship between adverse selection and efficiency contribution, might be that it is optimal to mandate trade in submarkets (risk classes in the insurance context) for which adverse selection is high. On the contrary, we show that it is optimal to mandate trade only in submarkets for which adverse selection is low.

Finally, our representation delivers a convenient parameterization of an information structure. In a Gaussian specification of the model, the economically relevant aspects of an information structure can be captured by a three dimensional space of parameters. These parameters

---

\(^3\)For a discussion see, e.g., Chiappori and Salanie (2000) and Einav, Finkelstein and Levin (2010).
have natural interpretations. The first is the $(Y_{II})$ information gap—a measure of the difference in the amount of information about $Y_{II}$ held by the informed and uninformed parties. The second parameter is informed party relevance—a measure of the relevance of the informed party’s estimate of its own value, $R_{\gamma_I}$, to the quantity that the uninformed parties seek to estimate, $G_{\gamma_I}$, given public information $T_{\gamma_I}$. The final parameter is a measure of the quality of public information about match quality.

Section 5 turns to the question of how overall efficiency aggregated across submarkets depends on the information structure, and how this relates to aggregate adverse selection $E[\text{AS}_{\gamma_I}]$. We use the Gaussian case, continuing the analysis at the end of Section 4.3 by exploring how expected adverse selection and efficiency contribution vary over the three dimensional parameter space. Expected adverse selection is shown to be increasing in each of these parameters. The relationship with expected efficiency is more complex. Comparing two information structures with the same information gap and quality of public information about match quality, the information structure for which informed party relevance is larger leads to less expected efficiency. On the other hand, holding the information gap and informed party relevance fixed (with the latter not too large), better public information about match quality leads to more expected efficiency. We also show that adverse selection and efficiency contribution satisfy a Spence-Mirrlees condition. Comparing two information structures with the same information gap and which generate the same quantity of expected adverse selection, the information structure with better public information about match quality leads to more expected efficiency.

It is of interest to understand which information structures impose a given amount of adverse selection with the least loss of efficiency. A primary motivation stems from the two-period employer learning application in which firms compete to hire workers in the first period with a view to capturing informational rents in the second period. Firms have incentives to capture these rents in an efficient way. In Section 6, we first consider the case where $\Gamma_I$ contains only (Gaussian) information structures with the same fixed inside information. The contribution here is the identification of the public information disclosures associated with efficient imposition of adverse selection. We establish that these efficient public information disclosures take a rather simple form—the estimates $R_{\gamma_I}$ and $G_{\gamma_I}$ are conditionally colinear given public information. Second, we explore the effect of acquiring better private information. We show that, providing there is sufficient freedom to transmit information publicly, improving private information facilitates greater efficiency in imposing adverse selection.

These results have an interesting implication that relates to our earlier discussion of empirical tests for adverse selection. If public information is disclosed efficiently, then under natural conditions the quantity of adverse selection in the aggregated market constructed by ignoring $T_{\gamma_I}$ will underestimate the average quantity of adverse selection across all markets (implying that the employer is a low wage firm). Thus, when imposed efficiently, adverse selection may be hard to find.
We complete the analysis by studying the choice of endogenous information structures. Firms make training contract offers specifying a training wage and an information structure. Since the worker is credit-constrained and unable to pay for the benefits of training up-front, choosing an information structure that imposes adverse selection efficiently becomes an effective way to claw back rent. If one firm has sufficient advantage over its competitors in training the worker and sufficient control of disclosure, it will acquire maximal information and disclose so as to efficiently impose positive adverse selection. In equilibrium, therefore, employers (wishing to claw back rents) will be low wage firms—workers who leave will generally go to higher paid jobs. This is counter to what might be expected given the classical literature starting from Gibbons and Katz (1991).

Related Literature Finkelstein and McGarry (2006) also argue that an extension of the classical literature is necessary. These authors document that there is asymmetric information in the market for long-term care insurance in the U.S., but that the standard positive correlation test of Chiappori and Salanie (2000) fails to provide evidence of adverse selection. Specifically, (1) conditional on insurance companies’ assessment of individuals’ risk types, individuals have residual private information that predicts their eventual risk, and (2) there is no positive correlation between individuals’ insurance coverage and their eventual risk, even controlling for the companies’ own assessment of the individuals’ risk types. Finkelstein and McGarry (2006) conclude that, to understand the equilibria that can arise in real-world markets, it is necessary to move beyond the standard unidimensional theory and allow for “multiple sources of correlated heterogeneity”.

Responding to this call, Einav, Finkelstein and Cullen (2010) discuss a model which is essentially identical to the one at the heart of our paper. Instead of a scalar type space, there is a more or less general probability space on which are defined functions (random variables) representing agents’ values and information. They make the important observation that conditional on public information (i.e. treating each realization of public information as a separate market) the presence of adverse selection manifests itself in a declining marginal cost curve (since those with the highest value of insurance are the costliest to supply). Welfare analysis is conducted in the usual way by integrating under the appropriate demand and cost curves. Although our model and that of Einav, Finkelstein and Cullen (2010) are similar, we take the analysis in a different but complimentary direction. A central objective of our paper is to understand how overall efficiency aggregated across submarkets depends on the information structure, and how this relates to aggregate adverse selection.4

4We explore this issue by adding efficiency losses across submarkets. This is natural in the employer learning framework, given the standard risk neutrality assumption. For insurance markets, ex ante welfare evaluation is significantly more complicated if it is deemed socially desirable to insure individuals against the underwriting risk of high insurance premiums. We do not address this welfare effect of information (Hirshleifer 1971), although this would be an interesting avenue for future research.
Levin (2001) also addresses the relationship between different types of asymmetric information, adverse selection and the probability of trade (as a surrogate for efficiency). In contrast to our paper, his model retains the classical scalar type assumption which is one of our main points of departure.

2 The Model

Here we present a game between firms which compete to employ workers. Firms compete, as usual, through wages, but also by choosing how information about worker productivity (and how well the worker is matched to the firm) flows onto the market. Readers primarily interested in exogenous information structures and in the insurance interpretation of the model can skim this section and focus on the material in Section 3.3 to Section 5.

2.1 Description

The economy consists of a finite number of firms indexed $I, J \in F = \{1, ..., n_F\}$, one worker, and a probability space on which is defined a collection of real valued random variables. There are two periods. We refer to first period employment as training and second period employment simply as employment.

Values. The value (productivity) of the worker to firm $I \in F$ during training is denoted by the random variable $Y_I$. The value of the worker during employment depends on where the worker trained. The random variable $Y_{IJ}$ denotes the value of the worker to firm $J$ during employment when the worker trained at firm $I$.

Firms all have the same information, normalized to null information, when hiring to training contracts. During training, firms acquire information according to an information structure which specifies what each firm knows at the time of hiring to employment contracts.

Information Structure. An information structure for firm $I$, indexed $\gamma_I \in \Gamma_I$, specifies an $n_{Q_{\gamma_I}} + n_{T_{\gamma_I}}$ dimensional vector $(Q_{\gamma_I}, T_{\gamma_I})$ of real valued random variables. $Q_{\gamma_I}$ is to be read as information private to the training firm, $T_{\gamma_I}$ represents information made public to all firms.

The collection of possible information structures $(\Gamma_I)_{I \in F}$ is a key parameter of the model and understanding how equilibrium outcomes depend on $(\Gamma_I)_{I \in F}$ is a central focus of the paper. We call the special case in which each $\Gamma_I, I \in F$ is a singleton, exogenous information.

Contracts. Firms compete to hire or retain the worker in the first and second periods by offering respectively training contracts and employment contracts:
1. A training contract \( \tau_I = (w^{1st}_I, \gamma_I) \) at firm \( I \in F \) specifies:
   (a) A first period (training) wage \( w^{1st}_I \in \mathbb{R}_+ \).
   (b) An information structure \( \gamma_I \in \Gamma_I \).

2. An employment contract at firm \( I \in F \) specifies a second period wage \( w^{2nd}_I \in \mathbb{R} \).

The existence of a lower bound (taken to be zero) on the training wage plays an important role in Section 6. Specifically, it is a firm’s desire to claw back future rents from a credit constrained worker which motivates its choice of information structure. If unboundedly negative training wages were possible, then familiar arguments, due to Becker (1962), imply that information structures would be chosen purely on grounds of efficiency. For simplicity, it is assumed there is no bound on second period wages. Our analysis focuses mainly on second period wages and to ease notation we write \( w^{2nd}_I = w_I \).

**Timing.** The timing of the game is as follows:

**First Period** Each firm \( J \in F \) simultaneously offers a training contract \( \tau_J = (w^{1st}_J, \gamma_J) \in \mathbb{R}_+ \times \Gamma_J \). The worker chooses a training contract from some firm \( J \) and training takes place. The worker is paid the training wage \( w^{1st}_I \) as specified in \( \tau_I \), a realization is drawn from the random pair \( (Q_{\gamma_I}, T_{\gamma_I}) \) specified in \( \gamma_I \), firm \( I \) privately observes the realization of \( Q_{\gamma_I} \) and the outside firms all observe the realization of \( T_{\gamma_I} \).

**Second Period** All outside firms \( J \in F, J \neq I \) simultaneously post employment wage contract offers \( w_J \in \mathbb{R} \) to the worker. Firm \( I \) observes the outside offers and then makes an employment wage counter offer\(^5\) \( w_I \in \mathbb{R} \). The worker chooses which employment offer to accept. Production takes place, employment wages are paid according to contracts and payoffs are realized. Worker payoffs are the undiscounted sum of wages received, firm payoffs are undiscounted profits, i.e. productivity less wages.

A behavioral strategy for the worker is a pair \( a = (a_1, a_2) \) in which \( a_1 \) determines which training contract to accept and \( a_2 \) determines which employment contract to accept. \( a_1 \) maps from the set of possible training contract offers \( \prod_{I \in F} (\mathbb{R}_+ \times \Gamma_I) \) to a training contract choice. \( a_2 \) maps from the set of possible histories of the game to an employment contract choice. Histories consist of a collection of first period contract offers, the identity of the firm whose contract was accepted, a realization of public information, and a collection of wage contract offers. A behavioral strategy for the firm is a triple consisting of: a first period contract, a wage offer

---

\(^5\) The second period competition for a worker effectively follows the procedure set out by Greenwald (1986) which has been adopted by much of the subsequent literature, e.g. Gibbons and Katz (1991) and Acemoglu and Pischke (1998). Variants on this wage setting protocol include Pinkston (2009) who studies ascending ‘button’ auctions and Li (2011) who studies first price auctions.
A Perfect Bayesian Equilibrium (PBE) of the game requires that the worker, and each firm, follows a sequentially optimal strategy for some Bayes consistent belief. In particular, in the second period: for the worker, \(a_2\) selects one of the highest wage employment contracts offered; for the training firm \(I\), its employment wage contract offer maximizes expected second period profit given the strategy of the worker and wage offers of the other firms \(J \neq I\); for each non-training firm \(J \neq I\), its wage offer at each realization \(t\) of \(T_{\gamma_I}\) maximizes its expected profit given its beliefs about the wage offers by other outside firms and the strategy of the training firm \(I\). In the first period: the worker’s contract choice maximizes her lifetime expected wages given her beliefs; and each firm’s contract offer maximizes its expected profit.

### 2.2 Statistical Assumptions and Definitions

As in the classical literature on adverse selection, symmetry assumptions simplify the analysis of competition. We make the following assumptions on the joint distribution of values and information.

**A1.** Interim exchangeability of outside firms: For any three distinct firms \(I, J, J' \in F\), \(\gamma_I \in \Gamma_I\), the random pairs \((Q_{\gamma_I}, Y_{IJ})\) and \((Q_{\gamma_I}, Y_{IJ'})\) have identical distributions.

**A2.** Existence and equality of expected employment value: For each \(I, J \in F\), the expectations \(\mathbb{E}[Y_J], \mathbb{E}[Y_{IJ}]\) exist. Furthermore, \(\mathbb{E}[Y_{IJ}] = \mathbb{E}[Y_{II}]\).

**A3.** Asymmetric Information: \(T_{\gamma_I}\) is less informative about employment value than \(Q_{\gamma_I}\). Specifically, for each \(I, J \in F\), \(\gamma_I \in \Gamma_I\), \(Y_{IJ}\) is conditionally independent of \(T_{\gamma_I}\) given \(Q_{\gamma_I}\), i.e. \(Y_{IJ} \perp \perp T_{\gamma_I} | Q_{\gamma_I}\).

Note, regarding Assumption A1, that once training takes place at firm \(I\) and \(Q_{\gamma_I}\) is realized, the conditional distributions of \(Y_{II}\) and \(Y_{IJ}\) given \(Q_{\gamma_I}\) are not assumed to be identical. Generally for each distinct \(I, J \in F\), and \(\gamma_I \in \Gamma_I\), \((Q_{\gamma_I}, Y_{II})\) is not equal in distribution to \((Q_{\gamma_I}, Y_{IJ})\). This reflects the fact that \(Q_{\gamma_I}\) generally contains information about how well matched the worker and training firm are compared to a randomly selected outside firm. Note that in Assumption A2 we do not assume \(\mathbb{E}[Y_{IJ}] = \mathbb{E}[Y_{IJ'}]\) for all \(I, I', J \in F\). As we discuss in Section 6, some firms may bestow especially valuable skills on the workers they choose to train. The assumption \(\mathbb{E}[Y_{IJ}] = \mathbb{E}[Y_{II}]\) can easily be relaxed. Finally, the fact that the information accruing to the training firm \(Q_{\gamma_I}\) contains \(T_{\gamma_I}\), in the sense of A3, is effectively a definition of \(T_{\gamma_I}\) being public information, and therefore also available to firm \(I\).
Notation 1 Given $A_1$ and $A_2$, for each $\gamma_I \in \Gamma_I$ we can uniquely define:\footnote{The conditional expectations exist by the first statement of A2. They are equal a.e. for all $J \in F, J \neq I$ by A1.}

\[ G_{\gamma_I} \overset{\text{def}}{=} \mathbb{E}[Y_{IJK}|Q_{\gamma_I}], \quad R_{\gamma_I} \overset{\text{def}}{=} \mathbb{E}[Y_{II}|Q_{\gamma_I}], \quad S_{\gamma_I} = R_{\gamma_I} - G_{\gamma_I}, J \neq I \in F. \]

The supports of $Q_{\gamma_I}, G_{\gamma_I}, R_{\gamma_I}$ and $T_{\gamma_I}$ are denoted respectively by $Q_{\gamma_I}, G_{\gamma_I}, R_{\gamma_I}, T_{\gamma_I}$. The support of the conditional distribution of $R_{\gamma_I}$ given the event $T_{\gamma_I} = t$ is $R_{\gamma_I}(t) = \text{supp}[R_{\gamma_I}|T_{\gamma_I} = t]$, similarly $G_{\gamma_I}(t) = \text{supp}[G_{\gamma_I}|T_{\gamma_I} = t]$.

The random variable $G_{\gamma_I}$ is the training firm’s estimate (the best estimate given all available information) of the value of the worker during employment in an outside firm. Similarly, the random variable $R_{\gamma_I}$ is the best estimate of the value of the worker in employment at the current, training firm. A feature of this paper which distinguishes it from much of the existing literature (e.g. Akerlof 1970, Levin 2001) is that we do not assume that $Y_{II}$ and $Y_{IJK}$ are increasing functions of some underlying scalar type (making them comonotone random variables). Neither do we assume that $G_{\gamma_I}$ and $R_{\gamma_I}$ are comonotone.\footnote{In general, $G_{\gamma_I}, R_{\gamma_I}$ may fail to be comonotone even if $Y_{IJK}, Y_{II}$ are comonotone. Also, $G_{\gamma_I}, R_{\gamma_I}$ may be comonotone when $Y_{IJK}, Y_{II}$ are not comonotone. The former case may arise due to curvature in the relationships, the latter naturally occurs if $Q_{\gamma_I}$ is scalar.}

Finally, the random variable $S_{\gamma_I}$ is the best estimate of the difference in the value of the worker in employment in the training firm versus an outside firm. In what follows, we refer to the expectation of $S_{\gamma_I}$ given public information, $\mathbb{E}[S_{\gamma_I}|T_{\gamma_I}]$, as the \textit{apparent match quality}.\footnote{Alternatively, one could view $\mathbb{E}[\cdot|S_{\gamma_I}]$ as the \textit{apparent grounds for trade}.}

A4. The conditional distribution of $R_{\gamma_I}$ given $T_{\gamma_I}$ is logconcave. Specifically, for each $t \in T_{\gamma_I}$,

\[ r \mapsto \log(\text{Pr}[R_{\gamma_I} < r|T_{\gamma_I} = t]) \text{ is an extended real valued concave function on } \mathbb{R}. \]

2.3 The Gaussian Case

**Definition 1** Information structures are Gaussian if for each $I \in F$ and $\gamma_I \in \Gamma_I$, there are integers $n_{Q_{\gamma_I}} \geq 0, n_{T_{\gamma_I}} \geq 0$ such that $U_{\gamma_I} = (Y_{I1},...,Y_{In_F},Q_{\gamma_I},T_{\gamma_I})$ has an $n_{\gamma_I} = n_F + n_{Q_{\gamma_I}} + n_{T_{\gamma_I}}$ dimensional Gaussian distribution.

Note that when information structures are Gaussian, they will generally be \textit{singular} Gaussian distributions for which a density does not exist. The distribution can, however, be represented by its characteristic function $\phi_{\gamma_I}(u) = \exp \left( u^T \mu_{\gamma_I} i - \frac{1}{2} u^T \Sigma_{\gamma_I} u \right)$, where $\Sigma_{\gamma_I}$ is a symmetric positive semidefinite matrix. Without loss of generality, we may normalize means of $Q_{\gamma_I}$ and $T_{\gamma_I}$ to be zero. Hence, if information structures are Gaussian, $\Gamma_I$ is identified with a set of symmetric positive semidefinite covariance matrices.
It will also become convenient, both for interpretation and reduction of dimensionality, to refer to Gaussian distributions in terms of regression coefficients.

Notation 2 Following standard notation, we denote the linear regression coefficient of \( G_{\gamma_I} \) on \( R_{\gamma_I} \) adjusting for \( T_{\gamma_I} \) by \( \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} \) and the total regression coefficient obtained by marginalizing over \( T_{\gamma_I} \) by \( \beta_{G_{\gamma_I}R_{\gamma_I}} \). Cochran’s (1938) identity is

\[
\beta_{G_{\gamma_I}R_{\gamma_I}} = \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} + \beta_{G_{\gamma_I}T_{\gamma_I}R_{\gamma_I}} \beta_{R_{\gamma_I}T_{\gamma_I}}.
\]

We write \( \sigma_{G_{\gamma_I}R_{\gamma_I}} \) for the covariance of \( G_{\gamma_I} \) and \( R_{\gamma_I} \), \( \sigma_{G_{\gamma_I}}^2 \) for the variance of \( G_{\gamma_I} \), \( \sigma_{R_{\gamma_I}|T_{\gamma_I}}^2 \) for the conditional variance of \( R_{\gamma_I} \) given \( T_{\gamma_I} \) and \( \sigma_{G_{\gamma_I}R_{\gamma_I}|T_{\gamma_I}} \) for the conditional covariance of \( G_{\gamma_I} \) and \( R_{\gamma_I} \) given \( T_{\gamma_I} \). If \( R_{\gamma_I} \) is contained in \( T_{\gamma_I} \), then we will set \( \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} = 0 \).

3 Equilibrium

A notable feature of the literature on adverse selection in labor markets is the introduction of more or less ad hoc devices for the purpose of ameliorating adverse selection. This is necessary in order to generate equilibria which (a) exist and (b) are interior to the support of the productivity distribution. Greenwald (1986), for instance, assumes that there is an exogenous separation probability independent of worker productivity. In our model, adverse selection is ameliorated endogenously by regression to the mean. This is very natural, a worker who performs below average in the training firm is likely to perform better elsewhere and so potential hirers are prepared to pay more than the minimum possible productivity level.

The crucial implication of regression to the mean is that the willingness to pay map, defined below, becomes a contraction. This fact not only allows us to settle existence and uniqueness questions, but it also supplies a means of verifying properties of adverse selection. We provide a definition of the required notion of regression to the mean in Section 3.1. In Section 3.2, we use this result to establish equilibrium market outcomes in the exogenous information version of our model. We present the insurance application in Section 3.3.

3.1 Regression to the Mean and Willingness to Pay

Assumption \( A_3 \) implies that outside firms in estimating their value for the worker need only do so second hand, via estimating \( G_{\gamma_I} \). In particular, the law of iterated expectations together with assumption \( A_1 \) implies

\[
E[Y_{I,J}|T_{\gamma_I}, R_{\gamma_I}] = E[G_{\gamma_I}|T_{\gamma_I}, R_{\gamma_I}], \quad I, J \in F, \ I \neq J,
\]

and consequently for each \( t \in T_{\gamma_I}, w \in R_{\gamma_I}(t) \),

\[
E[Y_{I,J}|T_{\gamma_I} = t, R_{\gamma_I} < w] = E[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} < w], \quad I, J \in F, \ I \neq J.
\]
It is natural to use (1) to define regression to the mean.

**Definition 2 (Regression to the mean)** We say $\gamma_I \in \Gamma_I$ induces regression to the mean if, for each $t \in T_{\gamma_I}$, $r \mapsto \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} = r]$ is increasing on $R_{\gamma_I}(t)$ and $r \mapsto r - \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} = r]$ is strictly increasing on $R_{\gamma_I}(t)$.

If information structures are Gaussian, $\gamma_I \in \Gamma_I$ induces regression to the mean if and only if $0 \leq \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} < 1$.

**Remark 1** It follows from Assumption A4 that regression to the mean implies $w \mapsto \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} < w]$ is increasing and $w \mapsto w - \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} < w]$ is strictly increasing.

**Proof.** Appendix A. ■

The quantity in equation (2) represents the expected value of the worker to outside firms in the event that the training firm is unwilling to retain the worker at wage $w$. It therefore represents outside firms’ willingness to pay for a released worker.

**Definition 3 (The willingness to pay map)** For $\gamma_I \in \Gamma_I$, let $C_{\gamma_I}$ be the set of real valued functions $\phi : T_{\gamma_I} \to \text{Conv}(G_{\gamma_I})$. We define the willingness to pay map $\Psi_{\gamma_I} : C_{\gamma_I} \to C_{\gamma_I}$, by

$$
\Psi_{\gamma_I}(\phi)(t) = \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} < \phi(t)], \text{ for } \phi(t) \geq \inf R_{\gamma_I}(t)
$$

$$
= \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} = \inf R_{\gamma_I}(t)], \text{ for } \phi(t) < \inf R_{\gamma_I}(t).
$$

We will make use of the fact (established as Lemma A1 in Appendix A) that under either of the conditions that information structures are Gaussian, or $G_{\gamma_I}$ is compact for each $I \in F$, that if $\gamma_I \in \Gamma_{\gamma_I}$ induces regression to the mean, then $\Psi_{\gamma_I}$ is **contractive**. Consequently, under these conditions, the willingness to pay map has a unique fixed point.

### 3.2 Characterization of Equilibrium Market Outcomes

Our objective in this section is to establish the analog of the familiar (Akerlof 1970) market outcomes as equilibria, rather than to provide an exhaustive characterization of all the strategies that constitute PBE.\(^{11}\) We begin by calling the unique fixed point of $\Psi_{\gamma_I}$ (in cases where it exists), the wage schedule.

**Definition 4** If the willingness to pay map $\Psi_{\gamma_I}$ has a unique fixed point, we call the fixed point the wage schedule $w_{\gamma_I} : T_{\gamma_I} \to R$ corresponding to $\gamma_I$. The random variable $W_{\gamma_I} = w_{\gamma_I}(T_{\gamma_I})$ is called the employment wage. The **expected total surplus** corresponding to $\gamma_I$ is defined as

$$
\mathbb{E}[TS_{\gamma_I}] = \mathbb{E}[Y_I] + \mathbb{E}[R_{\gamma_I}] - \mathbb{E}[S_{\gamma_I}1_{(R_{\gamma_I} < W_{\gamma_I})}].
$$

\(^{10}\)Throughout, we will call a function $f$ increasing if $x \geq y$ implies $f(x) \geq f(y)$. If the implication holds when the inequalities are replaced by strict ones, we call the function strictly increasing.

\(^{11}\)We briefly discuss equilibrium selection issues in Appendix A.
**Proposition 1** Assume A1-A4. Let information be exogenous, i.e. \( \Gamma_J = \{ \gamma_J \} \), for each \( J \in F \). Suppose either information structures are Gaussian, or for each \( J \in F \), \( G_{\gamma_J} \) is compact. Suppose that \( \gamma_J \) induces regression to the mean. It follows that, for each \( J \in F \), the willingness to pay map has a unique fixed point \( w_{\gamma_J} : T_{\gamma_J} \rightarrow \mathbb{R} \). Choose \( I \in F \) such that \( \mathbb{E}[TS_{\gamma_J}] \geq \mathbb{E}[TS_{\gamma_J}], J \in F \). There exists a PBE in which:

1. In period 1, each firm \( J \in F, J \neq I \) offers a training contract \( \tau_J = (\mathbb{E}[TS_{\gamma_J}] - \mathbb{E}[W_{\gamma_J}], \gamma_J ) \).
   Firm \( I \) offers a training contract \( \tau_I = \left( \max_{J \in F, J \neq I} (\mathbb{E}[TS_{\gamma_J}] - \mathbb{E}[W_{\gamma_J}])^+, \gamma_I \right) \). The worker chooses to train at firm \( I \).

2. In period 2, each outside firm \( J \in F, J \neq I \) makes the employment wage offer \( W_{\gamma_J} = w_{\gamma_J}(T_{\gamma_J}) \). The training firm \( I \) offers the same wage as outside firms if \( R_{\gamma_I} \geq W_{\gamma_I} \) but a lower wage if \( R_{\gamma_I} < W_{\gamma_I} \). The worker remains at the training firm \( I \) if its offer matches \( W_{\gamma_I} \) and otherwise moves to an outside firm chosen at random.

**Proof.** Appendix A.

In the PBE referred to in Proposition 1, equilibrium market outcomes are determined by a wage schedule that is the unique fixed point of the willingness to pay map. Each unsuccessful first period bidder, \( J \in F \), offers to pay the total surplus which would accrue were they successful, less the worker’s expected wage in the second period. The successful first period bidder, \( I \in F \), offers just as much as is needed to attract the worker subject to negative wages not being allowed.

It is an immediate consequence of Definition 2 that in equilibria identified by Proposition 1, if there is a positive probability of trade (i.e. if \( \Pr[R_{\gamma_I} < w_{\gamma_I}(t)|T_{\gamma_I} = t] > 0 \)), then the following wage equation (as displayed in the Introduction) holds

**Wage Equation**

\[
    w_{\gamma_I}(t) = \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} < w_{\gamma_I}(t)]. \quad (3)
\]

### 3.3 Insurance Interpretation

Individual \( I \), with von Neumann Morgenstern utility \( u \), initial wealth \( a \) and facing a random loss \( L_I \), will buy insurance for the loss at premium \( \pi \) if

\[
    \mathbb{E}[u(a - L_I)|Q_{\gamma_I}] \leq u(a - \pi) \Leftrightarrow R_{\gamma_I} \leq w,
\]

where \( w = -\pi \) and \( R_{\gamma_I} = -\left( \mathbb{E}[L_I|Q_{\gamma_I}] + \rho_I(Q_{\gamma_I}) \right) \), with \( \rho_I(Q_{\gamma_I}) = a - \mathbb{E}[L_I|Q_{\gamma_I}] - u^{-1}(\mathbb{E}[u(a - L_I)|Q_{\gamma_I}]) \) the Arrow-Pratt risk premium for the risky gamble \( a - L_I \) conditional on the information \( Q_{\gamma_I} \). An insurance firm’s expectation of the loss it will incur from bearing the risk is \( G_{\gamma_I} = -(c + \mathbb{E}[L_I|Q_{\gamma_I}]) \) where \( c \) is the administrative cost of selling the policy (the load). In the event that the individual is willing to purchase insurance at premium
the expected cost to the insurer of supplying the policy is \( E[G_{\gamma_I} | T_{\gamma_I} = t, R_{\gamma_I} < w] = \mathbb{E}[-L_I - c | T_{\gamma_I} = t, -\mathbb{E}[L_I | Q_{\gamma_I}] - \rho_I(Q_{\gamma_I}) < -\pi] \). Competition to sell insurance means that the wage (premium) equation (3) holds. Note that, in this case, \( S_{\gamma_I} = c - \rho_I(Q_{\gamma_I}) \), is the load less the risk premium.

4 Adverse Selection and Efficiency: Representation and Applications

We now turn our attention to defining and characterizing adverse selection as determined by the wage equation (3), exploring its relationship with the probability that trade occurs and with the efficiency of the economy in allocating the worker.

4.1 Definitions

We define the quantity of adverse selection as the difference between the average quality of the population at large and that part of the population which selects into a contract. This is consistent with the early usage in the insurance literature discussed in Akerlof (1970).

Definition 5 (Quantity of adverse selection) The quantity of adverse selection under information structure \( \gamma_I \in \Gamma_I \) is defined as the random variable

\[
AS_{\gamma_I} = \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}] - \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}, R_{\gamma_I} < W_{\gamma_I}].
\]  

The adverse selection schedule under information structure \( \gamma_I \) is defined as the map \( as_{\gamma_I} : T_{\gamma_I} \rightarrow \mathbb{R} \) with \( as_{\gamma_I}(t) = \mathbb{E}[G_{\gamma_I} | T_{\gamma_I} = t] - w_{\gamma_I}(t) \).

A unique adverse selection schedule exists under the conditions of Lemma A1. Since the adverse selection schedule differs from (minus) the employment wage schedule only by the given function \( t \mapsto \mathbb{E}[G_{\gamma_I} | T_{\gamma_I} = t] \), it follows from Lemma A1 that the adverse selection schedule may also be defined in terms of the fixed point of a contraction mapping. This has the evident consequence that any property preserved under the contraction mapping is a property which will be displayed by adverse selection. Much use is made of this insight later in the paper.

Although adverse selection is defined as the difference in average (i.e. expected) value between two populations, the competitive nature of equilibria and the fact that productivity and wages are measured in the same units means that the quantity of adverse selection can also be viewed as a difference in prices. We may view adverse selection as the difference between the equilibrium wage which would obtain, through competition, in a counterfactual game without private information (i.e. \( \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}] \)) and the equilibrium wage in the game itself (\( W_{\gamma_I} \)). Viewed in this way, adverse selection is a price effect.
A familiar thought, certainly since Akerlof (1970), is that adverse selection depresses prices and discourages trade, even when trade is warranted on efficiency grounds. To explore the equilibrium relationship between adverse selection, as quantified in Definition 5, the likelihood of trade and market efficiency, we introduce two further definitions.

**Definition 6 (Probability of trade)** The *probability of trade* under information structure \( \gamma_I \in \Gamma_I \) is defined as the random variable

\[
P_{\gamma_I} = \mathbb{E}[1\{R_{\gamma_I} < W_{\gamma_I}\}|T_{\gamma_I}] = \text{Pr}[R_{\gamma_I} < W_{\gamma_I}|T_{\gamma_I}]. \tag{5}
\]

The *probability of trade schedule* under information structure \( \gamma_I \) is defined as the map \( p_{\gamma_I} : T_{\gamma_I} \rightarrow \mathbb{R}_+ \) with \( p_{\gamma_I}(t) = \text{Pr}[R_{\gamma_I} < w_{\gamma_I}(t)|T_{\gamma_I} = t] \).

In contrast to Akerlof (1970) trade is not always warranted on efficiency grounds, so increased probability of trade is not a synonym for increased efficiency.

**Definition 7 (Efficiency contribution)** The *efficiency contribution* (relative to the no trade status quo) under information structure \( \gamma_I \in \Gamma_I \) is defined as the random variable

\[
EC_{\gamma_I} = -\mathbb{E}[S_{\gamma_I}1\{R_{\gamma_I} < W_{\gamma_I}\}|T_{\gamma_I}], \tag{6}
\]

The *efficiency contribution schedule* under information structure \( \gamma_I \) is defined as the map \( ec_{\gamma_I} : T_{\gamma_I} \rightarrow \mathbb{R} \) with \( ec_{\gamma_I}(t) = -\mathbb{E}[S_{\gamma_I}1\{R_{\gamma_I} < w_{\gamma_I}(t)\}|T_{\gamma_I} = t] \).

Recall \( \mathbb{E}[TS_{\gamma_I}] = \mathbb{E}[Y_I] + \mathbb{E}[R_{\gamma_I}] - \mathbb{E}[S_{\gamma_I}1\{R_{\gamma_I} < W_{\gamma_I}\}] \). Since \( \mathbb{E}[R_{\gamma_I}] = \mathbb{E}[Y_I] \) and \( \mathbb{E}[Y_I] \) are independent of \( \gamma_I \), \( \mathbb{E}[TS_{\gamma_I}] \) depends on the information structure only through its impact on \( EC_{\gamma_I} \). Note that \( EC_{\gamma_I} \) is always positive, this is established in the following remark, which is an easy consequence of the wage equation.

**Remark 2**

\[
EC_{\gamma_I} = \mathbb{E}[(W_{\gamma_I} - R_{\gamma_I})^+ |T_{\gamma_I}] \geq 0. \tag{7}
\]

We explore the relationship between ex ante (unconditional) expectations of \( AS_{\gamma_I} \) and \( EC_{\gamma_I} \) as a function of \( \gamma_I \in \Gamma_I \) in Sections 5 and 6 below. Here, we are interested in characterizations of the random variables \( AS_{\gamma_I}, P_{\gamma_I}, EC_{\gamma_I} \) and the relationship between them for fixed \( \gamma_I \in \Gamma_I \). A key simplification follows from identifying a natural orthogonality condition to which we now turn.

**4.2 Representation Results**

Given Assumptions A1-A3, the difference between the estimate of the value of the worker in employment in an outside firm based on public information and the estimate based on the
superior information of the training firm is $\mathbb{E}[Y_{ij}|T_{\gamma_j}] - \mathbb{E}[Y_{ij}|Q_{\gamma_j}] = \mathbb{E}[G_{\gamma_j}|T_{\gamma_j}] - G_{\gamma_j}$. Since inside information $Q_{\gamma_j}$ contains public information $T_{\gamma_j}$, $\mathbb{E}[G_{\gamma_j}|T_{\gamma_j}] - G_{\gamma_j}$ represents the error in the public estimate of the value of the worker in employment in an outside firm. Similarly, $\mathbb{E}[R_{\gamma_j}|T_{\gamma_j}] - R_{\gamma_j}$ is the corresponding error in the public estimate of the value of the worker in employment in the training firm.

**Notation 3** Errors in public estimates, compared to private estimates.

$$G_{\gamma_j} \triangleq \mathbb{E}[G_{\gamma_j}|T_{\gamma_j}] - G_{\gamma_j}, \quad R_{\gamma_j} \triangleq \mathbb{E}[R_{\gamma_j}|T_{\gamma_j}] - R_{\gamma_j}, \quad I \in F.$$  

The first thing to note is that these errors are by construction uncorrelated with $T_{\gamma_j}$. Since $AS_{\gamma_j}$ is a function of $T_{\gamma_j}$, it follows that these differences are also uncorrelated with $AS_{\gamma_j}$. For Gaussian information structures, in which lack of correlation implies statistical independence, we therefore have the following **orthogonality condition**.

**A5.**

$$(G_{\gamma_j}, R_{\gamma_j}) \perp T_{\gamma_j}. \quad (8)$$

Outside the Gaussian case, condition (8) becomes an assumption which will be key to some of the results which follow. A second point to note is that, in the Gaussian case, $G_{\gamma_j}$ and $R_{\gamma_j}$ have covariance equal to $\beta_{G,\gamma_j}R_{\gamma_j}$. So given regression to the mean they are positively correlated. In a not necessarily Gaussian case, in which (8) holds, Lemma B1 in Appendix B establishes that regression to the mean implies $G_{\gamma_j}$ and $R_{\gamma_j}$ are positively dependent\(^{12}\) random variables.

Using Definition 5, we can write the employment wage as $W_{\gamma_j} = \mathbb{E}[R_{\gamma_j}|T_{\gamma_j}] - \mathbb{E}[S_{\gamma_j}|T_{\gamma_j}] - AS_{\gamma_j}$. Hence, the trading event $R_{\gamma_j} < W_{\gamma_j}$ can be re-expressed as $AS_{\gamma_j} + \mathbb{E}[S_{\gamma_j}|T_{\gamma_j}] < R_{\gamma_j}$. That is, trade takes place when the error in the public estimate of the value of the worker in employment in the training firm exceeds the sum of adverse selection and the apparent match quality. Under the above orthogonality condition, therefore, the trading event is an inequality between two independent random variables. A number of important consequences follow.

**Proposition 2** In the equilibrium identified in Proposition 1, suppose the orthogonality condition (8) holds, then adverse selection, $AS_{\gamma_j}$, and the apparent match quality, $\mathbb{E}[S_{\gamma_j}|T_{\gamma_j}]$, are comonotone random variables. Specifically, there exists an increasing function $\hat{a}_{\gamma_j} : \mathbb{R} \to \mathbb{R}_+$ such that $AS_{\gamma_j} = a_{\gamma_j}(T_{\gamma_j}) = \hat{a}_{\gamma_j}(\mathbb{E}[S_{\gamma_j}|T_{\gamma_j}])$.

\(^{12}\)Specifically, Lemma B1 establishes positive dependence in a sense somewhat weaker than Lehmann’s (1966) concept of **positively quadrant dependent (PQD)**. Kowaleczcyk and Pleszcynska (1977) called the concept we use here **EQD\(^{+}\)**. Kowaleczcyk and Pleszcynska’s notion is itself rather closely related to our definition of adverse selection: $(X, Y)$ are said to be **EQD\(^{+}\)** if for all $y$ in the support of $Y$, $E[X|Y \geq y] \geq E[X]$. See Appendix B for further details.
Proof. Since $\gamma_I$ induces regression to the mean, by Lemma A1 and Definition 5, a unique adverse selection schedule exists, allowing us to write (4) recursively as

$$AS_{\gamma_I} = \mathbb{E}\left[\mathcal{G}_{\gamma_I}\mid T_{\gamma_I}, AS_{\gamma_I} + \mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}] < R_{\gamma_I}\right].$$

Using ($\mathcal{G}_{\gamma_I}, R_{\gamma_I}$) \perp T_{\gamma_I}, we may therefore write $a_{\gamma_I}(t) = \mathbb{E}[\mathcal{G}_{\gamma_I}\mid a_{\gamma_I}(t) + \mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I} = t] < R_{\gamma_I}]$. Hence, $a_{\gamma_I}(t)$ depends on $t$ only through $\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I} = t]$. This establishes the existence of $\tilde{a}_{\gamma_I} : \mathbb{R} \rightarrow \mathbb{R}$ such that $AS_{\gamma_I} = a_{\gamma_I}(T_{\gamma_I}) = \tilde{a}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}])$. Positivity and monotonicity follow from the positive dependence of $\mathcal{G}_{\gamma_I}$ and $R_{\gamma_I}$, details are in Appendix B. ■

Proposition 2 establishes two important properties of adverse selection. First, given regression to the mean, $AS_{\gamma_I}$ is a nonnegative random variable. This is intuitive: regression to the mean implies that the event that the worker is released is bad news for the value of the worker in employment in an outside firm. Second, $AS_{\gamma_I}$ depends on information $(Q_{\gamma_I}, T_{\gamma_I})$ only through the scalar $\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]$ and the relationship is a non-decreasing one. This is also intuitive: there is a genuine reason for trade if $\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]$ is small (negative) but not if it is large (positive). Note that in the insurance context, $-\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]$ is equal to the average risk premium for the risk class defined by $T_{\gamma_I}$ less the load $c$. Hence, adverse selection differs across risk classes according to their average risk premia.

Remark 3 Some of our results generalize to cases in which the regression to the mean assumption is relaxed. Specifically, if, for each $t \in T_{\gamma_I}$, $r \mapsto \mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I} = t, R_{\gamma_I} = r]$ is strictly increasing on $R_{\gamma_I}(t)$, the wage equation and adverse selection continue to be defined. However, adverse selection can now be negative.

Using the properties of $AS_{\gamma_I}$ established in Proposition 2, it becomes straightforward to provide a representation of the probability of trade $P_{\gamma_I}$ and the efficiency contribution $EC_{\gamma_I}$.

Proposition 3 Under the conditions of Proposition 2, the probability of trade $P_{\gamma_I}$ and the efficiency contribution $EC_{\gamma_I}$ depend on information $(Q_{\gamma_I}, T_{\gamma_I})$ only through the apparent match quality $\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]$. Specifically, there exist $\tilde{p}_{\gamma_I} : \mathbb{R} \rightarrow \mathbb{R}_+$ and $\tilde{c}_{\gamma_I} : \mathbb{R} \rightarrow \mathbb{R}_+$ such that

$$P_{\gamma_I} = p_{\gamma_I}(T_{\gamma_I}) = \tilde{p}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]),$$

$$EC_{\gamma_I} = ec_{\gamma_I}(T_{\gamma_I}) = \tilde{c}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]).$$

Moreover, $a \mapsto \tilde{p}_{\gamma_I}(a)$ and $a \mapsto \tilde{c}_{\gamma_I}(a)$ are decreasing. Hence, $P_{\gamma_I}, EC_{\gamma_I}, -AS_{\gamma_I}$ and $-\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}]$ are all comonotone.

Proof. We can write (5) as $p_{\gamma_I}(t) = \Pr\left[AS_{\gamma_I} + \mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}] < R_{\gamma_I}\mid T_{\gamma_I} = t\right]$. The orthogonality condition $R_{\gamma_I} \perp T_{\gamma_I}$ implies $p_{\gamma_I}(t) = \Pr\left[a_{\gamma_I}(t) + \mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I} = t] < R_{\gamma_I}\mid T_{\gamma_I} = t\right]$. Noting the obvious fact that this probability is decreasing in $a_{\gamma_I}(t) + \mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I} = t]$, Proposition 2 implies $P_{\gamma_I} = p_{\gamma_I}(T_{\gamma_I}) = \tilde{p}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}\mid T_{\gamma_I}])$ for some nonnegative decreasing function $\tilde{p}_{\gamma_I}$.

---

13 The possibility of asymmetric information leading to advantageous selection was pointed out by (de Meza and Webb, 2001). From our perspective advantageous selection is just negative adverse selection.
\[ \hat{p}_{\gamma_I} : \mathbb{R} \to \mathbb{R}_+ \] as required. Using Remark 2, the efficiency contribution at \( T_{\gamma_I} = t \) is \( ec_{\gamma_I}(t) = \mathbb{E}[(W_{\gamma_I} - R_{\gamma_I})^+ | T_{\gamma_I} = t] = \mathbb{E}\left( R_{\gamma_I} - AS_{\gamma_I} - \mathbb{E}[S_{\gamma_I} | T_{\gamma_I} = t] \right)^+ | T_{\gamma_I} = t \). The orthogonality condition (8) therefore implies \( ec_{\gamma_I}(t) = \varphi (\hat{a}s_{\gamma_I} \mathbb{E}[S_{\gamma_I} | T_{\gamma_I} = t]) + \mathbb{E}[S_{\gamma_I} | T_{\gamma_I} = t] \) for the decreasing convex function \( \varphi : \mathbb{R} \to \mathbb{R}_+ \), \( \varphi(z) = \mathbb{E}\left( R_{\gamma_I} - z \right)^+ \). Hence, \( ec_{\gamma_I}(t) \) is a decreasing transformation of \( \mathbb{E}[S_{\gamma_I} | T_{\gamma_I} = t] \). \[ \blacksquare \]

Propositions 2 and 3 highlight the important role played by the apparent match quality \( \mathbb{E}[S_{\gamma_I} | T_{\gamma_I}] \) in determining adverse selection and efficiency. As might be expected, higher adverse selection, as defined here, occurs with reduced trade probability and lower efficiency contribution.

### 4.3 Applications

#### 4.3.1 Released workers may be paid more than retained workers

The familiar intuition from adverse selection models with scalar types is that workers who are released by an employer possessing private information will be paid less than those who are retained (e.g. Gibbons and Katz 1991). In other words, the standard intuition is to expect \( \mathbb{E}[W_{\gamma_I} | R_{\gamma_I} \geq W_{\gamma_I}] > \mathbb{E}[W_{\gamma_I} | R_{\gamma_I} < W_{\gamma_I}] \). In this case, we say firm \( I \) is a high wage firm (under information structure \( \gamma_I \)). Conversely, if \( \mathbb{E}[W_{\gamma_I} | R_{\gamma_I} \geq W_{\gamma_I}] < \mathbb{E}[W_{\gamma_I} | R_{\gamma_I} < W_{\gamma_I}] \), we say firm \( I \) is a low wage firm (under information structure \( \gamma_I \)). In general, there is no reason for the standard intuition to obtain so it becomes of interest to understand what information structures lead to high wage versus low wage firms.\[14\]

**Proposition 4** Suppose the conditions of Proposition 2 hold.

1. If \( \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}] \) and \( \mathbb{E}[S_{\gamma_I} | T_{\gamma_I}] \) are negatively quadrant dependent (NQD)\[15\], then firm \( I \) is a low wage firm (under information structure \( \gamma_I \)).

2. If \((T_{\gamma_I}, R_{\gamma_I}, G_{\gamma_I})\) is affiliated and \( t \to \mathbb{E}[S_{\gamma_I} | T_{\gamma_I} = t] \) is increasing, then \( \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}] \) and \( \mathbb{E}[S_{\gamma_I} | T_{\gamma_I}] \) are positively quadrant dependent (PQD)\[16\], moreover, firm \( I \) is a high wage firm (under information structure \( \gamma_I \)).

**Proof.** 1. From the wage equation (3), \( \mathbb{E}[W_{\gamma_I} P_{\gamma_I}] = \mathbb{E}\left[ \mathbb{E} \left[ G_{\gamma_I} | T_{\gamma_I} \right] P_{\gamma_I} \right] - \mathbb{E}[AS_{\gamma_I} P_{\gamma_I}] \). If \( \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}] \) and \( \mathbb{E}[S_{\gamma_I} | T_{\gamma_I}] \) are NQD, then since \( P_{\gamma_I} \) is a decreasing function of \( \mathbb{E}[S_{\gamma_I} | T_{\gamma_I}] \), \( \mathbb{E}[G_{\gamma_I} | T_{\gamma_I}] \) and \( P_{\gamma_I} \) are PQD and, consequently, have a positive covariance. Therefore,

\[14\] Indeed, to briefly look ahead to Section 6 we will show there that, if information structure \( \gamma_I \) imposes adverse selection efficiently, then it satisfies the condition of Proposition 4 below for firm \( I \) to be a low wage firm under information structure \( \gamma_I \).

\[15\] In the Gaussian case, a sufficient condition for the conclusion is \( \sigma_{G_{\gamma_I} S_{\gamma_I}} > \sigma_{G_{\gamma_I} S_{\gamma_I}|T_{\gamma_I}} \), i.e. if disclosure of public information decreases the correlation between \( G_{\gamma_I} \) and \( S_{\gamma_I} \).

\[16\] In the Gaussian case, a sufficient condition for the conclusion is \( \beta_{G_{\gamma_I} T_{\gamma_I} R_{\gamma_I}} \geq 0, \beta_{G_{\gamma_I} S_{\gamma_I}} \geq 0, \beta_{S_{\gamma_I} T_{\gamma_I}} \geq 0, \) and \( T_{\gamma_I} \) is affiliated.
\[ E[E[G_{\gamma_1}|T_{\gamma_1}] P_{\gamma_1}] \geq E[G_{\gamma_1}|T_{\gamma_1}] E[P_{\gamma_1}] \]. Since Proposition 6 below asserts that \( E[AS_{\gamma_1} P_{\gamma_1}] \leq E[AS_{\gamma_1}] E[P_{\gamma_1}] \), it follows that \( E[W_{\gamma_1} P_{\gamma_1}] \geq E[W_{\gamma_1}] E[P_{\gamma_1}] \), i.e. the average wage of released workers is higher than the average wage, as required.

2. Suppose \( (T_{\gamma_1}, R_{\gamma_1}, G_{\gamma_1}) \) is affiliated. Theorem 5 of Milgrom and Weber (1981) implies that the willingness to pay map \( \Psi_{\gamma_1} \) of Definition 4 maps increasing functions into increasing functions. Hence, the wage schedule, which is the unique fixed point of this map, is an increasing function. If \( t \mapsto E[S_{\gamma_1}|T_{\gamma_1} = t] \) is increasing, then by Proposition 3, \( t \mapsto p_{\gamma_1}(t) \) is decreasing. Hence, since \( T_{\gamma_1} \) is affiliated, the covariance inequality \( E[w_{\gamma_1}(T_{\gamma_1})p_{\gamma_1}(T_{\gamma_1})] \leq E[w_{\gamma_1}(T_{\gamma_1})]E[p_{\gamma_1}(T_{\gamma_1})] \) holds. Since \( E[p_{\gamma_1}(T_{\gamma_1})] = \Pr[R_{\gamma_1} < W_{\gamma_1}] \), Bayes law gives \( E[W_{\gamma_1}] \geq E[W_{\gamma_1}|R_{\gamma_1} < W_{\gamma_1}] \) as required.

The proposition uses Lehmann’s (1966) notions of positive and negative quadrant dependence for pairs of random variables (PQD and NQD respectively). If a pair of random variables are PQD, they are also positively correlated. For the readers convenience, Appendix B gives a brief summary. While our framework allows for multidimensional public and private information, this result makes clear the importance of the correlation or otherwise of just two variables are PQD, they are also positively correlated. For the readers convenience, Appendix B gives a brief summary. While our framework allows for multidimensional public and private information, this result makes clear the importance of the correlation or otherwise of just two scalar statistics, \( E[G_{\gamma_1}|T_{\gamma_1}] \) and \( E[S_{\gamma_1}|T_{\gamma_1}] \). The basic intuition follows from our result that the probability of trade \( P_{\gamma_1} \) is high when \( E[S_{\gamma_1}|T_{\gamma_1}] \) is low. If \( E[G_{\gamma_1}|T_{\gamma_1}] \) and \( E[S_{\gamma_1}|T_{\gamma_1}] \) are ‘negatively correlated’, then it is intuitive that the workers more likely to leave will be those more highly valued by outside firms. The proposition makes this intuition precise.

### 4.3.2 Finding adverse selection when public information is not observed by the econometrician

In our unit-demand model, the positive correlation test for adverse selection (Chiappori and Salanie 2000) becomes a simple test of whether adverse selection is positive. As was stressed by Chiappori and Salanie, since theory implies different observable types should be offered different contracts, one should control for public information and hence test whether \( AS_{\gamma_1} = E[G_{\gamma_1}|T_{\gamma_1}] - E[G_{\gamma_1}|T_{\gamma_1}, R_{\gamma_1} < W_{\gamma_1}] > 0 \). However, if there is unobserved heterogeneity, i.e. \( T_{\gamma_1} \) is poorly observed by the econometrician, this control is difficult. A natural surrogate is to estimate the quantity of adverse selection in the aggregated market constructed by ignoring \( T_{\gamma_1} \), specifically \( E[G_{\gamma_1}] - E[G_{\gamma_1}|R_{\gamma_1} < W_{\gamma_1}] \).

Note that, since \( E[AS_{\gamma_1}] = E[G_{\gamma_1}] - E[W_{\gamma_1}] \), the difference between the quantity of adverse selection in the aggregated market and the average quantity of adverse selection can be written as

\[ E[G_{\gamma_1}] - E[G_{\gamma_1}|R_{\gamma_1} < W_{\gamma_1}] - E[AS_{\gamma_1}] = E[W_{\gamma_1}] - E[W_{\gamma_1}|R_{\gamma_1} < W_{\gamma_1}]. \] (9)

It follows that the quantity of adverse selection in the aggregated market will overestimate (underestimate) the average quantity of adverse selection for high (low) wage firms. Propo-
position 4 therefore has the immediate corollary that the aggregated market will overestimate (underestimate) the average quantity of adverse selection according to whether Condition 1, or 2. of Proposition 4 obtain. The following proposition strengthens the bound obtained in the first part of this corollary and provides conditions for a weaker bound in the second part.

**Proposition 5** Suppose the conditions of Proposition 2 hold. If $\mathbb{E}[G_{\gamma}]T_{\gamma}]$ and $\mathbb{E}[S_{\gamma}]T_{\gamma}]$ are NQD then

$$\mathbb{E}[G_{\gamma}] - \mathbb{E}[G_{\gamma}]T_{\gamma}] < W_{\gamma}] \leq \mathbb{E}[AS_{\gamma}]T_{\gamma}] < W_{\gamma}],$$

If $\mathbb{E}[G_{\gamma}]T_{\gamma}]$ and $\mathbb{E}[S_{\gamma}]T_{\gamma}]$ are PQD the direction of inequality is reversed.

**Proof.** By the definition of adverse selection and the wage equation $\mathbb{E}[AS_{\gamma}]T_{\gamma}] < W_{\gamma}] = \mathbb{E}[\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] - G_{\gamma}]T_{\gamma}] < W_{\gamma}]$. Hence, by straightforward manipulations $\mathbb{E}[G_{\gamma}]T_{\gamma}] - \mathbb{E}[G_{\gamma}]T_{\gamma}] < W_{\gamma}] = \mathbb{E}[AS_{\gamma}]T_{\gamma}] < W_{\gamma}] + \mathbb{E}[\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}]$. The result now follows from noting that $\mathbb{E}[\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}] = \mathbb{E}([\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}] = \mathbb{E}((\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}] - \mathbb{E}(\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}]P_{\gamma}] \mathbb{E}[P_{\gamma}]$. Hence, $\mathbb{E}[G_{\gamma}]T_{\gamma}] - \mathbb{E}[G_{\gamma}]T_{\gamma}] < W_{\gamma}] = \mathbb{E}[AS_{\gamma}]T_{\gamma}] < W_{\gamma}] - \mathbb{Cov}(\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}], \frac{P_{\gamma}}{\mathbb{P}[P_{\gamma}]}).$ Since $\mathbb{Cov}(\mathbb{E}[G_{\gamma}]T_{\gamma}]G_{\gamma}]T_{\gamma}] < W_{\gamma}], \frac{P_{\gamma}}{\mathbb{P}[P_{\gamma}]}$ is positive if $\mathbb{E}[G_{\gamma}]T_{\gamma}]$ and $\mathbb{E}[S_{\gamma}]T_{\gamma}]$ are NQD and negative if PQD, the result is established. 

This result raises the possibility of an effect reversal (Yule 1903) in which $\mathbb{E}[G_{\gamma}]T_{\gamma}] - \mathbb{E}[G_{\gamma}]T_{\gamma}] < W_{\gamma}] \leq 0$ notwithstanding the fact that $AS_{\gamma} > 0$. The intuition for this effect reversal is clearly similar to that of Proposition 4. The probability of trade $P_{\gamma}$ is high when $\mathbb{E}[S_{\gamma}]T_{\gamma}]$ is low. So if $\mathbb{E}[G_{\gamma}]T_{\gamma}]$ is high when $\mathbb{E}[S_{\gamma}]T_{\gamma}]$ is low, the population trading might be better than the population not trading.

Gibbons and Katz (1991) test for adverse selection by comparing wages for workers who are selectively ‘laid off’ and those who become unemployed due to exogenous plant closures. With plant closures, there is no stigma so wages should on average be $\mathbb{E}[G_{\gamma}]T_{\gamma}]$; however for workers who are laid off, expected wages are $\mathbb{E}[W_{\gamma}]T_{\gamma}] < W_{\gamma}] = \mathbb{E}[G_{\gamma}]T_{\gamma}] < W_{\gamma}]$. Hence, (absent the econometrician observing $T_{\gamma}$) the Gibbons and Katz test reduces to testing whether there is adverse selection in the aggregated market: $\mathbb{E}[G_{\gamma}]T_{\gamma}] - \mathbb{E}[G_{\gamma}]T_{\gamma}] < W_{\gamma}] > 0$. Our analysis suggests that this is indeed a necessary condition for adverse selection if $\mathbb{E}[G_{\gamma}]T_{\gamma}]$ and $\mathbb{E}[S_{\gamma}]T_{\gamma}]$ are PQD— in the Gaussian case $\mathbb{Cov}(\mathbb{E}[G_{\gamma}]T_{\gamma}], \mathbb{E}[S_{\gamma}]T_{\gamma}]) \geq 0$. Evidently, an over-sufficient condition for this is if $S_{\gamma}$ and $T_{\gamma}$ are orthogonal.

4.3.3 Adverse selection falls mainly on trades that do not take place

As noted above, in the empirical investigation of markets (potentially) impacted by adverse selection data may well be incomplete. For example, prices of used cars are generally only available for those used cars which are actually traded, while the cost of insurance for the uninsured will generally be harder to obtain than it is for those who actually buy insurance policies.
Proposition 6 Suppose the conditions of Proposition 2 hold. The average quantity of adverse selection conditional on trade taking place is an underestimate of the average quantity of adverse selection. Hence,

\[ 0 \leq \mathbb{E}[AS_{\gamma I} | R_{\gamma I} < W_{\gamma I}] \leq \mathbb{E}[AS_{\gamma I}] \leq \mathbb{E}[AS_{\gamma I} | R_{\gamma I} \geq W_{\gamma I}]. \]

Proof. Since \( AS_{\gamma I} \) and \( P_{\gamma I} \) are respectively increasing and decreasing functions of \( \mathbb{E}[S_{\gamma I} | T_{\gamma I}] \), they are negatively correlated, it follows that \( \mathbb{E}[AS_{\gamma I} | R_{\gamma I} < W_{\gamma I}] = \frac{\mathbb{E}[AS_{\gamma I} P_{\gamma I}]}{\mathbb{E}[P_{\gamma I}]} \leq \mathbb{E}[AS_{\gamma I}]. \)

This proposition establishes that adverse selection falls mainly on trades that do not take place. Hendren (2012) makes a similar point. He argues that severity of asymmetric information explains insurance rejections in non-group health insurance markets. Hence, failure of the positive correlation test for those who obtain (some) insurance does not necessarily mean that there is no adverse selection in the market overall. Hendren’s model retains the scalar type assumption and adverse selection is not always ‘interior’ (in the language of the current paper, for some realization of \( T_{\gamma I} \) the probability of trade is zero, hence insurance rejections). Although the models are different, both point to the obvious importance of taking into account trades which do not occur when estimating the amount of adverse selection in the economy and its impact on market efficiency.

Proposition 6 means that adverse selection concentrates disproportionately where trades do not take place, and we remarked that depending on data availability this might impede its detection. Propositions 4 and 5 are suggestive of another story which will unfold in Section 6. If firms appreciate the power of adverse selection and can influence its effect through information structures and prefer to become low wage firms, then standard tests for adverse selection will tend to fail to find evidence for adverse selection when the econometrician has worse information than market participants. The relationship between adverse selection and efficiency is such that, with endogenous information, adverse selection becomes difficult to detect empirically.

4.3.4 It is optimal to mandate when adverse selection is low rather than high

The canonical solution (Einav and Finkelstein, 2011) to the inefficiencies created by adverse selection is to have mandatory trades. Such policies are quite common in context of insurance markets, although they are seldom seriously proposed in labour markets. A naive view might be that it is optimal to mandate trade in submarkets (i.e. for risk classes in the insurance context) for which adverse selection is high. The following proposition warns against such a conclusion.
Proposition 7 Mandating trade in the submarket defined by \( T_{\gamma I} = t \) is optimal only when adverse selection is low. Specifically, if and only if

\[
E \left[ S_{\gamma I} | T_{\gamma I} \right] \leq -EC_{\gamma I}.
\]

**Proof.** The condition for mandated trade in a submarket to be more efficient than competitive provision is that

\[
E \left[ G_{\gamma I} | T_{\gamma I} = t \right] \geq E \left[ G_{\gamma I} . 1_{\{\text{trade}\}} | T_{\gamma I} = t \right] + E \left[ R_{\gamma I} . 1_{\{\text{no trade}\}} | T_{\gamma I} = t \right].
\]

(10)

The left hand side is the expected value achieved if trade is mandatory, the right hand side is the expected value achieved in the unregulated market. By elementary probability and straightforward rearrangement, this becomes \( E \left[ R_{\gamma I} . 1_{\{\text{trade}\}} | T_{\gamma I} \right] - E \left[ G_{\gamma I} . 1_{\{\text{trade}\}} | T_{\gamma I} \right] \geq E \left[ S_{\gamma I} | T_{\gamma I} \right] \). It is an implication of the wage equation that \( E \left[ G_{\gamma I} . 1_{\{\text{trade}\}} | T_{\gamma I} \right] = E \left[ W_{\gamma I} . 1_{\{\text{trade}\}} | T_{\gamma I} \right] \). Since \( R_{\gamma I} - W_{\gamma I} \leq 0 \) with probability one when trade takes place, inequality (10) becomes

\[
E \left[ S_{\gamma I} | T_{\gamma I} \right] \leq -E \left[ (W_{\gamma I} - R_{\gamma I})+ | T_{\gamma I} \right] = -EC_{\gamma I} < 0.
\]

This condition requires the apparent match quality to be low and so, by Proposition 2, adverse selection must also be low. It follows that, under the conditions of Proposition 2, it is not optimal to mandate trade in submarkets for which adverse selection is high—rather the opposite is true.

4.3.5 There are three dimensions of information (Gaussian case)

For Gaussian information structures, \( \Gamma_I \) is identified with a set of symmetric positive semidefinite covariance matrices. Notwithstanding the richness of this parameter space, the following corollary of Proposition 2 establishes that the marginal distribution of adverse selection depends on the information structure \( \gamma_I \in \Gamma_I \) only through a three dimensional parameter space.

Proposition 8 Suppose information structures are Gaussian and \( \gamma_I \in \Gamma_I \) induces regression to the mean. \( AS_{\gamma I} = \tilde{a}s_{\gamma I} (E[S_{\gamma I} | T_{\gamma I}]) \) with \( \tilde{a}s_{\gamma I} : \mathbb{R} \rightarrow \mathbb{R}_+ \) implicitly defined by

\[
\tilde{a}s_{\gamma I} = \beta_{G_{\gamma I} R_{\gamma I} | T_{\gamma I}} \sigma_{R_{\gamma I} | T_{\gamma I}} h \left( \frac{\tilde{a}s_{\gamma I} + E[S_{\gamma I} | T_{\gamma I} = t]}{\sigma_{R_{\gamma I} | T_{\gamma I}}} \right),
\]

(11)

where \( h : \mathbb{R} \rightarrow \mathbb{R}_+ \) is the inverse Mills ratio (i.e. the normal hazard function).

**Proof.** Recall from the proof of Proposition 2 that \( as_{\gamma I} (t) = E[G_{\gamma I} | as_{\gamma I} (t) + E[S_{\gamma I} | T_{\gamma I} = t] < R_{\gamma I}] \). Noting that, by the Law of Iterated Expectations \( E[G_{\gamma I} | z \leq R_{\gamma I}] = E[E[G_{\gamma I} | R_{\gamma I}] | z \leq R_{\gamma I}] \), \( E[G_{\gamma I} | R_{\gamma I}] = \beta_{G_{\gamma I} R_{\gamma I} | T_{\gamma I}} R_{\gamma I} \), \( R_{\gamma I} \sim N \left( 0, \sigma_{R_{\gamma I} | T_{\gamma I}} \right) \), and \( h (z) = E[R_{\gamma I} | z \leq R_{\gamma I}] \) is the inverse Mills ratio, establishes the implicit representation in (11). ■

These parameters have a natural economic interpretation:
• $\sigma_{R_{\gamma}|T_{\gamma}}$ is the $(Y_{II})$ information gap. It is a measure of the difference in the amount of information about $Y_{II}$ held by the informed and uninformed parties. For Gaussian distributions, the variance of the conditional expectation $\text{Var}(E[Y_{II}|Q_{\gamma}])$ is a bona fide measure of how informative $Q_{\gamma}$ is about $Y_{II}$, similarly $\text{Var}(E[Y_{II}|T_{\gamma}])$ measures how much information is in $T_{\gamma}$. By the law of total variance, we have $\sigma_{R_{\gamma}|T_{\gamma}}^2 = \text{Var}(E[Y_{II}|Q_{\gamma}]) - \text{Var}(E[Y_{II}|T_{\gamma}])$.

• $\beta_{G_{\gamma}R_{\gamma},T_{\gamma}}$ is informed party relevance. It is a measure of the relevance of $R_{\gamma}$ to $G_{\gamma}$ given public information $T_{\gamma}$. Note that we may also write $\beta_{G_{\gamma}R_{\gamma},T_{\gamma}}^2 = \text{Var}(E[Y_{II}|Q_{\gamma},T_{\gamma}]) - \text{Var}(E[Y_{II}|T_{\gamma}])$. Therefore given $\sigma_{R_{\gamma}|T_{\gamma}}^2$, $\beta_{G_{\gamma}R_{\gamma},T_{\gamma}}$ is a measure of how much extra information $(R_{\gamma}, T_{\gamma})$ contains about $G_{\gamma}$ than just the public information $T_{\gamma}$.

• $\sigma_{|S_{\gamma}|T_{\gamma}}$ is the quality of public information about match quality. It is a measure of how much information is available publicly about $Y_{II} - Y_{IJ}$.

Proposition 8 applies, more generally, if the information structure is conditionally Gaussian. That is, if $T_{\gamma} = \left(T_{\gamma}', T_{\gamma}''\right)$ such that $(Y_{II}, ..., Y_{Inp}, Q_{\gamma}, T_{\gamma}')$ has a Gaussian distribution conditional on $T_{\gamma}''$. A version of equation (11) holds conditional on each realization of $T_{\gamma}''$. For example, there may be one version of (11) for men and another for women. Notice (11) confirms that adverse selection is non-negative given regression to the mean ($0 \leq \beta_{G_{\gamma}R_{\gamma},T_{\gamma}} < 1$). A unique fixed point of the willingness to pay map $\Psi_{\gamma}$ exists if $\beta_{G_{\gamma}R_{\gamma},T_{\gamma}}^2 < 0$. In this case, $\tilde{s}_{\gamma}$ is a decreasing negative function; $G_{\gamma}$ and $R_{\gamma}$ are negatively correlated given public information so it is good news that $R_{\gamma} < W_{\gamma}$.

5 Adverse Selection and Overall (Expected) Efficiency

We now turn to the question of how overall efficiency aggregated across submarkets depends on the information structure, and how this relates to aggregate adverse selection. We use the Gaussian case, continuing the analysis of Section 4.3.5. As was shown there, as far as characterizing the marginal distribution of adverse selection, information can be captured by a three dimensional parameter. A similar result holds for the efficiency contribution. The following two subsections characterize how expected adverse selection and expected efficiency vary over this parameter space.

---

Note $\text{Var}(E[S_{\gamma}|T_{\gamma}]) = \text{Var}(E[Y_{II} - Y_{IJ}|T_{\gamma}])$ determines the distribution of $E[S_{\gamma}|T_{\gamma}]$. By construction, $E[E[S_{\gamma}|T_{\gamma}]] = E[Y_{II} - Y_{IJ}]$ independently of the information structure so, in the Gaussian case, its distribution is determined solely by its variance, $\text{Var}(E[S_{\gamma}|T_{\gamma}])$. 

---

25
5.1 General Results

**Proposition 9** Suppose information structures are Gaussian and \( \gamma_I \in \Gamma_I \) induces regression to the mean. \( \mathbb{E}[AS_{\gamma_I}] \) and \( \mathbb{E}[EC_{\gamma_I}] \) depend on the information structure \( \gamma_I \in \Gamma_I \) only through the three dimensional parameter \((\beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}}, \sigma_{R_{\gamma_I}T_{\gamma_I}}, \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}) \in [0, 1) \times \mathbb{R}_+^2 \). Furthermore,

1. \( \mathbb{E}[AS_{\gamma_I}] \) is increasing in \((\beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}}, \sigma_{R_{\gamma_I}T_{\gamma_I}}, \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}) \).
2. For each \((\sigma_{R_{\gamma_I}T_{\gamma_I}}, \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}) \in \mathbb{R}_+^2, \mathbb{E}[EC_{\gamma_I}] \) is decreasing in \( \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} \).
3. For each \((\beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}}, \sigma_{R_{\gamma_I}T_{\gamma_I}}) \in [0, \frac{7}{9}] \times \mathbb{R}_+, \mathbb{E}[EC_{\gamma_I}] \) is increasing in \( \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]} \).

**Proof.** Appendix C. \( \blacksquare \)

Proposition 9, Part 1, establishes that expected adverse selection is increasing in all three parameters. It is intuitive that \( \mathbb{E}[AS_{\gamma_I}] \) is increasing in the information gap and informed party relevance parameters. The reason why \( \mathbb{E}[AS_{\gamma_I}] \) is increasing in the quality of public information about match quality is slightly less obvious but it arises via Jensen’s inequality from a natural convexity—when apparent match quality is high adverse selection will be positive and large, but when apparent match quality is low adverse selection remains bounded from below by zero.

Parts 2 and 3 of the proposition establish the effect of the parameters on expected efficiency. Comparing two information structures with the same \( \sigma_{R_{\gamma_I}T_{\gamma_I}} \) and \( \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]} \)\(^{18} \), the information structure for which \( \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} \) is larger leads to less expected efficiency. Hence, there is a trade off between \( \mathbb{E}[AS_{\gamma_I}] \) and \( \mathbb{E}[EC_{\gamma_I}] \) for fixed \( \sigma_{R_{\gamma_I}T_{\gamma_I}} \) and \( \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]} \). This is consistent with the familiar intuition from models with scalar types and no public information. On the other hand, holding \( \sigma_{R_{\gamma_I}T_{\gamma_I}} \) and \( \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}} \) fixed (with the latter not too large), better public information about the match quality increases both expected adverse selection and expected efficiency.

**Proposition 10** Suppose information structures are Gaussian and \( \gamma_I \in \Gamma_I \) induces regression to the mean. For each fixed \( \sigma_{R_{\gamma_I}T_{\gamma_I}}, \mathbb{E}[EC_{\gamma_I}] \) and \( \mathbb{E}[AS_{\gamma_I}] \) satisfy a strict Spence-Mirrlees condition in the remaining parameters. Specifically,

\[
\frac{\partial \mathbb{E}[AS_{\gamma_I}]}{\partial \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}} / \frac{\partial \mathbb{E}[AS_{\gamma_I}]}{\partial \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}}} > \frac{\partial \mathbb{E}[EC_{\gamma_I}]}{\partial \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}} / \frac{\partial \mathbb{E}[EC_{\gamma_I}]}{\partial \beta_{G_{\gamma_I}R_{\gamma_I}T_{\gamma_I}}}.
\]

**Proof.** Appendix C. \( \blacksquare \)

Comparing two information structures with \( \sigma_{R_{\gamma_I}T_{\gamma_I}} \) fixed and which generate the same amount of expected adverse selection, the information structure with better public information about match quality leads to higher expected efficiency.

---

\(^{18}\)For instance, this would arise under two information structures with the same public information and with private information equally good for estimating \( Y_{II} \). To see this, note first that private information \( Q_{\gamma_I} \) does not affect \( \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]} \) (recall \( \text{Var}(\mathbb{E}[S_{\gamma_I}|T_{\gamma_I}]) = \text{Var}(\mathbb{E}[Y_{II} - Y_{II}|T_{\gamma_I}]) \)). Second, \( \sigma_{R_{\gamma_I}T_{\gamma_I}} \) is a measure of the quality of information about \( Y_{II} \).
Propositions 9 and 10 have a useful implication which is illustrated in Figure 1. The figure illustrates the level sets of $E[AS_{\gamma}]$ and $E[EC_{\gamma}]$, their general shape are as established in the Propositions. It is assumed that the vertex of the cone which is at $(0, \beta)$ satisfies $\beta \leq \frac{\pi}{4}$. On the positively oriented extreme ray of the cone, Proposition 10 applies to ensure that there is a direction of increase of both $E[AS_{\gamma}]$ and $E[EC_{\gamma}]$ pointing into the interior of the cone. Similarly, on the negatively oriented extreme ray of the cone, Proposition 9 applies to ensure that there is a direction of increase of both $E[AS_{\gamma}]$ and $E[EC_{\gamma}]$ pointing into the interior of the cone. Hence, there is always a direction pointing into the interior of the cone in which both $E[AS_{\gamma}]$ and $E[EC_{\gamma}]$ are increasing. Moreover, as illustrated, this direction may be chosen such that $\sigma_{E[S_{\gamma_1}T_{\gamma_1}]}$ is increasing and $|\beta - \beta_{G_{\gamma_1}R_{\gamma_1}T_{\gamma_1}}|$ is decreasing. This is stated formally in Corollary 1 below.

**Definition 8** Let $C$ be a cone in $R_+ \times R$ with vertex $(0, \beta)$ and non-empty interior which contains the horizontal line segment through its vertex, i.e. $\{(a_1, \beta) \in R_2^2|a_1 \geq 0\}$. Let $\hat{a} = (\hat{a}_1, \hat{a}_2)$ lie on an extreme ray of $C$. We say the path $p : R_+ \rightarrow C, p(a_1) = (a_1, \varphi(a_1))$ with $\varphi : R_+ \rightarrow R$ is from $\hat{a}$ into the body of the cone if $p(\hat{a}_1) = (\hat{a}_1, \hat{a}_2)$ and $a_1 \rightarrow |\varphi(a_1) - \beta|$ is decreasing on $[a_1, \infty)$.

**Corollary 1** Let $C$ be a closed convex cone satisfying the conditions of Definition 8 for some $0 \leq \beta \leq \frac{\pi}{4}$. Suppose $\hat{a} = (\sigma_{E[S_{\gamma_1}T_{\gamma_1}]}, \beta_{G_{\gamma_1}R_{\gamma_1}T_{\gamma_1}}) \in R_+ \times [0, 1)$ lies on an extreme ray of $C$, there is a path from $\hat{a}$ into the body of the cone along which both $E[AS_{\gamma}]$ and $E[EC_{\gamma}]$ are increasing.
Proof. On the positively oriented extreme ray of the cone, Proposition 10 together with the fact that \( \mathbb{E}[AS_{\gamma_1}] \) is increasing in both \( \sigma_{p, S_{\gamma_1}|T_{\gamma_1}} \) and \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \) applies to ensure that there is a direction of increase of \( \mathbb{E}[S_{\gamma_1}|T_{\gamma_1}] \) and decrease of \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \) which increases both \( \mathbb{E}[AS_{\gamma_1}] \) and \( \mathbb{E}[EC_{\gamma_1}] \). This establishes to conclusion if \( \alpha \) is on the positively oriented extreme ray. Similarly, on the negatively oriented extreme ray of the cone, Proposition 9 applies—increasing \( \sigma_{p, S_{\gamma_1}|T_{\gamma_1}} \) with \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \) constant suffices. ■

5.2 A Simple Case: Public Information Orthogonal to Match Quality

Recall that both adverse selection and efficiency contribution depend on the realization of public information only through the scalar \( \mathbb{E}[S_{\gamma_1}|T_{\gamma_1}] \). Hence, for the slice of the parameter space with \( \sigma_{p, S_{\gamma_1}|T_{\gamma_1}} = 0 \), both adverse selection and efficiency contribution are constants independent of \( T_{\gamma_1} \). Since, \( \frac{AS_{\gamma_1}}{\sigma_{R_{\gamma_1}|T_{\gamma_1}}} = \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} h \left( \frac{AS_{\gamma_1}}{\sigma_{R_{\gamma_1}|T_{\gamma_1}}} \right) \), we can define \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \) as an explicit, increasing, function of \( \frac{AS_{\gamma_1}}{\sigma_{R_{\gamma_1}|T_{\gamma_1}}} \). Denoting this function \( x \mapsto \phi(x) = x/h(x) \), adverse selection has in this case the simple, more-or-less explicit representation \( AS_{\gamma_1} = \sigma_{R_{\gamma_1}|T_{\gamma_1}} \phi^{-1}(\beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}}) \).

It can also be shown that

\[
EC_{\gamma_1} = \sigma_{R_{\gamma_1}|T_{\gamma_1}} \left( 1 - \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \right) f \left( \phi^{-1}(\beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}}) \right),
\]

where, \( f \) is the \( \mathcal{N}(0,1) \) density. Hence, \( EC_{\gamma_1} \) is decreasing in \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \) on \([0,1]\) and increasing in \( \sigma_{R_{\gamma_1}|T_{\gamma_1}} \). It follows from regression to the mean and orthogonality of \( T_{\gamma_1} \) and \( S_{\gamma_1} \) that \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \) decreases as more information is publicly disclosed. As we have remarked, it follows from the law of total variance that \( \sigma_{R_{\gamma_1}|T_{\gamma_1}} \) is decreasing in information disclosure. Therefore, increased information disclosure decreases adverse selection.

We conclude this subsection by elaborating on the sampling example from the Introduction to illustrate how changes in the information structure impact upon adverse selection and efficiency via changes in \( \sigma_{R_{\gamma_1}|T_{\gamma_1}} \) and \( \beta_{G_{\gamma_1}, R_{\gamma_1}, T_{\gamma_1}} \).

Example 1. The parties acquire noisy samples of the value of the worker in employment. For each firm \( I, J \in F \)

\[
Q_{\gamma_1} = (Y_{II} + \varepsilon_1, ..., Y_{II} + \varepsilon_M, X_I + \nu_1, ..., X_I + \nu_N),
\]

\[
T_{\gamma_1} = (X_I + \nu_1, ..., X_I + \nu_n), \quad n \leq N.
\]

Private information \( Q_{\gamma_1} \) consists of \( M \) samples of the worker’s productivity at the training firm \( Y_{II} = X_I + Z_I \) and \( N \) samples of her general ability \( X_I \). Some of the samples of general ability are publicly available in \( T_{\gamma_1} \). Hence, elements of \( \Gamma_I \) are identified with three numbers: \( M, N, \) and \( n, \Gamma_I \subset \{(M, N, n) \in \mathbb{N}^3 \mid 0 \leq n \leq N, 0 \leq M \} \). All random variables are independently distributed, and assumed to be \( \mathcal{N}(0,1) \) except for \( X_I \sim \mathcal{N}(\mu_X, \sigma_X^2) \) and \( Z_I \sim \mathcal{N}(0, \sigma_Z^2) \).
Below, we describe how changes in the information structure, that increase the asymmetry in information between the training and outside firms, affect adverse selection and efficiency.

a. **Disclosing fewer samples of general ability.**

Suppose \( T_{\gamma_I} \) contains fewer samples of general ability \( X_I \) than \( T_{\gamma_f} \) (lower \( n \) with no change in \( M \) and \( N \)). Both the information gap and informed party relevance are bigger under \( \gamma'_I \) than \( \gamma_I \). Hence, \( AS_{\gamma'_I} > AS_{\gamma_I} \). To see that \( EC_{\gamma'_I} < EC_{\gamma_I} \), note that 

\[
\beta_{G_{\gamma_I}R_{\gamma_I},T_{\gamma_I}} = 1 - \beta_{S_{\gamma_I}R_{\gamma_I},T_{\gamma_I}}.
\]

Using the orthogonality of \( S_{\gamma_I} \) and \( T_{\gamma_I} \), it follows that 

\[
1 - \beta_{G_{\gamma_I}R_{\gamma_I},T_{\gamma_I}} = \frac{\sigma_{S_{\gamma_I}R_{\gamma_I}}}{\sigma_{R_{\gamma_I}|T_{\gamma_I}}}
\]

therefore, since \( \sigma_{S_{\gamma_I}R_{\gamma_I}} \) is unaffected by information disclosure, 

\[
\sigma_{R_{\gamma_I}|T_{\gamma_I}} \left( 1 - \beta_{G_{\gamma_I}R_{\gamma_I},T_{\gamma_I}} \right) = \frac{\sigma_{S_{\gamma_I}R_{\gamma_I}}}{\sigma_{R_{\gamma_I}|T_{\gamma_I}}}
\]

is decreasing in the information gap. This gives the result.

b. **Acquiring more samples of general ability.**

Suppose \( Q_{\gamma_I} \) contains more samples of \( X_I \) than \( Q_{\gamma_f} \) (higher \( N \) with no change in \( M \) and \( n \)). It can be shown that both the information gap and informed party relevance are bigger under \( \gamma'_I \) than \( \gamma_I \). Hence, it is immediate that \( AS_{\gamma'_I} > AS_{\gamma_I} \). Using (12), we can also show that \( EC_{\gamma'_I} < EC_{\gamma_I} \). As in part a., this is established by showing that collecting more samples of \( X_I \) decreases \( \beta_{S_{\gamma_I}R_{\gamma_I},T_{\gamma_I}} \sigma_{R_{\gamma_I}|T_{\gamma_I}} \). The intuition is that collecting more samples of \( X_I \) makes \( R_{\gamma_I} \) a noisier measure of \( Z_I \) given public information. Note that 

\[
\beta_{S_{\gamma_I}R_{\gamma_I},T_{\gamma_I}} \sigma_{R_{\gamma_I}|T_{\gamma_I}}
\]

is a measure of how much information \( R_{\gamma_I} \) contains about \( Z_I \).

c. **Acquiring more samples of the worker’s productivity during training.**

In a. and b. above, adverse selection and efficiency go in opposite directions. Here is a case where they go together. Suppose \( Q_{\gamma_I} = T_{\gamma_I} = X_I + \nu_1 \), then \( AS_{\gamma_I} = EC_{\gamma_I} = 0 \). However, with \( Q_{\gamma'_I} = (X_I + Z_I + \epsilon_1, X_I + \nu_1) \), \( T_{\gamma'_I} = X_I + \nu_1 \) (higher \( M \) with no change in \( N \) and \( n \)) there is both positive adverse selection and efficiency contribution.

6 **Imposing Adverse Selection Efficiently**

The results in the previous section tell us how \( \mathbb{E}[AS_{\gamma_I}] \) and \( \mathbb{E}[EC_{\gamma_I}] \) vary over different information structures. However, when considering different information structures there are also constraints. For instance, firms cannot disclose information which they do not possess. In this section, we introduce these constraints and explore the question of which information structures impose a given amount of adverse selection with the least loss of efficiency. The motivation for doing so stems from the employer learning application. For firms strictly constrained by the lower bound on training wages, expected adverse selection determines expected second period wages and therefore the expected present value of the worker’s career if she chooses to train at the firm. Hence, for such firms, attracting the worker requires expected adverse selection to be below some level determined by competition in the labour market. Adverse selection
also determines the distribution of surplus between the training firm and the worker. It will generally be optimal to impose the amount of expected adverse selection determined by this competitive constraint, yet do so in a way that generates as much surplus as possible.

Having established some preliminaries, we break the analysis into two stages. First, we consider the case where $\Gamma_I$ contains only information structures with the same private information. Hence, we analyze different public information disclosures. The main result is the identification of the public information disclosures associated with efficient imposition of adverse selection. We then explore the effect of acquiring inside information. Here, we are able to show that, under certain conditions and providing there is sufficient freedom to transmit information publicly, improving private information always facilitates a weakly more efficient way of imposing adverse selection. This is very natural—more information is better—but the statement is not obviously true a priori.

### 6.1 Preliminaries

In the Gaussian case, $\Gamma_I$ can be identified one-to-one with a set of symmetric positive semidefinite covariance matrices partitioned as

$$
\Sigma_{\gamma_I} = \begin{bmatrix}
\Sigma_Y & \Sigma_{YQ_{\gamma_I}} & \Sigma_{YT_{\gamma_I}} \\
\Sigma_{Q_{\gamma_I}Y} & \Sigma_{Q_{\gamma_I}} & \Sigma_{Q_{\gamma_I}T_{\gamma_I}} \\
\Sigma_{T_{\gamma_I}Y} & \Sigma_{T_{\gamma_I}Q_{\gamma_I}} & \Sigma_{T_{\gamma_I}} \\
\end{bmatrix},
$$

(13)
each containing $\Sigma_Y$ as first $n_Y \times n_Y$ elements and with $\Sigma_{Q_{\gamma_I}}$ and $\Sigma_{T_{\gamma_I}}$ positive definite (if necessary delete redundant elements). $Y$ denotes a vector of employment values. Note that we write $\Sigma_Y$ rather than $\Sigma_{YY}$ for later convenience. Given Assumption A1, and the equilibrium established in Proposition 1, it is clear that we may simplify and take $Y$ to be the two dimensional vector $(Y_{II}, Y_{IJ})$ without loss of generality. Using this notation, Assumption A3 may be written as $\Sigma_{YT_{\gamma_I}} = \Sigma_{YQ_{\gamma_I}} \Sigma_{Q_{\gamma_I}^{-1}} \Sigma_{Q_{\gamma_I}T_{\gamma_I}}$.

**Definition 9** Firm $I \in F$ has fixed private information $Q_I$ if for each $\gamma_I \in \Gamma_I$, $(Y_{\gamma_I}, Q_{\gamma_I}, T_{\gamma_I})$ is equal in distribution to $(Y, Q, T)$. Hence, $R_{I\gamma_I} = \mathbb{E}[Y_{IJ}|Q_I] = R_I$ and $G_{I\gamma_I} = \mathbb{E}[Y_{II}|Q_I] = G_I$, for each $\gamma_I \in \Gamma_I$.

**Notation 4** Let $S_n$ denote the set of $n \times n$ symmetric matrices, $S^+_n \subset S_n$ positive semidefinite, $S^{++}_n \subset S^+_n$ positive definite. For $n = n_Y + n_{Q_I} + n_{T_{\gamma_I}}$, $S^+ (\Sigma_{(Y,Q_{\gamma_I})})$ denotes the set of all finite symmetric positive semidefinite matrices partitioned as in (13) with

$$
\Sigma_{(Y,Q_{\gamma_I})} = \begin{bmatrix}
\Sigma_Y & \Sigma_{YQ_{\gamma_I}} \\
\Sigma_{Q_{\gamma_I}Y} & \Sigma_{Q_{\gamma_I}} \\
\end{bmatrix}
$$

(14)as the first $(n_Y + n_{Q_I}) \times (n_Y + n_{Q_I})$ elements. For $A, B \in S_n$, we write $A \preceq B$ if $B - A \in S^+_n$. 


and \( A < B \) if \( B - A \in \mathcal{S}_{n}^{+} \). \( 0 \in \mathcal{S}_{n} \) denotes the \( n \)-dimensional null matrix, hence \( A \succ 0 \) means \( A \) is positive definite. \( \Sigma_{(R_{I}, G_{I})} \) and \( \Sigma_{(R_{I}, G_{I})|T_{\gamma_{I}}} \) denote the \( 2 \times 2 \) unconditional and conditional covariance matrices:

\[
\Sigma_{(R_{I}, G_{I})} = \begin{pmatrix}
\sigma_{R_{I}R_{I}} & \sigma_{G_{I}R_{I}} \\
\sigma_{G_{I}R_{I}} & \sigma_{G_{I}G_{I}}
\end{pmatrix}, \quad \Sigma_{(R_{I}, G_{I})|T_{\gamma_{I}}} = \begin{pmatrix}
\sigma_{R_{I}R_{I}|T_{\gamma_{I}}} & \sigma_{G_{I}R_{I}|T_{\gamma_{I}}} \\
\sigma_{G_{I}R_{I}|T_{\gamma_{I}}} & \sigma_{G_{I}G_{I}|T_{\gamma_{I}}}
\end{pmatrix}.
\]

We now make precise what it means for an information structure to contain more private information. Suppose \( Q_{I} \sim \mathcal{N}(0, \Sigma_{\gamma_{I}}) \) and \( Q'_{I} = (Q_{I}, Q'_{I}) \sim \mathcal{N}(0, \Sigma_{\gamma_{I}'}) \). Then \( Q'_{I} \) contains \( Q_{I} \) and so is evidently more informative. With \( R_{I} = \mathbb{E}[Y_{I}|Q_{I}], \ R'_{I} = \mathbb{E}[Y_{I}|Q'_{I}] \), the law of iterated expectations implies \( R_{I} = \mathbb{E}[R'_{I}|Q_{I}] \) and \( \text{Var}(R_{I}) = \text{Var}(\mathbb{E}[R'_{I}|Q_{I}]). \) The law of total variance states \( \text{Var}(\mathbb{E}[R'_{I}|Q_{I}]) = \text{Var}(R'_{I}) - \mathbb{E}[\text{Var}(R'_{I}|Q_{I})]. \) Hence, \( \text{Var}(R'_{I}) = \text{Var}(R_{I}) + \mathbb{E}[\text{Var}(R'_{I}|Q_{I})]. \) More generally, in the Gaussian case, the covariance matrix of \( (R'_{I}, G'_{I}) \) equals the covariance matrix of \( (R_{I}, G_{I}) \) plus the conditional covariance matrix of \( (R'_{I}, G'_{I}) \) given \( Q_{I} \); that is, \( \Sigma_{(R'_{I}, G'_{I})} = \Sigma_{(R_{I}, G_{I})} + \Sigma_{(R'_{I}, G'_{I})|Q_{I}} \). Adding information to \( Q_{I} \) therefore has the effect of increasing, in the positive semidefiniteness order \( \succeq \), the covariance matrix of value estimates, \( \Sigma_{(R'_{I}, G'_{I})} \succeq \Sigma_{(R_{I}, G_{I})} \). This means that \( \Sigma_{(R'_{I}, G'_{I})} \succeq \Sigma_{(R_{I}, G_{I})} \) is the natural criterion to represent that \( Q'_{I} \) is more informative about the vector \((Y_{I}, Y'_{I})\), than is \( Q_{I} \).

**Definition 10** If information structures are Gaussian, \( \gamma'_{I} \in \Gamma_{I} \) has more private information than \( \gamma_{I} \in \Gamma_{I} \) if \( \Sigma_{(R_{I}, G_{I})} \succeq \Sigma_{(R'_{I}, G'_{I})} \).

**6.2 Disclosure with Fixed Private Information**

With fixed private information, different elements of \( \Gamma_{I} \) correspond to different public information disclosures. We are especially interested in the case of Gaussian information structures where there are no further constraints on information disclosure apart from the fact that one cannot disclose what is not known (Assumption A3). We write \( \Gamma_{I} = \Gamma_{Q_{I}} \) to denote the set of such Gaussian information structures with fixed private information \( Q_{I} \). \( \Gamma_{Q_{I}} \) can be identified one-to-one with the subset of \( \mathcal{S}^{+}(\Sigma_{(Y_{I}, Q_{I})}) \) satisfying \( \Sigma_{YT_{\gamma_{I}}} = \Sigma_{YQ_{I}} \Sigma_{Q_{I}^{-1}} \Sigma_{Q_{I}T_{\gamma_{I}}} \). We can therefore define the set of possible expected adverse selection-efficiency pairs as

\[
\Omega(\Gamma_{Q_{I}}) = \left\{ (\mathbb{E}[AS_{\gamma_{I}}], \mathbb{E}[EC_{\gamma_{I}}]) \mid \Sigma_{\gamma_{I}} \in \mathcal{S}^{+}(\Sigma_{(Y_{I}, Q_{I})}), \Sigma_{YT_{\gamma_{I}}} = \Sigma_{YQ_{I}} \Sigma_{Q_{I}^{-1}} \Sigma_{Q_{I}T_{\gamma_{I}}} \right\}.
\]

The set \( \mathcal{S}^{+}(\Sigma_{(Y_{I}, Q_{I})}) \) of positive semidefinite matrices with (14) as initial block of elements is a rather large one from which to choose, especially since the overall dimensionality is not specified a priori. However, Proposition 9 established that adverse selection \( \mathbb{E}[AS_{\gamma_{I}}] \) and efficiency \( \mathbb{E}[EC_{\gamma_{I}}] \) depend on a much lower dimensional parameter space. The following lemma achieves a similar reduction in dimensionality to three (the elements of a symmetric \( 2 \times 2 \) matrix) bounded by semidefinite constraints.
Lemma 1  The set $\Omega(\Gamma_{\gamma, G_I})$ can be parameterized equivalently as

$$\Omega(\Gamma_{Q_I}) = \left\{ \left( \mathbb{E}[AS_{\gamma_I}], \mathbb{E}[EC_{\gamma_I}] \right) \mid \Sigma_{(R_I,G_I)|T_{\gamma_I}} \in \mathcal{S}_2, 0 \leq \Sigma_{(R_I,G_I)|T_{\gamma_I}} \preceq \Sigma_{(R_I,G_I)} \right\}.$$  

**Proof.** Appendix D. □

Given inside information $Q_I$, what type of public information $T_{\gamma_I}$ generates the maximum expected efficiency for a given expected quantity of adverse selection? Lemma 1 implies this question can be formulated as the following nonlinear semidefinite program:

$$V(\Sigma_{(R_I,G_I)}, \mathcal{S}) = \max_{\Sigma_{(R_I,G_I)|T_{\gamma_I}} \in \mathcal{S}_2} \mathbb{E}[EC_{\gamma_I}]$$

subject to $\mathbb{E}[AS_{\gamma_I}] \geq \mathcal{S}$, and $0 \preceq \Sigma_{(R_I,G_I)|T_{\gamma_I}} \preceq \Sigma_{(R_I,G_I)}$.

Thus the problem is reduced to choosing a $2 \times 2$ conditional covariance matrix $\Sigma_{(R_I,G_I)|T_{\gamma_I}} \in \mathcal{S}_2$ for $(R_I,G_I)$. In what follows, we will say $\gamma_I^* \in \Gamma_I$ solves the program P1 if $\Sigma_{(R_I,G_I)|T_{\gamma_I}} \in \mathcal{S}_2$ solves P1. Our next result establishes conditions for this optimum. Since the statistic $T_{\gamma_I}$ is not uniquely defined (because any full rank linear transformation of $T_{\gamma_I}$ carries the same information), the $T_{\gamma_I}^*$ identified in the statement of this proposition should be understood as unique up to equivalence classes.

**Proposition 11** Suppose information structures are Gaussian with fixed inside information $Q_I$, and suppose $\beta_{G_R_I} \leq \frac{4}{M}$. If $\gamma_I^* \in \Gamma_I$ solves program P1 for some $\mathcal{S} \geq 0$, then $\Sigma_{(R_I,G_I)|T_{\gamma_I}}$ is singular. Furthermore,

1. If $G_I$ and $R_I$ are colinear ($G_I = \beta_0 + \beta_{G,R_I} R_I$), then for some $\varepsilon \perp G_I, \varepsilon \sim N(0, \sigma_\varepsilon)$, $T_{\gamma_I} = G_I + \varepsilon$.

2. If $G_I$ and $R_I$ are not colinear, $T_{\gamma_I}^*$ is two dimensional and for some $\alpha \in \mathbb{R}^2$, $G_I = \beta_{G,I} R_I + \alpha \cdot T_{\gamma_I}^*$.

**Proof.** Appendix D. □

The result identifies (up to equivalence classes) the public information disclosures associated with efficient imposition of adverse selection. The constraints in P1 bind in a particular way—such that the matrix $\Sigma_{(R_I,G_I)|T_{\gamma_I}}$ is singular. This condition implies that, given $T_{\gamma_I}^*$, $R_I$ and $G_I$ have a singular covariance matrix and therefore that there is an affine relationship between them.

The basic idea of the proof follows easily from Corollary 1. Either the constraint $\Sigma_{(R_I,G_I)|T_{\gamma_I}} \preceq \Sigma_{(R_I,G_I)}$ binds, or $0 \preceq \Sigma_{(R_I,G_I)|T_{\gamma_I}}$ binds. In the latter case $\Sigma_{(R_I,G_I)|T_{\gamma_I}}$ is evidently singular as claimed. The constraint $\Sigma_{(R_I,G_I)|T_{\gamma_I}} \preceq \Sigma_{(R_I,G_I)}$ can be expressed as a conical one to which Corollary 1 applies, implying there is a path of increasing $\mathbb{E}[AS_{\gamma_I}]$ and $\mathbb{E}[EC_{\gamma_I}]$ into the body of the cone—in other words, the constraint $\Sigma_{(R_I,G_I)|T_{\gamma_I}} \preceq \Sigma_{(R_I,G_I)}$ does not bind (in isolation). Hence, the result.
Empirical Implications  Suppose $G_I$ and $R_I$ are not colinear, which is the interesting case. By Proposition 11, $G_I$ and $S_I$ have a positive conditional correlation given $T_{γ^* I}$. It follows from the law of total probability (law of total covariance) that $\text{Cov} \left( \mathbb{E} \left[ G_I | T_{γ^* I} \right], \mathbb{E} \left[ S_I | T_{γ^* I} \right] \right) < \text{Cov}(G_I, S_I)$. Hence, if $\text{Cov}(G_I, S_I) \leq 0$, as will be reasonable to assume in certain applications—e.g. it is true in the specification of Example 1 when private information is very good (in the limit of large $M, N$)—then $\text{Cov} \left( \mathbb{E} \left[ G_I | T_{γ^* I} \right], \mathbb{E} \left[ S_I | T_{γ^* I} \right] \right) < 0$. Since negatively correlated Gaussian random variables are NQD (Lehmann 1996, p.1139), we may apply Proposition 4 to obtain the following result.

**Corollary 2** Suppose the conditions of Proposition 11 hold and $\text{Cov}(G_I, S_I) \leq 0$. If adverse selection is induced efficiently, then the training firm $I$ is a low wage firm.

**Proof.** See preceding text. ■

Corollary 2 is intuitive. Competition implies that only the training firm makes profits. Consequently, the efficiency contribution $\mathbb{E}[EC_{γ I}]$ determines the amount of surplus to be distributed between the training firm and the worker. Since the quantity of adverse selection $\mathbb{E}[AS_{γ I}]$ determines the expected employment wage of the worker, it also determines the distribution of this surplus between the training firm and worker. Firm profits are maximized by inducing adverse selection efficiently. Applying Propositions 4 and 11, the way to do this is to impose adverse selection on those workers that it is efficient to keep and have higher wages for those workers who are more productively employed elsewhere.

In the specification of Example 1, if the training firm has maximal private information, then $\text{Cov}(G_I, S_I) = 0$. Thus, applying Corollary 2, the training firm will be a low wage firm. Recall from Section 4 that this implies that adverse selection in the aggregated market will underestimate expected adverse selection. Effect reversals are therefore a possibility, implying that the standard test (Gibbons and Katz 1991) is no longer a necessary condition for adverse selection. Of course, this raises the question as to whether it is actually profitable for the training firm to acquire maximal private information.

6.3 Information Acquisition

In this section, we explore how the quality of private information affects the trade-off between adverse selection and efficiency. Hence, we are interested in the incentives for firms’ to acquire information. We begin by establishing that it is not always possible to ‘neutralize’ the effect of acquiring more private information via a public information disclosure. Notwithstanding this result, we are still able to show that information acquisition expands the expected adverse selection-efficiency frontier and hence (absent significant acquisition costs) could be profitable for the training firm.

---

19 To see this note that, using $R_I = G_I + S_I$, we can rewrite the expression in the Proposition as $G_I = \frac{\beta_{G_I R_I T_{γ^* I}}}{\beta_{G_I R_I T_{γ^* I}} + S_I} + \alpha \cdot T_{γ^* I}$, where to satisfy regression to the mean, $\beta_{G_I R_I T_{γ^* I}} \in [0, 1)$.
The Difficulty of Neutralizing Private Information  It should be obvious that the effect of acquiring more inside information on expected adverse selection and efficiency will depend on disclosure, i.e. whether this information is passed into the public domain. A more subtle point is that acquiring more inside information and disclosing it will typically result in a different combination of expected adverse selection and efficiency relative to not acquiring the information in the first place. To see this, it is helpful to return to our concrete example.

Example 1 (continued) Suppose that under information structure $\gamma_I$, $Q_{\gamma_I} = Y_I$ and $T_{\gamma_I} = \emptyset$ (i.e. large $M$, $N = n = 0$), and under information structure $\gamma'_{I}$, $Q_{\gamma'_{I}} = (Y_I, X_I)$, and $T_{\gamma'_{I}} = X_I$ (i.e. large $M$, large $N = n$). The change from information structure $\gamma_I$ to $\gamma'_{I}$ therefore corresponds to the situation where the training firm learns, and discloses, the worker’s general ability; i.e. it passes all of its extra private information into the public domain. Under information structure $\gamma'_{I}$, there is full efficiency and no adverse selection. In contrast under information structure $\gamma_I$, there is positive adverse selection but not full efficiency (even relative to the worse information).

In light of this example, it is natural to ask whether there is some (other) disclosure policy that will neutralize the acquisition of more inside information.

Remark 4 In the Gaussian case, suppose $Q'_{I}$ is more informative than $Q_I$. The following statements are equivalent.

1. For any $\gamma_I \in \Gamma_{Q_I}$, there is a $\gamma'_{I} \in \Gamma_{Q'_{I}}$ such that $(AS_{\gamma'_{I}}, EC_{\gamma'_{I}})$ is equal in distribution to $(AS_{\gamma_{I}}, EC_{\gamma_{I}})$.

2. The improvement in information leaves unchanged the correlation between $R_I$ and $G_I$, that is $\sigma_{R_I, S_I} = \sigma_{R'_{I}, S'_{I}}$.

Proof. Appendix D.

The condition in Remark 4 is strong. Typically, acquisition will improve information about match quality, and it will not be possible to neutralize the effects of more private information simply by passing it on. This difficulty arises because of the parameter:

$$\sigma^2_E[S_{\gamma_I} | T_{\gamma_I}] = \sigma^2_E[S_{\gamma_I} | T_{\gamma_I}] - \sigma^2_{S_{\gamma_I} | T_{\gamma_I}}.$$ 

To retain this at its previous level when $\Sigma_{(R_I, G_I)}$ increases to $\Sigma_{(R'_{I}, G'_{I})}$, $\sigma^2_{G_{\gamma_I} | T_{\gamma_I}}$ must also increase by the same amount.

Information Acquisition Increases Efficiency If it were possible to neutralize private information via disclosure, then establishing that information acquisition (weakly) increases efficiency would be trivial. Notwithstanding Remark 4, we can use the results of Section 5 via Corollary 1 to establish that private information acquisition can expand the adverse selection-efficiency frontier.
Proposition 12  With free disclosure of information, information acquisition increases efficiency. Specifically, $V(\Sigma(R_I,G_I),AS)$ is increasing in $\Sigma(R_I,G_I)$ in positive semidefinite order in the region $\beta_{G_I R_I} \leq \pi / T$.

Proof. Appendix D.

Proposition 12 establishes that, if $\beta_{G_I R_I}$ is not too large, private information acquisition expands the adverse selection-efficiency frontier. Hence, absent significant acquisition costs and given sufficiently flexible disclosure possibilities, acquiring maximal information could be profitable for the training firm. We pursue this logic in the following section where we return to discussion of the application of these results to the employer learning application.

As with Proposition 11 the gist of the proof follows easily given Corollary 1. The constraints $\Sigma(R_I,G_I)|T_{\gamma_I} \leq \Sigma(R_I,G_I)$ and $0 \leq \Sigma(R_I,G_I)|T_{\gamma_I}$ boil down respectively to a cone satisfying the conditions of Corollary 1 and a convex elliptical set containing the vertex of the cone. Improving private information for fixed $T_{\gamma_I}$, expands the elliptical set in the sense of set inclusion, but generally shifts the vertex of the conical set. So the intersection of the two constraint sets are not generally ordered by set inclusion as private information improves (Remark 4 already implies this). Notwithstanding this, it is very easy to show that the set of attainable $\beta_{G_I R_I T_{\gamma_I}}$ under better private information contains the set of attainable $\beta_{G_I R_I T_{\gamma_I}}$ under worse private information. Given this fact, Corollary 1 is easily seen to imply that from any point in the intersection of constraint sets with worse private information, there is a path along which which $\mathbb{E}[AS_{\gamma_I}]$ and $\mathbb{E}[EC_{\gamma_I}]$ are both increasing and which passes into the intersection of the constraint sets for better private information. Hence, the conclusion of the proposition is established.

6.4 Endogenous Information and its Consequences for Wages and Tests for Adverse Selection

We conclude our analysis by showing how Propositions 11 and 12 can be used to characterize equilibrium training contracts when firms face a non-trivial choice of information structure. We begin by considering the case where each firm has full control of disclosure; that is, for each firm $I \in F$, $\Gamma_I$ is identified one to one with the set of symmetric positive semidefinite covariance matrices partitioned as in (13). We then explore the effects of constraints on disclosure. To enable us to state the results in a simple fashion, in both cases we introduce the following assumption.

A6. One firm is skill enhancing and every other firm in $F$ is competitive. For each pair of competitive firms, $J, J' \in F$, and for each $K \in F$, the random pairs $(Q_{\gamma_J}, Y_{J+K})$ and $(Q_{\gamma_{J'}}, Y_{J'+K})$ have identical distributions. For the skill enhancing firm $I$, for each $K \in F$, and competitive firms $J \in F$, $(Q_{\gamma_J}, Y_{IK})$ has the same distribution as $(Q_{\gamma_J}, Y_{JK} + \Delta)$ for some $\Delta > 0$. 

35
Thus it is known that workers trained at the skill enhancing firm are $\Delta$ more valuable than competitive firms. This is the only difference between firms; i.e. we do not assume that firms are ex ante identical but make the next simplest assumption which is to distinguish between one skill enhancing firm and a fringe of identical competitive firms.

In any PBE, the contract offered by the training firm $I$ must maximize expected profit (from training and employment) subject to the worker’s participation constraint and credit constraint. Since we can write expected employment profit, $E[(R_{\gamma_I} - W_{\gamma_I})^+]$, as the sum of expected adverse selection and the expected efficiency contribution, the contract offered by any firm $I \in F$ must therefore solve

\[
\begin{align*}
\text{Maximize} & \quad E[Y_I] - w^{1st}_I + E[EC_{\gamma_I}] + E[AS_{\gamma_I}] \\
\text{subject to} & \quad w^{1st}_I + E[W_{\gamma_I}] \geq \max_{J \in F, J \neq I} \{E[TS_{\gamma_J}]\}.
\end{align*}
\]

It follows that an information structure is optimal for firm $I$ only if it solves this program. We will say that $\gamma_I \in \Gamma_I$ has maximal private information acquisition if, for $\gamma_I, \gamma'_I \in \Gamma_I$, $\Sigma E[Y|Q_{\gamma'_I}] \geq \Sigma E[Y|Q_{\gamma_I}]$ implies $\Sigma E[Y|Q_{\gamma'_I}] = \Sigma E[Y|Q_{\gamma_I}]$. Appealing to Propositions 11 and 12, we have the following result.

**Corollary 3** In the Gaussian case where each firm $I \in F$ has full control of disclosure, there exists a PBE in which:

1. Each firm $I \in F$ chooses an information structure $\gamma^*_I$ with maximal private information acquisition.

2. For each competitive firm public information disclosure is $T_{\gamma^*_I} = G_{\gamma^*_I}$.

3. For the skill enhancing firm,

   (a) If $\Delta$ is sufficiently large, public information disclosure satisfies Conditions 1. and 2. of Proposition 11.

   (b) If $\Delta$ is small, public information disclosure is $T_{\gamma^*_I} = G_{\gamma^*_I}$.

This result can be related to our earlier discussion of low wage firms and tests for adverse selection. In equilibrium, a sufficiently skill enhancing firm will induce positive adverse selection via efficient disclosure and (as discussed after Corollary 2) become a low wage firm. The possibility of an effect reversal therefore arises an equilibrium phenomenon. In short, adverse selection endogenously creates the forces by which it hides itself.

Corollary 3 describes the situation when firms have complete (subject to the maintained Gaussian framework) freedom in deciding how much of what they will learn about their worker enters the public domain. This is a natural benchmark case, but our model allows investigation
of other interesting cases. We conclude this section by supplying an example in which firms face constraints on disclosure and, as a result, it is not optimal to acquire maximal information.

Example 1 (continued for the case with constrained disclosure)  Recall, an information structure is characterized by three numbers: $M$, $N$, and $n$. To capture the idea that disclosure possibilities are circumscribed we assume that $n$ are fixed, and to make the idea of maximal information acquisition concrete we assume that $M$ and $N$ are bounded, respectively, by $\bar{M}$ and $\bar{N}$. Since it is important that some information is disclosed, we set $n = 1$. In this setting, there exists a PBE in which the skill enhancing and competitive firms take the following choices.

- Competitive firms acquire no further information about general ability ($M = \bar{M}$ and $N = 1$).
  Competitive firms choose their information structure to maximize the efficiency contribution. We saw in Section 5 that, with $n$ and $M$ fixed, increasing $N$ decreases the efficiency contribution and so, when disclosure opportunities are constrained, competitive firms collect the minimum number of measurements of general ability. On the other hand, we also saw in Section 5 that with $m$ and $N$ fixed, increasing $M$ increases the efficiency contribution and so competitive firms continue to collect the maximum number of measurements of the worker’s productivity during training.

- For $\Delta$ sufficiently large, the skill enhancing firm acquires only information about general ability ($M = 0$ and $N = \bar{N}$).
  When $\Delta$ is greater than the equilibrium training wage offered by competitive firms, the skill enhancing firm has an incentive to choose an information structure that induces (more) adverse selection to claw back rent from the worker. Initially, as $\Delta$ increases above the competitive training wage, the skill enhancing firm collects more samples of general ability. This raises informed party relevance and the information gap and, as a result, both adverse selection and profit increase. Once $N = \bar{N}$, the skill enhancing firm can only induce further increases in adverse selection by collecting fewer samples of the worker’s productivity during training. Although this decreases efficiency, profit continues to increase. For $\Delta$ sufficiently large, the skill augmenting firm collects the minimum number of samples of the worker’s productivity during training and the maximum number of samples of general ability.

7  Conclusion

This paper has developed a framework for the analysis of how asymmetric information impacts on adverse selection and market efficiency. We adopted Akerlof’s (1970) unit-demand model
extended to a setting in which valuations of market participants are not all increasing functions of some underlying scalar type and there is multidimensional public and private information.

We identified two conditions, orthogonality and regression to the mean, as key to making the analysis tractable. Under these conditions, adverse selection depends on multidimensional public information via a scalar statistic, the apparent match quality. Adverse selection is an increasing function of the apparent match quality, while both the probability of trade and the efficiency contribution are decreasing functions of the apparent match quality. We showed that the relationship between the apparent match quality and another scalar statistic, the apparent employment value to outside firms, has important consequences. In particular, if these statistics are negatively quadrant dependent (Lehmann 1966) then released workers may be paid more than retained workers (the training firm is a low wage firm) and, absent the econometrician observing public information, standard positive correlation tests for adverse selection lack power.

Adapting our framework to the Gaussian case, we also made progress in understanding how expected efficiency depends on the information structure, and how this relates to expected adverse selection. We showed that information structures could be represented in a three-dimensional parameter space, and how expected adverse selection and expected efficiency are related within this space. Using these insights, we characterized endogenous information structures in our employer learning setting. In equilibrium, the training firm’s estimates of its own employment value for the worker and the worker’s employment value in an outside firm are conditionally colinear given public information. As previously established, it is in precisely these cases that the power of standard positive correlation tests is reduced making it harder for the econometrician, when less informed than market participants, to test for and find adverse selection.

References


[23] Li, J. (2011) “Job Mobility, Wage Dispersion and Asymmetric Information”, mimeo Kellogg School of Management.


Remark 1  It follows from Assumption A4 that regression to the mean implies $w \mapsto E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} < w]$ is increasing and $w \mapsto w - E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} < w]$ is strictly increasing.

Proof of Remark 1. It suffices to show that for each $t \in T_{\gamma_1}$, for all $w \in \mathbb{R}$, $0 \leq \frac{\partial}{\partial w} E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} \leq w]$. Note that

$$
\frac{\partial}{\partial w} E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} \leq w] = \left( \int_{R}^w \left( E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} = w] - E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} = r] \right) f_{R_{\gamma_1}|T_{\gamma_1}}(r|t)dr \right) \frac{f_{R_{\gamma_1}|T_{\gamma_1}}(w|t)}{F_{R_{\gamma_1}|T_{\gamma_1}}(w|t)}.
$$

Regression to the mean implies that for $r \leq w$, $0 \leq E[G_{I}|T = t, R_{I} = w] - E[G_{I}|T = t, R_{I} = r] < w - r$. Hence,

$$
0 \leq \frac{\partial}{\partial w} E[G_{\gamma_1}|T_{\gamma_1} = t, R_{\gamma_1} \leq w] \leq \left( \int_{R}^w \frac{(w - r) f_{R_{\gamma_1}|T_{\gamma_1}}(r|t)dr}{F_{R_{\gamma_1}|T_{\gamma_1}}(w|t)dr} \right) \frac{f_{R_{\gamma_1}|T_{\gamma_1}}(w|t)}{F_{R_{\gamma_1}|T_{\gamma_1}}(w|t)} = \int_{R}^w \frac{F_{R_{\gamma_1}|T_{\gamma_1}}(r|t)dr}{F_{R_{\gamma_1}|T_{\gamma_1}}(w|t)} \frac{f_{R_{\gamma_1}|T_{\gamma_1}}(w|t)}{F_{R_{\gamma_1}|T_{\gamma_1}}(w|t)} \leq 1.
$$

The equality following from integration by parts and the final inequality following directly from the logconcavity assumption.

The willingness to pay map

Define the distance $d : C_{\gamma_1} \times C_{\gamma_1} \rightarrow [-\infty, \infty]$, $d(\phi, \varphi) = \sup_{t \in T_{\gamma_1}} |\phi(t) - \varphi(t)|$. It is classical that the set of bounded functions mapping from an arbitrary set into the reals is a complete metric space when endowed with the sup norm (see e.g. Dunford and Schwartz p.258). If $G_{\gamma_1}$ is compact, therefore, $M_{\gamma_1} = (C_{\gamma_1}, d)$ is a complete metric space. In the Gaussian case, apart from the trivial case where $Q_{\gamma_1}$ is orthogonal to $Y_{II}$, $G_{\gamma_1}$ is not compact. Evidently, given some $\phi_{\gamma_1}^* \in C_{\gamma_1}$ if we denote by $C_{\gamma_1}^* \subset C_{\gamma_1}$ the set of $\phi \in C_{\gamma_1}$ such that $\phi - \phi_{\gamma_1}^*$ is bounded, $(C_{\gamma_1}^*, d)$ is a complete metric space. Specifically, Lemma A1 below takes

$$
\phi_{\gamma_1}^*(t) = E[G_{\gamma_1}|T_{\gamma_1} = t] - \frac{\beta_{G_{\gamma_1} R_{\gamma_1} T_{\gamma_1}}}{1 - \gamma_{G_{\gamma_1} R_{\gamma_1} T_{\gamma_1}}} \max\{E[S_{\gamma_1}|T_{\gamma_1} = t], 0\}.
$$

Lemma A1  Suppose $\gamma_1$ induces regression to the mean.

1. If A4 holds and $G_{\gamma_1}$ is compact for each $I \in F$, then the map $\Psi_{\gamma_1} : C_{\gamma_1} \rightarrow C_{\gamma_1}$, is contractive.\(^{20}\) $\Psi_{\gamma_1}$ has a unique fixed point in $C_{\gamma_1}$.

\(^{20}\)The terminology in the literature is not entirely consistent, we say a map $\Psi$ from a metric space $M = (C, d)$ to itself is contractive if for each $\phi, \phi' \in C$, $d(\Psi(\phi), \Psi(\phi')) < d(\phi, \phi')$. It is a contraction if there exists a Lipschitz constant $-1 < \beta < 1$ such that for each $\phi, \phi' \in C$, $d(\Psi(\phi), \Psi(\phi')) \leq \beta d(\phi, \phi')$. It is non-expansive if for each $\phi, \phi' \in C$ $d(\Psi(\phi), \Psi(\phi')) \leq d(\phi, \phi')$. 

41
2. In the Gaussian case, $\Psi_{\gamma_I}$ is a contraction mapping on $(C_{\gamma_I}^*, d)$, $\Psi_{\gamma_I}$ has a unique fixed point in $C_{\gamma_I}^*$.

**Proof.** Part 1. Remark 1 asserts that $\varphi(t) > \phi(t)$ implies

$$\varphi(t) - \phi(t) > \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} \leq \varphi(t)] - \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} \leq \phi(t)] \geq 0.$$ 

It follows that for any $\varphi, \phi \in C_{\gamma_I}$, and each $t \in T_{\gamma_I}$, $|\varphi(t) - \phi(t)| \geq |\Psi_{\gamma_I}(\varphi) - \Psi_{\gamma_I}(\phi)|$. Hence, $d\left(\Psi_{\gamma_I}(\phi), \Psi_{\gamma_I}(\varphi)\right) < d(\varphi, \phi)$ on $C_{\gamma_I}$. Existence follows from a straightforward application of Brouwer’s fixed point theorem, given compactness of $G_{\gamma_I}$ and continuity of $w \mapsto \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t, R_{\gamma_I} \leq w]$ for each $t \in T_{\gamma_I}$. Uniqueness follows since if $\phi^*, \phi^{**}$ are distinct fixed points, $d\left(\Psi_{\gamma_I}(\phi^*), \Psi_{\gamma_I}(\phi^{**})\right) < d(\phi^*, \phi^{**})$ contradicting the fixed point property $\Psi_{\gamma_I}(\phi^*) = \phi^*$, $\Psi_{\gamma_I}(\phi^{**}) = \phi^{**}$.

**Part 2.** The map $\Psi_{\gamma_I}$ is defined by

$$\Psi_{\gamma_I}(\phi)(t) = \mathbb{E}[G_{\gamma_I}|T_{\gamma_I} = t] - \beta_{G_{\gamma_I}} R_{\gamma_I} \Phi_{\gamma_I} \sigma_{R_{\gamma_I}}|T_{\gamma_I} \frac{h}{\sigma_{R_{\gamma_I}}|T_{\gamma_I}} \left( \frac{\mathbb{E}[R_{\gamma_I}|T_{\gamma_I} = t] - \phi(t)}{\sigma_{R_{\gamma_I}}|T_{\gamma_I}} \right),$$

where $h$ is the Gaussian $\mathcal{N}(0,1)$ hazard function. Confirmation of this fact appears in Section 4. The hazard is continuously differentiable and satisfies $0 < h' < 1$, therefore it follows immediately from the mean value theorem that $|\Psi_{\gamma_I}(\phi)(t) - \Psi_{\gamma_I}(\varphi)(t)| \leq \beta_{G_{\gamma_I}} R_{\gamma_I} \Phi_{\gamma_I} \sigma_{R_{\gamma_I}}|T_{\gamma_I} \frac{h}{\sigma_{R_{\gamma_I}}|T_{\gamma_I}} \left( \frac{\mathbb{E}[R_{\gamma_I}|T_{\gamma_I} = t] - \phi(t)}{\sigma_{R_{\gamma_I}}|T_{\gamma_I}} \right)$

Proposition 1 Assume A1-A4. Let information be exogenous, i.e. $\Gamma_J = \{\gamma_J\}$, for each $J \in F$. Suppose either information structures are Gaussian, or for each $J \in F$, $G_{\gamma_J}$ is compact. Suppose that $\gamma_J$ induces regression to the mean. It follows that, for each $J \in F$, the willingness to pay map has a unique fixed point $w_{\gamma_J} : T_{\gamma_J} \to \mathbb{R}$. Choose $I \in F$ such that $\mathbb{E}[TS_{\gamma_I}] \geq \mathbb{E}[TS_{\gamma_J}], J \in F$. There exists a PBE in which:

1. In period 1, each firm $J \in F, J \neq I$ offers a training contract $\tau_J = \left(\mathbb{E}[TS_{\gamma_J}] - \mathbb{E}[W_{\gamma_J}], \gamma_J\right)$. Firm $I$ offers a training contract $\tau_I = \left(\max_{J \in F, J \neq I} \left(\mathbb{E}[TS_{\gamma_J}] - \mathbb{E}[W_{\gamma_J}]\right)^+ \gamma_I\right)$. The worker chooses to train at firm $I$.

2. In period 2, each outside firm $J \in F, J \neq I$ makes the employment wage offer $W_{\gamma_J} = w_{\gamma_J}(T_{\gamma_J})$. The training firm $I$ offers the same wage as outside firms if $R_{\gamma_I} \geq W_{\gamma_J}$ but a lower wage if $R_{\gamma_J} < W_{\gamma_J}$. The worker remains at the training firm $I$ if its offer matches $W_{\gamma_I}$ and otherwise moves to an outside firm chosen at random.

**Proof.** We start by confirming that there is no incentive to deviate from second period strategies. Clearly, the worker quits only if the training firm $I$ fails to match the best outside
offer. Anticipating this, the training firm $I$ releases the worker (equivalently fails to match the best outside offer) if and only if $R_{\gamma_I} < W_{\gamma_I}$. It follows that it cannot be optimal for outside firms to deviate from posting an offer that is a fixed point of the willingness to pay map in Definition 4. To see this, take the first case in the statement of the proposition and suppose $\mathbb{E}[Y_{IJ}|T_{\gamma_I} = t, R_{\gamma_I} < w_{\gamma_I}(t)] \neq w_{\gamma_I}(t)$ for some $t \in T_{\gamma_I}$ with $w_{\gamma_I}(t) \geq \inf R_{\gamma_I}(t)$. Here, either the successful bidder will make an expected loss, or there is a wage bid that is a profitable deviation by some firm. Turning to the second case, suppose $\mathbb{E}[Y_{IJ}|T_{\gamma_I} = t, R_{\gamma_I} < w_{\gamma_I}(t)] \neq \mathbb{E}[Y_{IJ}|T_{\gamma_I} = t, R_{\gamma_I} = \inf R_{\gamma_I}(t)]$ for some $t \in T_{\gamma_I}$ with $w_{\gamma_I}(t) < \inf R_{\gamma_I}(t)$. Here, the outside bidder fails to hire the worker with positive probability and so there is nothing to be gained from deviating.

We now argue that there is no incentive to deviate from first period strategies (given second period strategies). The worker has no incentive to choose an alternative first period employer because training at firm $I$ generates the (weakly) highest expected lifetime utility $w^{1st}_I + \mathbb{E}[W_{\gamma_I}] = \max_{J \in F, J \neq I} \{\mathbb{E}[T_S_{\gamma_J}]\} = w^{1st}_I + \mathbb{E}[W_{\gamma_I}]$. In equilibrium, each firm $J \in F, J \neq I$ fails to hire the worker and makes zero profit in the first period (and expects to make zero profit in the second period). Clearly, there is nothing to be gained from deviating to a lower training wage. A higher training wage would entail offering the worker an expected lifetime utility in excess of $\mathbb{E}[T_S_{\gamma_J}]$ and so cannot be profitable deviation. In equilibrium, firm $I$ makes a weakly positive expected profit. If firm $I$ deviates to a lower training wage, it would fail to hire the worker and so would make zero profit in the first period (and expect to make zero profit in the second period). If firm $I$ deviates to a higher training wage, it would still hire the worker but for a lower expected profit. 

**Discussion of equilibrium selection**

If outside firms always expect to make a loss from hiring the worker, then any bid below the lowest possible retained value $\inf R_{\gamma_I}(t)$ can be part of an equilibrium. This is because trade occurs with zero probability and so outside firms expect to make zero profit given any bid $w < \inf R_{\gamma_I}(t)$. In this case, we may have an equilibrium wage outcome that is not a fixed point of the willingness to pay map. Relatively little in the paper depends on this but such equilibrium wage outcomes are an inconvenience that we prefer to avoid. In what follows, we focus attention on the case where equilibrium wage outcomes are determined by the wage schedule $w_{\gamma_I}$, the unique fixed point of $\Psi_{\gamma_I}$. Here, we present a refinement based on the following idea which may be offered as justification for this focus. Suppose that at the time of wage setting, the current employer faces a cost (which may be negative) $\sigma \varepsilon$ for releasing the worker with $\varepsilon \sim N[0, 1]$, so $R_{\gamma_I} - \sigma \varepsilon < w$ leads to release of the worker. Since $\sigma \varepsilon$ has infinite support, worker release is never off the equilibrium path. As $\sigma \to 0$, the conditional distribution of $R_{\gamma_I}$ given $T_{\gamma_I} = t$ and the event $R_{\gamma_I, \sigma} = R_{\gamma_I} - \sigma \varepsilon < w$ converges in distribution to the random variable degenerate at $\inf R_{\gamma_I}(t)$. Lemma A2 contains a formal statement and proof of this assertion. It follows
from this fact, given \( r \mapsto \mathbb{E}[G_{\gamma_1} | T_{\gamma_1} = t, R_{\gamma_1} = r] \) is continuous on \( R_{\gamma_1}(t) \), that for any \( w < \inf R_{\gamma_1} R(t), \lim_{t \to 0} \mathbb{E}[G_{\gamma_1} | T_{\gamma_1} = t, R_{\gamma_1} < w] = \mathbb{E}[G_{\gamma_1} | T_{\gamma_1} = t, R_{\gamma_1} = \inf R_{\gamma_1} R(t)] \).

Notwithstanding this refinement, since there are many firms in our model, there are inevitably a large number of Perfect Bayesian Equilibria.21 However, the substance of the above discussion is that, given regression to the mean, an equilibrium wage schedule exists and given our refinement is unique.

**Lemma A2** Let \( R_{\gamma_1} \) have compact support \( R_{\gamma_1} \), cdf \( F : R_{\gamma_1} \to [0,1] \) absolutely continuous with respect to Lebesgue measure, and density \( f \). Let \( R_{\gamma_1,\sigma} = R_{\gamma_1} + \sigma \varepsilon \), with \( \sigma \in \mathbb{R}_+ \), \( R_{\gamma_1} \) and \( \varepsilon \) independently distributed with \( \varepsilon \sim N[0,1] \). For any \( w < \inf R_{\gamma_1} \), the conditional distribution of \( R_{\gamma_1} \) given the event \( R_{\gamma_1,\sigma} \leq w \) converges in probability to the distribution degenerate at \( \inf R_{\gamma_1} \).

**Proof.** Denote the unit normal cdf as \( \Phi \). Denote \( \inf R_{\gamma_1} = a \), \( \text{supp} R_{\gamma_1} = b \). We need to show that for each \( x > a > w \), the cdf of the conditional distribution converges to 1. That is, for each \( 0 < \delta < 1 \), \( x > a > w \), there is an \( \sigma' \) such that for all \( \sigma < \sigma' \)

\[
\Pr[R_{\gamma_1} \leq x \mid R_{\gamma_1,\sigma} \leq w] = \frac{\int_a^x f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr}{\int_a^b f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr} > 1 - \delta.
\]

Equivalently,

\[
\delta \int_a^x f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr > (1 - \delta) \int_x^b f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr
\]

\[= (1 - \delta) \left( \int_x^b f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr + \int_y^b f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr \right).\]

Since \( \Phi\left(\frac{w-r}{\sigma}\right) \) is decreasing in \( r \), it suffices to show that for some \( y \in (x,b) \)

\[
\delta \int_a^x f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr > (1 - \delta) \left( \int_x^y f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr + \int_y^b f(r) \Phi\left(\frac{w-r}{\sigma}\right) \, dr \right).
\]

That is,

\[
\delta F(x) > (1 - \delta) \left( F(y) - F(x) + \frac{\Phi\left(\frac{w-y}{\sigma}\right)}{\Phi\left(\frac{w-x}{\sigma}\right)} (1 - F(y)) \right)
\]

\[
F(x) > (1 - \delta) \left( F(y) + \frac{\Phi\left(\frac{w-y}{\sigma}\right)}{\Phi\left(\frac{w-x}{\sigma}\right)} (1 - F(y)) \right).
\]

Choosing \( y \in (x,b) \) such that \( F(x) > (1 - \delta) F(y) \), possible by continuity, this establishes the

---

21For instance, providing that two firms \( J, J' \in F, J \neq I \) bid as stated in Proposition 1, other firms can post arbitrary (lower) training contract and employment wage offers.
result on noting that \((1 - F(y))\) is evidently bounded, \(\Phi(\frac{w-y}{\sigma}) < \frac{1}{\sqrt{2\pi}} \left(1 + \frac{(w-y)^2}{2\sigma^2}\right)^{-1}\) by log concavity of \(\Phi\) and \(e^{-\frac{1}{2} \left(\frac{w-y}{\sigma}\right)^2}\) converges monotonically to zero as \(\sigma\) tends to zero. Hence, as required, there is some \(\sigma'\) such that \(\Phi(\frac{w-y}{\sigma'}) < \frac{F(x) - F(y)(1 - \delta)}{(1 - \delta)(F(R) - F(y))}\) for all \(\sigma < \sigma'\).

Appendix B: Material Omitted from Section 4

**Lemma B1** Suppose the orthogonality condition (8) holds. Suppose \(\gamma_I \in \Gamma_I\) induces regression to the mean and \(z \leq z'\), then \(E[G_{\gamma_I} | z < R_{\gamma_I}] \leq E[G_{\gamma_I} | z' < R_{\gamma_I}]\). Hence, \(0 \leq E[G_{\gamma_I} | z < R_{\gamma_I}]\).

**Proof.** By the law of iterated expectations

\[
E[G_{\gamma_I} | z \leq R_{\gamma_I}] = E \left[ E \left[ G_{\gamma_I} | T_{\gamma_I} = t \right] - E \left[ G_{\gamma_I} | T_{\gamma_I} = t, R_{\gamma_I} \right] | z \leq R_{\gamma_I} \right].
\]

Hence, by the law of iterated expectations again,

\[
E[G_{\gamma_I} | z \leq R_{\gamma_I}] = \int E \left[ E \left[ G_{\gamma_I} | T_{\gamma_I} = t \right] - E \left[ G_{\gamma_I} | T_{\gamma_I} = t, R_{\gamma_I} \right] | T_{\gamma_I} = t, z \leq R_{\gamma_I} \right] dF_{T_{\gamma_I}} | z \leq R_{\gamma_I}(t).
\]

Where, \(F_{T_{\gamma_I}} | z \leq R_{\gamma_I}\) is the conditional cdf of \(T_{\gamma_I}\) given the event \(z \leq R_{\gamma_I}\). By the orthogonality condition, \(F_{T_{\gamma_I}} | z \leq R_{\gamma_I} = F_{T_{\gamma_I}}\), the unconditional cdf. Hence,

\[
E[G_{\gamma_I} | z \leq R_{\gamma_I}] = \int E \left[ E \left[ G_{\gamma_I} | T_{\gamma_I} = t \right] - E \left[ G_{\gamma_I} | T_{\gamma_I} = t, R_{\gamma_I} \right] | T_{\gamma_I} = t, z \leq R_{\gamma_I} \right] dF_{T_{\gamma_I}}(t).
\]

The result now follows from the fact that, given regression to the mean, for each fixed \(t \in T_{\gamma_I}\) the integrand is increasing in \(z\).

**Brief Notes on Statistical Dependence Concepts**

The pair of scalar random variables \((X_1, X_2)\) are said to be *positively quadrant dependent* (PQD) Lehmann (1966) if for all \(x_1, x_2\) in the support of \((X_1, X_2)\)

\[
Pr[X_1 \leq x_1, X_2 \leq x_2] \geq Pr[X_1 \leq x_1] Pr[X_2 \leq x_2].
\]

If the inequality is reversed \((X_1, X_2)\) is said to be *negatively quadrant dependant* (NQD). If \((X, Y)\) are PQD, then for any non-decreasing functions \(\varphi_1, \varphi_2 (\varphi_1(X_1), \varphi_2(X_2))\) is PQD. (Lehmann 1966, Lemma 1). Lehmann Lemma 3 establishes if \((X_1, X_2)\) are PQD, for all non-decreasing functions \(\varphi_1, \varphi_2\) such that the expectations exist

\[
E[\varphi_1(X_1)\varphi_2(X_2)] \geq E[\varphi_1(X_1)]E[\varphi_2(X_2)].
\]
\( \varphi_1(X_1) \) and \( \varphi_2(X_2) \) have a positive covariance. This condition is an equivalence so may serve as an alternative definition of PQD.

Applying Bayes rule, PQD evidently implies \( \Pr[X_1 \leq x_1 | X_2 \leq x_2] \geq \Pr[X_1 \leq x_1] \) which in turn implies, if the expectations exist, \( \mathbb{E}[X_1 | X_2 \leq x_2] \leq \mathbb{E}[X_1] \). Kowalczyk and Pleszcynska (1977) say \((X_1, X_2)\) is EQD\(^+\) if for all \( x_2 \) in the support of \( X_2 \),

\[
\mathbb{E}[X_1 | X_2 \leq x_2] \leq \mathbb{E}[X_1].
\]

If the inequality is reversed, then Kowalczyk and Pleszcynska (1977) say \((X_1, X_2)\) are EQD\(^-\).

Let the vector of random variables \( X = (X_1, ..., X_n) \) have a density with respect to Lebesgue measure \( f \), \( X \) is said to be affiliated if \( f(x \vee x')f(x \wedge x') \geq f(x)f(x') \) a.e. on \( \mathbb{R}^n \). Equivalently, for each pair of nondecreasing functions and sublattice \( S \), \( \mathbb{E}[\varphi_1(X)\varphi_2(X)|S] \geq \mathbb{E}[\varphi_1(X)|S]\mathbb{E}[\varphi_2(X)|S] \).

It is known that: \((X_1, X_2)\) affiliated \(\Rightarrow\) \((X_1, X_2)\) PQD \(\Rightarrow\) \((X_1, X_2)\) EQD\(^+\).

**Proposition 2** In the equilibrium identified in Proposition 1, suppose the orthogonality condition (8) holds, then adverse selection, \( AS_{g_1} \), and the apparent match quality, \( \mathbb{E}[S_{g_1}|T_{g_1}] \), are comonotone random variables. Specifically, there exists an increasing function \( \tilde{a}s_{g_1} : \mathbb{R} \rightarrow \mathbb{R}_+ \) such that \( AS_{g_1} = a_{g_1}(T_{g_1}) = \tilde{a}s_{g_1}(\mathbb{E}[S_{g_1}|T_{g_1}]) \).

**Proof.** (continued) It remains to confirm positivity and monotonicity. Lemma A1 implies that adverse selection at \( T_{g_1} = t \), \( a_{g_1}(t) \), is uniquely defined as the limit of the convergent sequence \( a_0, a_1, a_2, ... \) with \( a_{i+1} = \mathbb{E}[G_{g_1}|a_i + \mathbb{E}[S_{g_1}|T_{g_1} = t] < R_{g_1}] \). \( a_{g_1}(t') \) is defined by a similarly defined convergent sequence \( a_0, a'_1, a'_2, ... \). Suppose that \( \mathbb{E}[S_{g_1}|T_{g_1} = t'] > \mathbb{E}[S_{g_1}|T_{g_1} = t] \).

It follows immediately from the positive dependence of \((G_{g_1}, R_{g_1})\) (Lemma B1 in the appendix) that \( a'_1 = \mathbb{E}[G_{g_1}|a_0 + \mathbb{E}[S_{g_1}|T_{g_1} = t'] < R_{g_1}] \geq a_1 \), hence \( a'_2 \geq a'_1 \) and \( a'_i \geq a_i \) for all \( i \), therefore \( a_{g_1}(t') \geq a_{g_1}(t) \). Positivity follows directly from the last statement in Lemma B1.

**Appendix C: Material Omitted from Section 5**

**Properties of the normal hazard function**

It is convenient to collect here some mainly familiar properties of the normal hazard function (inverse Mills ratio).

**Lemma C1** The normal hazard function \( h : \mathbb{R} \rightarrow \mathbb{R}_+ \), satisfies:

1. Level bounds. \( |x|^+ < h(x) < \sqrt{\frac{x^2}{2}} \) \(< |x + \frac{1}{2}| \) on \( R \). This implies \( \lim_{x \rightarrow \infty} \frac{h(x)}{x} = \lim_{x \rightarrow -\infty} h'(x) = 1 \).

2. Gradient bounds. \( x 
\mapsto h(x) \) and \( x \mapsto x - h(x) \) are both strictly increasing on \( \mathbb{R} \). \( 0 < h' < 1 \) on \( \mathbb{R} \).
3. Convexity bounds. $h$ is strictly convex and log concave on $\mathbb{R}$. $0 < \frac{d}{dx} \frac{h(x)}{h'(x)} < 1$.

4. Starshaped. $x \mapsto h(x)$ and $x \mapsto \frac{h(x)}{h'(x)}$ are both strictly starshaped on $\mathbb{R}$. We call a function $g : \mathbb{R} \to \mathbb{R}$ with $g(0) > 0$ strictly starshaped if for all $x \in \mathbb{R}$, $0 < \alpha < 1$, $\alpha g(x) < g(\alpha x)$.

**Proof.** Property 1. The lower bound and larger upper bound are in Gordon (1941). The tighter upper bound was established by Birnbaum (1942). Property 2. One easily verifies $h'(x) = h(x)(h(x) - x)$ which, given Property 1, is positive. Since convexity implies the derivative is increasing, the final statement of 1. implies $h' < 1$. Property 3. Convexity was conjectured by Birnbaum (1950), and proved by Sampford (1953). Logconcavity is established in Marshall and Olkin (2007) Lemma B.4 p.437. Marshall and Olkin establish the result using Property 2. That $h/\beta$ is increasing follows from the logconcavity of $h$. That $\frac{d}{dx} \frac{h(x)}{h'(x)} < 1$, follows from the convexity of $h$ follows immediately from observing $\frac{d}{dx} \frac{h(x)}{h'(x)} = 1 - \frac{h(x)h''(x)}{(h'(x))^2}$. Property 4. A continuous positive function is starshaped on $\mathbb{R}$ if it is starshaped on $(0, \infty)$. $h$ is starshaped if $\frac{h(x)}{h'(x)} \leq \frac{1}{x}$ on $(0, \infty)$. One easily verifies $h'(x) = h(x)(h(x) - x)$. Hence, it suffices to show that $(h(x) - x)x \leq 1$, the result therefore follows from 1. To establish $x \mapsto \frac{h(x)}{h'(x)}$ is starshaped on $(0, \infty)$, we require that $h/\beta'x$ is decreasing on $x > 0$, this follows from $h$ being starshaped ($h/x$ is decreasing) and by the convexity of $h$, $\frac{1}{h}$ is decreasing. Hence, the result follows from the fact that the product of two decreasing positive functions is decreasing.

**Proposition 9** Suppose information structures are Gaussian and $\gamma_I \in \Gamma_I$ induces regression to the mean. $\mathbb{E}[AS_{\gamma_I}]$ and $\mathbb{E}[EC_{\gamma_I}]$ depend on the information structure $\gamma_I \in \Gamma_I$ only through the three dimensional parameter $(\beta_{G_{\gamma_I}R_{\gamma_I}}T_{\gamma_I}, \sigma_{R_{\gamma_I}|T_{\gamma_I}}, \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}) \in [0, 1] \times [0, 1] \times [0, 1]$. Furthermore,

1. $\mathbb{E}[AS_{\gamma_I}]$ is increasing in $(\beta_{G_{\gamma_I}R_{\gamma_I}}T_{\gamma_I}, \sigma_{R_{\gamma_I}|T_{\gamma_I}}, \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]})$.

2. For each $(\sigma_{R_{\gamma_I}|T_{\gamma_I}}, \sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}) \in [0, 1] \times [0, 1]$, $\mathbb{E}[EC_{\gamma_I}]$ is increasing in $\beta_{G_{\gamma_I}R_{\gamma_I}}T_{\gamma_I}$.

3. For each $(\beta_{G_{\gamma_I}R_{\gamma_I}}T_{\gamma_I}, \sigma_{R_{\gamma_I}|T_{\gamma_I}}) \in [0, 1] \times [0, 1]$, $\mathbb{E}[EC_{\gamma_I}]$ is increasing in $\sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}$.

**Proof.** Part 1. It is immediate from (11) that for each $t \in T_{\gamma_I}$, $\widehat{a}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}|T_{\gamma_I} = t])$ is increasing in $\beta_{G_{\gamma_I}R_{\gamma_I}}T_{\gamma_I}$. Similarly, $\widehat{a}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}|T_{\gamma_I} = t])$ is seen to be increasing in $\sigma_{R_{\gamma_I}|T_{\gamma_I}}$ by the star-shaped property of the normal hazard function (Lemma C1). The convexity property of the normal hazard (Lemma C1) is inherited by $\widehat{a}_{\gamma_I}$, hence it follows from Jensen’s inequality that $\mathbb{E}[AS_{\gamma_I}] = \mathbb{E}[\widehat{a}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}|T_{\gamma_I}])]$ is increasing in $\sigma_{E[S_{\gamma_I}|T_{\gamma_I}]}$.

Part 2. Straightforward manipulations show that $EC_{\gamma_I}$ has the representation

$$EC_{\gamma_I} = \sigma_{R_{\gamma_I}|T_{\gamma_I}} \Psi \left( \frac{\mathbb{E}[S_{\gamma_I}|T_{\gamma_I}] + \widehat{a}_{\gamma_I}(\mathbb{E}[S_{\gamma_I}|T_{\gamma_I}])}{\sigma_{R_{\gamma_I}|T_{\gamma_I}}} \right)$$
where, \( \Psi(x) = \mathbb{E}[(\varepsilon - x)^+] \) with \( \varepsilon \sim \mathcal{N}(0, 1) \). Note that \( \Psi \) is a decreasing function. Since \( \tilde{a}s_{\gamma I} \) is increasing in \( \beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}} \), \( EC_{\gamma I} \) is made smaller in first-order stochastic dominance order and therefore has smaller expectation.

**Part 3.** This is more delicate. Using the above representation, we can write \( \mathbb{E}[EC_{\gamma I}] \) as

\[
\mathbb{E}[EC_{\gamma I}] = \sigma_{R_{\gamma I}|T_{\gamma I}} \mathbb{E} \left[ \Psi \left( \frac{\sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}] \varepsilon} + \tilde{a}s_{\gamma I} \left( \frac{\sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}] \varepsilon}}{\sigma_{R_{\gamma I}|T_{\gamma I}}} \right)}{\sigma_{R_{\gamma I}|T_{\gamma I}}} \right) \right].
\]

(16)

Differentiation gives

\[
\frac{\partial \mathbb{E}[EC_{\gamma I}]}{\partial \sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}]}} = \mathbb{E} \left[ \Psi' \left( \tilde{a}s_{\gamma I} \varepsilon + \varepsilon \right) \right].
\]

Differentiating the representation for \( \tilde{a}s_{\gamma I} \) in (11) gives

\[
\tilde{a}s'_{\gamma I} = \tilde{a}s'_{\gamma I} \left( \sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}] \varepsilon} \right) = \frac{\beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}} h'}{1 - \beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}} h'}.
\]

Using this together with the fact that \( \Psi' = -(1 - F) \) (\( F \) is the standard normal cdf) gives

\[
\frac{\partial \mathbb{E}[EC_{\gamma I}]}{\partial \sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}]}} = -\mathbb{E} \left[ \frac{1 - F}{1 - \beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}} h'} \varepsilon \right].
\]

Setting \( k(x, \beta) = \frac{1 - F(x)}{1 - \beta h'(x)} \) and \( a(\varepsilon) = \frac{\sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}] \varepsilon + \tilde{a}s_{\gamma I} \left( \frac{\sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}] \varepsilon}}{\sigma_{R_{\gamma I}|T_{\gamma I}}} \right)}{\sigma_{R_{\gamma I}|T_{\gamma I}}} \), this becomes

\[
\frac{\partial \mathbb{E}[EC_{\gamma I}]}{\partial \sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}]}} = -\int_{-\infty}^{\infty} k(a(\eta), \beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}}) \eta f(\eta) d\eta.
\]

Using symmetry of the normal density around zero,

\[
\frac{\partial \mathbb{E}[EC_{\gamma I}]}{\partial \sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}]}} = \int_{0}^{\infty} \left( k(a(-\eta), \beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}}) - k(a(\eta), \beta_{G_{\gamma I} R_{\gamma I}, T_{\gamma I}}) \right) \eta f(\eta) d\eta.
\]

The desired result now follows from the following facts. For each \( 0 \leq \beta < 1 \), \( x \mapsto k(x, \beta) \) is quasiconcave and for each \( x \in \mathbb{R}, \beta \mapsto k(x, \beta) \) is increasing on \([0, 1]\) and satisfies \( \lim_{x \to -\infty} k(x, \beta) = 1 \) and \( k(0, \beta) = 0.5 \left( 1 - \frac{2\beta}{\pi} \right)^{-1} \). Hence, \( k(0, \frac{\pi}{4}) = 1 \). It follows that for \( \beta \leq \frac{\pi}{4}, (1) \) \( \min_{x \leq 0} k(x, \beta) \geq \max_{x \geq 0} k(x, \beta) \), (2) \( x \mapsto k(x, \beta) \) is decreasing on \( \mathbb{R}_+ \). Hence, \( x > 0, x' < x \) imply \( k(x', \beta) > k(x, \beta) \). Since \( a \) is an increasing function and it is positive on \( \mathbb{R}_+ \), it follows that \( k(a(-t), \beta) > k(a(t), \beta) \) for \( t \in \mathbb{R}_+ \), therefore \( \frac{\partial \mathbb{E}[EC_{\gamma I}]}{\partial \sigma_{\mathbb{E}[S_{\gamma I}|T_{\gamma I}]}} > 0 \) as required.

**Proposition 10** Suppose information structures are Gaussian and \( \gamma I \in \Gamma_I \) induces regression to the mean. For each fixed \( \sigma_{R_{\gamma I}|T_{\gamma I}}, \mathbb{E}[EC_{\gamma I}] \) and \( \mathbb{E}[AS_{\gamma I}] \) satisfy a strict Spence-Mirlees
condition in the remaining parameters. Specifically,

\[
\frac{\partial \mathbb{E}[AS_{\gamma_1}]}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} \cdot \frac{\partial \mathbb{E}[AS_{\gamma_1}]}{\partial \beta_{\gamma_1 T_{\gamma_1}}} > \frac{\partial \mathbb{E}[EC_{\gamma_1}]}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} \cdot \frac{\partial \mathbb{E}[EC_{\gamma_1}]}{\partial \beta_{\gamma_1 T_{\gamma_1}}}.
\]

**Proof.** Implicit differentiation of (16) gives

\[
\frac{\partial \mathbb{E}[EC_{\gamma_1}]}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} = \mathbb{E} \left[ \Psi' \left( \frac{\partial AS_{\gamma_1}}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} \right) + \varepsilon \right],
\]

\[
\frac{\partial \mathbb{E}[EC_{\gamma_1}]}{\partial \beta_{\gamma_1 T_{\gamma_1}}} = \mathbb{E} \left[ \Psi' \left( \frac{\partial AS_{\gamma_1}}{\partial \beta_{\gamma_1 T_{\gamma_1}}} \right) \right] < 0,
\]

the inequality following from \( \Psi' < 0 \). Similarly, implicit differentiation of the representation for \( \hat{a}_{\sigma_1} \) in (11) gives

\[
\frac{\partial \mathbb{E}[AS_{\gamma_1}]}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} = \mathbb{E} \left[ \frac{\partial AS_{\gamma_1}}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} \right] = \mathbb{E} \left[ \frac{\beta_{\gamma_1 R_{\gamma_1}.T_{\gamma_1}} h'}{1 - \beta_{\gamma_1 R_{\gamma_1}.T_{\gamma_1}} h'} \right] > 0,
\]

\[
\frac{\partial \mathbb{E}[AS_{\gamma_1}]}{\partial \beta_{\gamma_1 T_{\gamma_1}}} = \mathbb{E} \left[ \frac{\partial AS_{\gamma_1}}{\partial \beta_{\gamma_1 T_{\gamma_1}}} \right] = \mathbb{E} \left[ \frac{\sigma_{T_{\gamma_1}} h}{1 - \beta_{\gamma_1 R_{\gamma_1}.T_{\gamma_1}} h'} \right] > 0.
\]

The marginal rate of substitution between \( \sigma_E[S_{\gamma_1}|T_{\gamma_1}] \) and \( \beta \) for the function \( \mathbb{E}[AS] \) is given by

\[
\frac{\partial \mathbb{E}[AS_{\gamma_1}]}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} = \beta \frac{\mathbb{E} \left[ \frac{h'}{1 - \beta T_{\gamma_1}} \right]}{\mathbb{E} \left[ \frac{h}{1 - \beta T_{\gamma_1}} \right]} = k > 0.
\]

Thus, with \( k \) so defined, the operator \( D_k \) defined as \( D_k = \left( \frac{\partial}{\partial \sigma_E[S_{\gamma_1}|T_{\gamma_1}]} - k \frac{\partial}{\partial \beta} \right) \) satisfies

\[
D_k (\mathbb{E}[AS_{\gamma_1}]) = \mathbb{E}[D_k (AS_{\gamma_1})] = 0.
\]

We will show

\[
D_k (\mathbb{E}[EC]) = \mathbb{E}[D_k (EC)] = \mathbb{E} \left[ \Psi' \left( D_k (AS_{\gamma_1}) + \varepsilon \right) \right] > 0.
\]

First, noting that \( \Psi' \) is increasing in \( \varepsilon \), hence, \( \Psi' \) and \( \varepsilon \) have a positive covariance \( \mathbb{E} [\Psi', \varepsilon] > 0 \). It suffices therefore to show that

\[
\mathbb{E} \left[ \Psi' . D_k (AS_{\gamma_1}) \right] > 0.
\]

Note that

\[
D_k (AS_{\gamma_1}) = \frac{\beta_{\gamma_1 R_{\gamma_1}.T_{\gamma_1}} h'}{1 - \beta_{\gamma_1 R_{\gamma_1}.T_{\gamma_1}} h'} \left( \varepsilon - \frac{k \sigma_{T_{\gamma_1}} h}{\beta_{\gamma_1 R_{\gamma_1}.T_{\gamma_1}} h'} \right).
\]
The function \( x \mapsto h(x)/h'(x) \) is increasing and starshaped (Lemma C1). It is a consequence of Lemma A1 that \( \frac{\delta_k}{\sigma_{R_{\gamma_1}T_{\gamma_1}}} \) is starshaped in \( \varepsilon \) (which is inherited from \( h \)). Hence, since starshaped functions are closed under composition, \( D_k (AS_{\gamma_1}) \) has a single sign change from negative to positive as \( \varepsilon \) traverses from \(-\infty\) to \(\infty\). The rest of the proof is standard. Suppose \( D_k (AS_{\gamma_1}) \) is zero at some realisation \( \varepsilon = \varepsilon' \), then setting \( c \) equal to the value taken by \( \Psi' \) at \( \varepsilon' \) and using the fact that \( \Psi' \) is increasing in \( \varepsilon \), it is established that \( (\Psi' - c) . D_k (AS_{\gamma_1}) \geq 0 \) on \( \mathbb{R} \), the inequality being strict except at \( \varepsilon' \). Therefore,

\[
0 < \mathbb{E} [ (\Psi' - c) . D_k (AS_{\gamma_1}) ] = \mathbb{E} [ \Psi' . D_k (AS_{\gamma_1}) ] - c \mathbb{E} [ D_k (AS_{\gamma_1}) ] = \mathbb{E} [ \Psi' . D_k (AS_{\gamma_1}) ] .
\]

This establishes \( D_k (\mathbb{E} [ EC_{\gamma_1} ]) > 0 \). Since \( \frac{\partial \mathbb{E} [ EC_{\gamma_1} ]}{\partial G_{\gamma_1} R_{\gamma_1} T_{\gamma_1}} < 0 \), this implies the result. ■

**Appendix D: Material Omitted from Section 6**

**Lemma 1** The set \( \Omega(\Gamma_{Q_1}) \) can be parameterized equivalently as

\[
\Omega(\Gamma_{Q_1}) = \left\{ \left[ \mathbb{E}[AS_{\gamma_1}], \mathbb{E}[EC_{\gamma_1}] \right] \mid \Sigma_{(R_1,G_1)|T_{\gamma_1}} \in \mathcal{S}_2, 0 \preceq \Sigma_{(R_1,G_1)|T_{\gamma_1}} \preceq \Sigma_{(R_1,G_1)} \right\}.
\]

**Proof.** Assumptions A1 and A2 simply allow definition of \( R_1, G_1 \). A3 (\( \Sigma_{YQ_1} = \Sigma_{YQ_1} \Sigma_{Q_1}^{-1} \Sigma_{Q_1}T_{\gamma_1} \)) means that Proposition 8 applies wherever \( AS_{\gamma_1} \) is defined. Hence, the marginal distribution of \( AS_{\gamma_1} \) is determined by the three quantities \( \sigma_{R_{\gamma_1}T_{\gamma_1}}, \beta_{G_{\gamma_1}R_{\gamma_1}T_{\gamma_1}}, \mathbb{E}[S_{\gamma_1}T_{\gamma_1}] \). Similarly for \( EC_{\gamma_1} \). By the law of total variance \( \sigma_{S_{\gamma_1}T_{\gamma_1}}^2 = \sigma^2_{S_{\gamma_1}T_{\gamma_1}} - \sigma^2_{S_{\gamma_1}T_{\gamma_1}} = \sigma^2_{S_{\gamma_1}T_{\gamma_1}} - (\sigma_{R_{\gamma_1}T_{\gamma_1}} - \sigma_{G_{\gamma_1}T_{\gamma_1}})^2 \). \( \sigma_{S_{\gamma_1}T_{\gamma_1}}^2 \) is determined only by the private information of training firm, \( Q_1 \), so remains constant. Hence \( \mathbb{E}[S_{\gamma_1}T_{\gamma_1}] \) is determined by the conditional covariance matrix. It follows immediately from the standard formula for univariate regression that \( \beta_{G_{\gamma_1}R_{\gamma_1}T_{\gamma_1}} = \sigma_{G_{\gamma_1}R_{\gamma_1}T_{\gamma_1}}^2 / \sigma_{R_{\gamma_1}T_{\gamma_1}}^2 \). Therefore, \( \beta_{G_{\gamma_1}R_{\gamma_1}T_{\gamma_1}} \) is also determined by the conditional covariance matrix. Hence, \( \mathbb{E}[EC_{\gamma_1}] \) and \( \mathbb{E}[AS_{\gamma_1}] \) depend on \( \Sigma_{\gamma_1} \in \mathcal{S}_2 \) only through \( \Sigma_{(R_1,G_1)|T_{\gamma_1}} \in \mathcal{S}_2 \). This establishes the first part of the parameterization.

It remains to establish the semidefiniteness conditions. The constraint \( 0 \preceq \Sigma_{(R_1,G_1)|T_{\gamma_1}} \) arises because \( \Sigma_{(R_1,G_1)|T_{\gamma_1}} \) is a covariance matrix and so must be positive semidefinite. The second semidefinite constraint states that \( \Sigma_{(R_1,G_1)|T_{\gamma_1}} \) must also be smaller in the semidefinite order than the unconditional covariance matrix for \( (R_{\gamma_1}, G_{\gamma_1}) \). To see this, note by the law of total variance \( \text{Var} \left( (R_{\gamma_1}, G_{\gamma_1}) \right) = \mathbb{E}[\text{Var} \left( (R_{\gamma_1}, G_{\gamma_1}) \mid T_{\gamma_1} \right)] + \text{Var}(\mathbb{E}[R_{\gamma_1} \mid T_{\gamma_1}], \mathbb{E}[G_1 \mid T_{\gamma_1}]) \). In the Gaussian case, \( \mathbb{E}[\text{Var} \left( (R_{\gamma_1}, G_{\gamma_1}) \mid T_{\gamma_1} \right)] = \Sigma_{(R_1,G_1)|T_{\gamma_1}} \) and \( \text{Var} \left( (R_{\gamma_1}, G_{\gamma_1}) \right) = \Sigma_{(R_1,G_1)} \). The second semidefinite constraint arises because \( \text{Var}(\mathbb{E}[R_{\gamma_1} \mid T_{\gamma_1}], \mathbb{E}[G_1 \mid T_{\gamma_1}]) \) is a covariance matrix and so must be positive semidefinite. ■

**Proposition 11** Suppose information structures are Gaussian with fixed inside information \( Q_1 \), and suppose \( \beta_{G_1 R_1} \leq \frac{2}{3} \). If \( \gamma_1^* \in \Gamma_1 \) solves program \( P1 \) for some \( AS_{\gamma_1} \geq 0 \), then \( \Sigma_{(R_1,G_1)|T_{\gamma_1}} \) is singular. Furthermore,
1. If \( G_I \) and \( R_I \) are colinear (\( G_I = \beta_0 + \beta_G R_I R_I \)), then for some \( \varepsilon \perp G_I, \varepsilon \sim N(0, \sigma_\varepsilon) \), 
\[ T_{\gamma_I} = G_I + \varepsilon. \]

2. If \( G_I \) and \( R_I \) are not colinear, \( T_{\gamma_I} \) is two dimensional and for some \( \alpha \in \mathbb{R}^2 \), \( G_I = \beta_G R_I R_I + \alpha \cdot T_{\gamma_I} \).

**Proof.** We apply Lemma C1 to establish that \( \Sigma_{(R_I, G_I) | T_{\gamma_I}} \) is singular. Proposition 10 implies that at least one of the constraints \( 0 \leq \Sigma_{(R_I, G_I) | T_{\gamma_I}} \) or \( \Sigma_{(R_I, G_I) | T_{\gamma_I}} \leq \Sigma_{(R_I, G_I)} \) must bind, therefore it suffices to show that \( \Sigma_{(R_I, G_I) | T_{\gamma_I}} \leq \Sigma_{(R_I, G_I)} \) does not bind alone. Equivalently, that \( \Sigma_{(R_I, S_I) | T_{\gamma_I}} \leq \Sigma_{(R_I, S_I)} \) does not bind alone. The constraint \( \Sigma_{(R_I, S_I) | T_{\gamma_I}} \leq \Sigma_{(R_I, S_I)} \) expressed in terms of \( \beta_{G_I R_I T_{\gamma_I}}, \sigma_{E[S_I | T_{\gamma_I}]}, \sigma_{R_I S_I | T_{\gamma_I}} \) becomes

\[
0 \leq \begin{pmatrix} \sigma_{R_I S_I | T_{\gamma_I}}^2 - \sigma_{R_I S_I | T_{\gamma_I}}^2 & (1 - \beta_{G_I R_I T_{\gamma_I}}) \sigma_{R_I S_I | T_{\gamma_I}}^2 \\ (1 - \beta_{G_I R_I T_{\gamma_I}}) \sigma_{R_I S_I | T_{\gamma_I}}^2 & \sigma_{S_I | T_{\gamma_I}}^2 - \sigma_{E[S_I | T_{\gamma_I}]}^2 \end{pmatrix}.
\]

That is, \( (\sigma_{E[S_I | T_{\gamma_I}]}, \beta_{G_I R_I T_{\gamma_I}}) \) is contained in a cone with vertex at \( (0, \overline{\beta}) \) with \( (1 - \beta) \sigma_{R_I S_I | T_{\gamma_I}} = \sigma_{R_I S_I}, \) i.e. \( \overline{\beta} = 1 - \frac{\sigma_{R_I S_I}}{\sigma_{R_I T_{\gamma_I}}} \leq 1 - \frac{\sigma_{R_I S_I}}{\sigma_{R_I T_{\gamma_I}}} \leq \frac{1}{\sigma_{R_I T_{\gamma_I}}} = \beta_{G_I R_I} \). If \( \beta_{G_I R_I} \leq \frac{\sigma_{R_I S_I}}{\sigma_{R_I T_{\gamma_I}}} \), this cone satisfies the conditions of Corollary 1 and applying the corollary, there is a direction of increase for \( \mathbb{E}[AS_{\gamma_I}], \mathbb{E}[EC_{\gamma_I}] \) pointing into the interior of the cone. Hence the constraint \( \Sigma_{(R_I, S_I) | T_{\gamma_I}} \leq \Sigma_{(R_I, S_I)} \) does not bind alone. Therefore \( \Sigma_{(R_I, G_I) | T_{\gamma_I}} \) is singular as required. This argument establishes the result for the nondegenerate cases \( \sigma_{R_I T_{\gamma_I}}^2 \), \( 0 < \sigma_{R_I S_I | T_{\gamma_I}}^2 < \sigma_{R_I}^2 \). If \( \sigma_{R_I T_{\gamma_I}}^2 = \sigma_{R_I}^2 \), \( T_{\gamma_I} \) is orthogonal to \( R_I \) and the cone has no interior. However, in this case the cone degenerates into the line segment \( \beta = \beta_{G_I R_I} \). Proposition 9 implies that traversing this line segment in a direction of increasing \( \sigma_{E[S_I | T_{\gamma_I}]} \) increases \( \mathbb{E}[EC_{\gamma_I}] \) and \( \mathbb{E}[AS_{\gamma_I}] \), so the result still holds. If \( \sigma_{R_I S_I | T_{\gamma_I}}^2 = 0 \), then \( \Sigma_{(R_I, G_I) | T_{\gamma_I}} \) is singular so there is nothing to prove. ■

**Remark 4** In the Gaussian case, suppose \( Q_I' \) is more informative than \( Q_I \). The following statements are equivalent.

1. For any \( \gamma_I \in \Gamma_Q \), there is a \( \gamma_I' \in \Gamma_{Q_I'} \) such that \( (AS_{\gamma_I'}, EC_{\gamma_I'}) \) is equal in distribution to \( (AS_{\gamma_I}, EC_{\gamma_I}) \).

2. The improvement in information leaves unchanged the correlation between \( R_I \) and \( G_I \), that is \( \sigma_{R_I S_I} = \sigma_{R_I S_I'} \).

**Proof.** It is required that for each disclosure \( T_{\gamma_I} \) under private information \( Q_I \), there exists a choice of \( T_{\gamma_I'} \) under private information \( Q_I' \) such that (1) \( \mathbb{E}[S_I' | T_{\gamma_I'}] \) has the same distribution as \( \mathbb{E}[S_I | T_{\gamma_I}] \), i.e. \( \text{Var}(\mathbb{E}[S_I' | T_{\gamma_I'}]) = \text{Var}(\mathbb{E}[S_I | T_{\gamma_I}]) \). (2) the functional forms of \( \hat{a}s \) and \( \hat{e}c \) are the same in the two cases. Hence, we require it be possible to set \( \sigma_{R_I S_I | T_{\gamma_I'}}^2 = \sigma_{R_I S_I | T_{\gamma_I}}^2 \) and \( \beta_{G_I R_I S_I | T_{\gamma_I'}} = \beta_{G_I R_I S_I | T_{\gamma_I}} \) and \( \text{Var}(\mathbb{E}[S_I' | T_{\gamma_I'}]) = \text{Var}(\mathbb{E}[S_I | T_{\gamma_I}]) \). Equivalently,
\[ \sigma_{S_j' R_j'|T_{\gamma_1}} = \sigma_{S_{\gamma_1} R_{\gamma_1}|T_{\gamma_1}}, \quad \sigma_{R_j'|T_{\gamma_1}}^2 = \sigma_{R_{\gamma_1}|T_{\gamma_1}}^2 \] and \( \text{Var}(\mathbb{E}[S_j'|T_{\gamma_1}]) = \text{Var}(\mathbb{E}[S_j|T_{\gamma_1}]). \) By the law of total variance \( \text{Var}(\mathbb{E}[S_j'|T_{\gamma_1}]) = \sigma_{S_j'|T_{\gamma_1}}^2 - \sigma_{S_j|^T_{\gamma_1}}^2, \) similarly for \( \text{Var}(\mathbb{E}[S_j'|T_{\gamma_1}]). \) Hence, we require \( \sigma_{S_j'|T_{\gamma_1}}^2 = \sigma_{S_j|^T_{\gamma_1}}^2 + \sigma_{S_j'|T_{\gamma_1}}^2. \) It is also required that the semidefinite constraints \( 0 \preceq \Sigma(R'_j, S'_j)|T_{\gamma_1} \preceq \Sigma(R_j, S_j) \) are satisfied. In sum, we require that for any \( 2 \times 2 \) symmetric matrix \( \Sigma(R_j, S_j)|T_{\gamma_1} \) satisfying \( 0 \preceq \Sigma(R_j, S_j)|T_{\gamma_1} \preceq \Sigma(R_j, S_j), \) it is also true that \( 0 \preceq \Sigma(R_j, S_j)|T_{\gamma_1} \preceq \Sigma(R_j, S_j) \) is evidently satisfied so the condition reduces to \( \Sigma(R_j, S_j)|T_{\gamma_1} \preceq \Sigma(R_j, S_j) + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{S_j}'^2 - \sigma_{S_j}^2 \end{pmatrix} \preceq \Sigma(R_j, S_j). \)

Given \( \Sigma(R_j, S_j)|T_{\gamma_1} \preceq \Sigma(R_j, S_j), \) it suffices that \( \Sigma(R_j, S_j) + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{S_j}'^2 - \sigma_{S_j}^2 \end{pmatrix} \preceq \Sigma(R_j, S_j). \) Equivalently, \( 0 \preceq \begin{pmatrix} \sigma_{R_j}'^2 - \sigma_{R_j}^2 & \sigma_{R_j'}^2\sigma_{S_j}' - \sigma_{R_j}^2\sigma_{S_j} \\ \sigma_{R_j'}^2\sigma_{S_j}' - \sigma_{R_j}^2\sigma_{S_j} & 0 \end{pmatrix} \) which holds if and only iff \( \sigma_{R_j}'^2 \geq \sigma_{R_j}^2 \) (which is ensured by \( Q'I \) is more informative than \( QI \)) and \( \sigma_{R_j}'^2\sigma_{S_j}' = \sigma_{R_j}^2\sigma_{S_j}, \) which is the desired result.

The following simple Lemma together with Corollary 1 is key to proving Proposition 12.

**Lemma 2** For each \( 0 \preceq \sigma_{R_j}|T_{\gamma_1} \preceq \sigma_{R_j}^2, \) \( 0 \preceq \Sigma(R_j, S_j) \preceq \Sigma(R_j, S_j), \) implies:

\[
\left\{ \beta \mid 0 \preceq \begin{pmatrix} \sigma_{R_j}|T_{\gamma_1} \\ (1 - \beta)\sigma_{R_j}|T_{\gamma_1} \\ \sigma_{S_j}|T_{\gamma_1} \end{pmatrix} \leq \Sigma(R_j, S_j), \sigma_{S_j}|T_{\gamma_1} \geq 0 \right\} \subseteq \left\{ \beta \mid 0 \preceq \begin{pmatrix} \sigma_{R_j}|T_{\gamma_1} \\ (1 - \beta)\sigma_{R_j}|T_{\gamma_1} \\ \sigma_{S_j}|T_{\gamma_1} \end{pmatrix} \leq \Sigma(R_j, S_j), \sigma_{S_j}|T_{\gamma_1} \geq 0 \right\}.
\]

**Proof.** Immediate. If \( \beta \) satisfies the more restrictive condition, it satisfies the less restrictive condition.
Proposition 12. With free disclosure of information, information acquisition increases efficiency. Specifically, $V(\Sigma_{(R_l,G_l)}, A_S)$ is increasing in $\Sigma_{(R_l,G_l)}$ in positive semidefinite order in the region $\beta_{G_l R_l} \leq \frac{\pi}{4}$.

Proof. As in the proof of Proposition 11, for fixed $\sigma_{R_l,T_{\gamma_l}}$, the constraint $\Sigma_{(R_l,G_l)\mid T_{\gamma_l}} \preceq \Sigma_{(R_l,G_l)}$ represents, in $(\sigma_{E[S_{\gamma_l} \mid T_{\gamma_l}], \beta_{G_l R_l} R_{\gamma_l} T_{\gamma_l}})$ space as a cone satisfying the conditions of Corollary 1. The corresponding region for the constraint $0 \preceq \Sigma_{(R_l,G_l)\mid T_{\gamma_l}}$ is a convex elliptical region containing the vertex of the cone. Improving the information from $\Sigma_{(R_l,G_l)}$ to $\Sigma_{(R_0,G_0)} \succeq \Sigma_{(R_l,G_l)}$ enlarges the elliptical region by set inclusion, but shifts the cone so that neither of the two constraint sets generally contains the other. The general situation is as illustrated in Figure 2 in which the cone with vertex at $(0, \beta)$ represents the constraint set with less private information and the one with vertex at $(0, \beta')$ represents the constraint set with better private information. The two constraint sets have been drawn to satisfy the conclusion of Lemma 2 which asserts that the set of $\beta_{G_{\gamma_l} R_{\gamma_l} T_{\gamma_l}}$ achievable with less private information is also achievable with more private information. By Corollary 1 from any point in the constraint with vertex at $(0, \beta)$, there must be a path along which both $E[EC_{\gamma_l}]$ and $E[AS_{\gamma_l}]$ are increasing which eventually passes into the constraint set with vertex at $(0, \beta')$. (Note, one can always construct a new cone which is a subset of the one with vertex at $(0, \beta)$—the construction is shown for the path initiating on the elliptical part of the constraint.) This establishes the conclusion. \[53\]