On the Robustness of ‘Good News and Bad News’∗

Christopher P. Chambers† Paul J. Healy‡

Abstract

In situations where noisy signals of some underlying variable are observed, Paul Milgrom (The Bell Journal of Economics, 12(2): 380–391) showed that a higher signal implies a higher posterior (in the sense of first-order stochastic dominance) for every non-degenerate prior if and only if the conditional distribution satisfies the strict monotone likelihood ratio property (MLRP). We show that for any non-degenerate prior with bounded support there exists a symmetric and quasiconcave conditional distribution such that a higher signal implies a lower posterior. Thus, when the prior is known but the conditional distribution is unrestricted it is always possible that “good” signals may be “bad” news.

Keywords: Signal extraction, Bayes’s rule, MLRP, stochastic dominance, updating

Draft Copy: December 5, 2007
1 Introduction

Suppose that a single-dimensional parameter is observed with noise. For two possible observations $z$ and $z'$ with $z < z'$ when is it possible to say that the posterior distribution of the parameter after observing $z'$ first-order stochastically dominates the distribution after observing $z$, regardless of the structure of the prior distribution of the parameter?\footnote{1} Milgrom’s celebrated “good news, bad news” result (Milgrom (1981)) establishes that it is necessary and sufficient that the noise structure satisfy the strict monotone likelihood ratio property (MLRP).\footnote{2}

The MLRP assumption is frequently employed in the mechanism design literature because of its resulting monotonicity properties. In many settings, however, the exact structure of the noise is not known or not well understood. In these cases it may be unduly restrictive to assume the MLRP; any strong assumption about the structure of noise may lead to models that are either misspecified or not robust. The prior, on the other hand, is often an object of choice for a principal (as in classical Bayesian statistics and decision theory). In this note we reconsider Milgrom’s result taking the prior as fixed and allowing for flexibility in the distribution of noise. In this sense, our framework can be interpreted as the dual of Milgrom’s framework.

When the prior is fixed (and has a bounded support) but the noise distribution is unspecified we obtain the following negative conclusion: There exists a noise distribution independent of the parameter realization whose density is symmetric and quasiconcave and there is a pair of signals $z$ and $z'$ with $z' > z$ such that the posterior after observing $z'$ is strictly dominated by the posterior after observing $z$; the high signal is interpreted as “bad news” and the low signal is interpreted as “good news”. Surprisingly, this noise distribution can be constructed based only on the support of the prior; it need not depend in any way on the shape of the prior distribution.

The contrapositive of Milgrom’s necessity result says that for any noise distribution that does not satisfy strict MLRP there exists some non-degenerate prior such that monotonicity of the posterior will fail. Our result differs in two ways; first, we take as given any non-degenerate prior with bounded support and construct a corresponding noise distribution that does not satisfy strict MLRP (this is the sense in which our result is the dual of Milgrom’s). Second, we show that monotonicity will not only fail, but will be reversed for some pair of

\begin{itemize}
  \item[A cumulative distribution function $F$ strictly first-order stochastically dominates $G$ if $F(x) \leq G(x)$ for all $x$ with strict inequality for some $x$.]
  \item[If $x$ and $x'$ are two realizations of the parameter, $z$ and $z'$ are two realizations of the noisy signal, and $f(z|x)$ is the density of the conditional distribution of the signal, then the strict MLRP requires that $f(z'|x')f(z|x) > f(z'|x)f(z|x')$ for all $x' > x$ and $z' > z$.]
\end{itemize}
To apply our result, consider a modeler who selects a (bounded) prior but is unable or unwilling to specify a distribution of the signal noise. This is in the spirit of the robust mechanism design literature, which follows the “Wilson doctrine” (Wilson (1987)) that models should place as few restrictions as possible on agents’ common knowledge and on the underlying distributions used in the model. The modeler may consider it reasonable to admit any additive noise distribution that is independent of the underlying parameter and whose density is symmetric about zero and quasiconcave. With these specifications our result implies that there must be an admissible noise distribution and a pair of signals such that the higher signal’s posterior is stochastically dominated by the lower signal’s posterior.

The intuition for our result is simple. For any prior distribution with bounded support consider a mean-zero error distribution whose support is significantly larger than the support of the prior. If this error distribution has sufficiently “fat” tails then any extremely high observation $z'$ is likely due to a very large error term, indicating a relatively small parameter value. For a less extreme observation $z$ that falls in the support of the prior it becomes more likely that the observation is indicative of a large parameter value. By carefully constructing the noise distribution one can guarantee that the posterior after observing $z'$ is stochastically dominated by the posterior after observing $z$. In fact, the construction of the noise distribution needs only to depend on the support of the prior. For the case where the prior has support on $[-10, 10]$ the constructed noise distribution is shown in Figure 1. With that support the posterior after observing $z' = 30$ is dominated by the posterior after observing $z = 10$.

2 The model and main result

Consider a state space $\Omega$ with probability measure $P$ defined on some $\sigma$-algebra of $\Omega$. Let $X$ be a random variable mapping $\Omega$ into $\mathbb{R}$. We say that $X$ is bounded if there exist $a, b \in \mathbb{R}$ such that $P(\{\omega : X(\omega) \in [a, b]\}) = 1$. In what follows we refer to the support of $X$ as the

---

3 Recall that the stochastic dominance ordering is not complete; simply observing that monotonicity will fail implies that there are two posterior distributions that cannot be ordered according to the dominance ordering, but not that their orderings will be reversed relative to the ordering of the signals.

4 This family is known by statisticians to provide robust Bayesian hypothesis tests; see Berger (1994). In a related paper (Chambers and Healy (2007)) we show that with this family of noise distributions the posterior mean lies between the prior mean and the signal observation if and only if the prior distribution is symmetric and quasiconcave.

5 All random variables are assumed to be measurable.
Figure 1: An example of the noise distribution used in the proof.

Theorem. Let $X$ be a non-degenerate bounded random variable. There exists a noise term $\tilde{\varepsilon}$ and real numbers $z, z' \in \mathbb{R}$ with $z < z'$ such that $X|\{X + \tilde{\varepsilon} = z\}$ strictly first order stochastically dominates $X|\{X + \tilde{\varepsilon} = z'\}$.

---

6This differs slightly from the standard definition since it includes any subset $S \subset [a, b]$ on which $P(\{\omega : X(\omega) \in S\}) = 0$.

7Symmetry about zero means that for all $x$, $f_{\tilde{\varepsilon}}(\varepsilon) = f_{\tilde{\varepsilon}}(-\varepsilon)$. Quasiconcavity means that for all $\alpha$, $\{\varepsilon : f_{\tilde{\varepsilon}}(\varepsilon) \geq \alpha\}$ is a convex set.

8Quasiconcavity is also known in the statistical literature as unimodality.
Proof. Let \([a, b] \subset \mathbb{R}\) (with \(a < b\)) be the support of \(X\) and let \(\varepsilon\) be a noise term distributed according to the density

\[
f_\varepsilon(\varepsilon) = \begin{cases} 
\frac{1}{4(b-a)+(b-a)^2} & \text{for } \varepsilon \in [-2(b-a), -(b-a)] \cup [(b-a), 2(b-a)] \\
\frac{1}{4(b-a)+(b-a)^2} (b-a+1+\varepsilon) & \text{for } \varepsilon \in (-a, 0) \\
\frac{1}{4(b-a)+(b-a)^2} (b-a+1-\varepsilon) & \text{for } \varepsilon \in (0, (b-a)).
\end{cases}
\]

(An example of this distribution is shown in Figure 1.) Let \(Z = X + \varepsilon\) and consider \(z = b\) and \(z' = 2b - a\), which are two feasible realizations of \(Z\) such that \(z < z'\). Fix any \(w \in [a, b]\) and note that the cumulative distribution function of \(X|z\) is given by

\[
F_{X|z}(w) = \frac{\int_a^w (x-a+1)dF_X(x)}{\int_a^b (x-a+1)dF_X(x)}.
\]

Moreover, note that \(X|z'\) is simply distributed the same as \(X\), so that \(F_{X|z'} \equiv F_X\).

Separately integrating the numerator and denominator of (1) by parts and rearranging, we obtain

\[
F_{X|z}(w) = \frac{(w-a+1)F_X(w) - \int_a^w F_X(x)dx}{(b-a+1) - \int_a^b F_X(x)dx}
= \frac{F_X(w) - \int_a^w [-F_X(w) + F_X(x)]dx}{1 - \int_a^b [-1 + F_X(x)]dx}
= \frac{F_X(w) + \int_a^w [F_X(w) - F_X(x)]dx}{1 + \int_a^b [1 - F_X(x)]dx}.
\]

Clearly, if \(F_X(w) = 0\) then this expression calculates to 0 at \(w\) and hence \(F_{X|z}(w) \leq F_{X|z'}(w)\), consistent with \(F_{X|z}\) stochastically dominating \(F_{X|z'}\). If \(F_X(w) = 1\) then obviously \(F_{X|z}(w) \leq F_{X|z'}(w)\) since \(F_{X|z'}(w) = 1\).

Finally, consider the case where \(F_X(w) \in (0, 1)\). For these values of \(w\) the following is true of the numerator of (2):

\[
F_X(w) + \int_a^w [F_X(w) - F_X(x)]dx = F_X(w) \left(1 + \int_a^w \left[1 - \frac{F_X(x)}{F_X(w)}\right] dx\right)
\leq F_X(w) \left(1 + \int_a^w [1 - F_X(x)]dx\right)
\leq F_X(w) \left(1 + \int_a^b [1 - F_X(x)]dx\right)
\]
If \( w > a \) then the first inequality is strict since \( F_X(w) < 1 \). If \( w = a \) then the second inequality is strict because \( b > a \).\(^9\) Dividing by the term in parentheses, we thus establish that

\[
\frac{F_X(w) + \int_a^w [F_X(w) - F_X(x)]dx}{1 + \int_a^b [1 - F_X(x)]dx} < F_X(w).
\]

Recalling equation (2) and the fact that \( F_{X|z} \equiv F_X \), the above inequality implies \( F_{X|z} < F_{X|z'} \). Therefore, \( F_{X|z} \) strictly first order stochastic dominates \( F_{X|z'} \).

\[\square\]

References


\(^9\)Recall that this case assumes \( F_X(w) > 0 \), so \( w = a \) implies a point mass at \( a \). Since \( b > a \) it cannot be that \( F_X(w) = 1 \).