TREATMENT HETEROGENEITY AND INDIVIDUAL QUALITATIVE INTERACTION

by

ROBERT S. POULSON

B.A., Brigham Young University, 1991
M.B.A., University of California at Sacramento, 1999
M.S., Utah State University, 2002

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Abstract

The potential for high variability in treatment effects across individuals has been recognized as an important consideration in clinical studies. Surprisingly, little attention has been given to evaluating this variability in design of clinical trials or analyses of resulting data. High variation in a treatment's efficacy or safety across individuals (referred to herein as treatment heterogeneity) may have important consequences because the optimal treatment choice for an individual may be different from that suggested by a study of average effects. We call this an individual qualitative interaction (IQI), borrowing terminology from earlier work - referring to a qualitative interaction (QI) being present when the optimal treatment varies across 'groups' of individuals. At least three techniques have been proposed to investigate treatment heterogeneity: techniques to detect a QI, use of measures such as the density overlap of two outcome variables under different treatments, and use of cross-over designs to observe 'individual effects.' Connections, limitations, and the required assumptions are compared among these techniques through a quantity frequently referred to as subject-treatment (S-T) interaction, but shown here to be the probability of an IQI (PIQI). Their association is studied utilizing a potential outcomes framework that can add insights to results from usual data analyses and to study design features to more directly assess treatment heterogeneity.

Particular attention is given to the density overlap of two outcome variables, each representing an individual's 'potential' response under a different treatment. Connections are made between the overlap quantified as the proportion of similar responses (PSR) and the PIQI. Given a bivariate normal model, the maximum PIQI is shown to be an upper bound for ½ the PSR. Additionally, the characterization of a conditional PSR allows for the PIQI boundaries to

be developed within subgroups defined over observable covariates so that the subset contribution to treatment heterogeneity may be identified. The possibility of similar boundaries is explored outside the normal model using the skew normal distribution. Furthermore, a bivariate PIQI is developed along with its PSR counterpart to help characterize treatment heterogeneity resulting from a bivariate response such as the efficacy and safety of a treatment.

TREATMENT HETEROGENEITY AND INDIVIDUAL QUALITATIVE INTERACTION

by

ROBERT S. POULSON

B.A., Brigham Young University, 1991
M.B.A., University of California at Sacramento, 1999
M.S., Utah State University, 2002

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Approved by:

Major Professor
Gary L. Gadbury

Copyright

Abstract

The potential for high variability in treatment effects across individuals has been recognized as an important consideration in clinical studies. Surprisingly, little attention has been given to evaluating this variability in design of clinical trials or analyses of resulting data. High variation in a treatment's efficacy or safety across individuals (referred to herein as treatment heterogeneity) may have important consequences because the optimal treatment choice for an individual may be different from that suggested by a study of average effects. We call this an individual qualitative interaction (IQI), borrowing terminology from earlier work - referring to a qualitative interaction (QI) being present when the optimal treatment varies across 'groups' of individuals. At least three techniques have been proposed to investigate treatment heterogeneity: techniques to detect a QI, use of measures such as the density overlap of two outcome variables under different treatments, and use of cross-over designs to observe 'individual effects.' Connections, limitations, and the required assumptions are compared among these techniques through a quantity frequently referred to as subject-treatment (S-T) interaction, but shown here to be the probability of an IQI (PIQI). Their association is studied utilizing a potential outcomes framework that can add insights to results from usual data analyses and to study design features to more directly assess treatment heterogeneity.

Particular attention is given to the density overlap of two outcome variables, each representing an individual's 'potential' response under a different treatment. Connections are made between the overlap quantified as the proportion of similar responses (PSR) and the PIQI. Given a bivariate normal model, the maximum PIQI is shown to be an upper bound for ½ the PSR. Additionally, the characterization of a conditional PSR allows for the PIQI boundaries to

be developed within subgroups defined over observable covariates so that the subset contribution to treatment heterogeneity may be identified. The possibility of similar boundaries is explored outside the normal model using the skew normal distribution. Furthermore, a bivariate PIQI is developed along with its PSR counterpart to help characterize treatment heterogeneity resulting from a bivariate response such as the efficacy and safety of a treatment.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I first want to thank the faculty in the Department of Statistics at Kansas State University for the opportunity to study and learn under their guidance and supervision. In particular, I would like to express my appreciation to John Boyer and Jim Neill for their support and leadership in the department and for me personally for making this work possible.

A special thanks to my committee Gary Gadbury, Paul Nelson, Weixing Song, and David Renter. Since before I arrived at Kansas State, Paul Nelson's guidance has proven to be sure and steady. His suggestions, thoughts, and comments have helped to direct both this work and my education in general. Foremost, a great thank you to my advisor, Gary Gadbury, for his constant support in countless ways. Our work together has been stimulating, enjoyable, challenging, and fruitful. His efforts have left an indelible impression on my life.

Finally, I want to thank my wife, Annette, whose support makes this work not only possible, but worthwhile.

Chapter 1 - Introduction

> "..,it appears that white sheep and pigs are injured by certain plants, whilst dark- coloured individuals escape." ~ Charles Darwin
>
> "What is food to one to some becomes Fierce poison" ~ Lucretius

The quotations above illustrate that individual differences in response to stimuli or 'treatments' have been the subject of interest throughout recorded history. They further illustrate two kinds of interactions. Darwin points out an interaction in which one type of animal is harmed by a certain treatment whereas other animals are not harmed, but are not necessarily helped. In contrast, Lucretius points out a more dramatic type of interaction in which what is helpful to some is actually harmful to others. More formally, treatment heterogeneity is present when the effect of a treatment, say T, with respect to a reference treatment, R, varies across subsets or individuals in a population. A consequence of this heterogeneity is that different individuals or groups of individuals may respond to treatment in opposite directions, with treatment T having higher efficacy for some and treatment R having higher efficacy for others. At times this form of treatment heterogeneity may be accounted for by group or subset identification. The term qualitative interaction (QI) has been used to describe this condition at the subset level (Peto, 1982). Gail and Simon (1985) developed a test to detect a QI, and when such tests are significant, optimal treatments may differ across subsets (Byar and Corle, 1977).

Taking the idea of subsets to its limit, every person is unique and can be considered a separate subset. Individual qualitative interaction (IQI) is present when at least two individuals respond in opposite direction to treatment. However, since individuals can receive only one

treatment at any given time, T or R, IQIs are more difficult to detect than QIs and, as such, formal tests are nonexistent. Still, the very existence of QIs suggests, as was recorded by Lucretius, IQIs are not only plausible, but probable. Although the number of individuals in a population that may experience an IQI for any given T/R combination may be small relative to the number of individuals in the population, the extremity of this type of treatment heterogeneity motivates the need for further development in this area. The work presented in this dissertation provides a structure and a set of related procedures for both studying and quantifying the presence of IQI within populations.

Notwithstanding the serious nature of IQIs, the challenges related to the study of treatment heterogeneity at the individual level have hindered advancement in this area. Consequently, researchers have focused on studying treatment heterogeneity at the subset level through the detection of subset interaction, of which QI is one type. Simon (1982, p. 474) stated that despite the fact that no two patients are exactly alike "At some point it is necessary to settle upon a target population for whom we are willing to attempt to reach a reliable overall conclusion about therapeutic effects." Currently, the study of subset interaction alone may be too restrictive in light of existing research objectives in areas such as personalized nutrition, health care, and behavioral therapy (Lewis and Burton-Freeman, 2010; Marshall, 1997). For example, Kent and Hayward (2007, p. 1209) report, "There remain important differences between individuals in each treatment group that can dramatically affect the likelihood of benefiting from or being harmed by a therapy." The challenge of evaluating treatment heterogeneity from an individual perspective has been addressed by some. The statement by Senn (2001, p.1479) that personalized care, such as those mentioned above, "May be rather more difficult to realize than has been supposed," is an example. An inherent problem with

quantifying individual treatment heterogeneity is that the response to only T or R may be observed at any given time while the unobserved response is considered missing since it is potentially observable under the alternative treatment (see Rubin, 1983). This makes directly estimating and displaying evidence of treatment heterogeneity at the individual level and, in particular IQIs, impossible without assumptions. These ideas are illustrated in the soy treatment example introduced in the next section.

1.1. Soy Treatment Trial

1.1.1. Description and Initial Results

Allison et al. (2003) published the results on the efficacy of a soy-based meal replacement program to treat obesity, which, among other improvements, indicated a significant reduction in average cholesterol levels for the soy-treatment group T over the reference or control group R. A portion of the data from this trial is given in Table 1.1 including 16 of the original 73 observations (8 from each treatment group). The measured response is base line minus end of trial cholesterol level over the 12 week trial period so positive values represent a decrease in cholesterol in that period. The soy treatment group received both a soy-based meal diet and training to help control the patients' diets, while the patients on R received the nutritional training only. For the data listed in Table 1.1, the first 8 patients were on treatment T and measured by X, while patients 9 through 16 were on treatment R and were measured by Y. Since the first 8 patients did not receive treatment R, their response to R was not recorded. Likewise, the response to T for patients 9 through 16 was not recorded. To emphasize that these values would have been observed had treatment assignment gone the other way, the non-observed or missing data for the 16 patients is indicated by an NA. The average cholesterol reduction for the soy group was 21.75 (mg/dl), while observed average reduction in the

cholesterol level for the reference group was 6.75 (mg/dl). A t-test from the observed data for equal means indicates that on average the soy treatment was significantly ($p - value = 0.044$) more effective than the reference treatment in helping to reduce the average cholesterol level.

Table 1.1: Soy treatment data
This table provides the observed responses to a portion of the soy treatment trial, while at the same time illustrates the difficulty of assessing individual treatment heterogeneity with observable data. The NAs represent the unobserved or missing values that could not be attained since that particular treatment was not given and the question marks indicate that no direct individual comparisons can be made for the individual.

Soy-Treatment Example				
Observed and Missing Responses			Unattainable Individual Comparisons	
Patient	Soy(X)	Ref. (Y)	X-Y	$I(Y>X)$
1	17	NA	?	?
2	41	NA	?	?
3	35	NA	?	?
4	-4	NA	?	?
5	22	NA	?	?
6	48	NA	?	?
7	14	NA	?	?
8	1	NA	?	?
9	NA	1	?	?
10	NA	22	?	?
11	NA	-5	?	?
12	NA	11	?	?
13	NA	31	?	?
14	NA	0	?	?
15	NA	4	?	?
16	NA	-10	?	?
Average	21.75	6.75	?	?
Variance	344	192	?	?

1.1.2. Treatment Heterogeneity Assessment

The question motivated by this work is whether there is any indication of an IQI in the population, given the sample data. That is, despite $\hat{\mu}_X = 21.75$ being significantly larger than $\hat{\mu}_Y = 6.75$, were there any individuals for which $X < Y$? The question marks in Table 1.1 indicate that detection of $X < Y$ is not possible due to the missing data. Likewise the determination of $I(X < Y)$ is not possible, where $I(\cdot)$ is an indicator function used to count the number of IQIs in the sample, i.e. those individuals responding opposite to the average. Note that the increase in cholesterol from patient 4 despite having been treated with T may seem to indicate that this patient may have done better on R. On the other hand, had patient 4 been given R, the increase in cholesterol may have been even more profound. Likewise, the large drop in cholesterol recorded for patient 10 who was on R may have had an even more significant reduction if this patient was treated with T. The important point is that without assumptions there is no way to know whether there exists an IQI, which would be detected by the occurrence of $X < Y$.

Figure 1.1 provides a graphical illustration of the problem created by the missing data and also serves as a backdrop for possible solutions. The observed $x_1, x_2, \cdots, x_8$ values are denoted with open triangles along the x axis, which is displayed as the horizontal dotted line, and the observed $y_9, y_{10}, \cdots, y_{16}$ values are denoted with open circles along the y axis, displayed as the vertical dotted line. The sample averages are plotted with the bullet at the point $(\hat{\mu}_X,\ \hat{\mu}_Y) = (21.75, 6.75)$. Note that the data for the observed values are considered independent of each other since they originate from different individuals, and assignment to treatment was random. However, pairs of responses coming from the 16 individuals would necessitate that X and Y originate from a bivariate distribution since $(x_1, y_1), \cdots, (x_{16}, y_{16})$ would each be considered

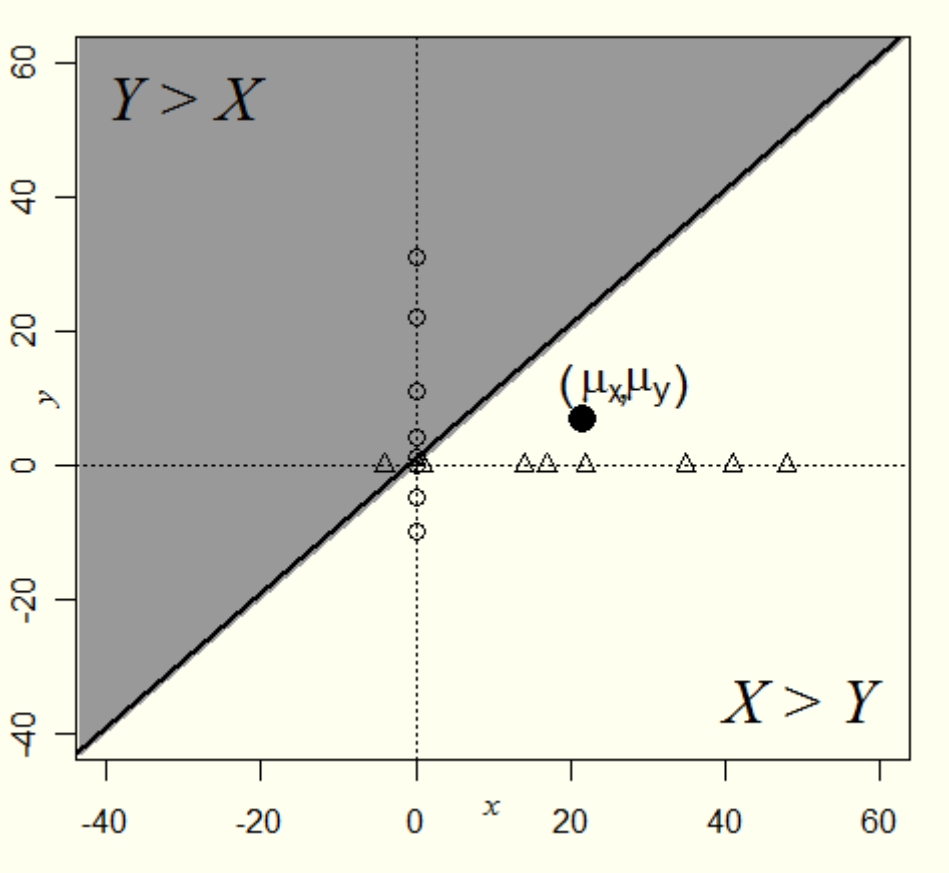


Figure 1.1: Soy treatment data and potential pairs
The observed x values are shown as triangles along the horizontal axis shown as the dotted line, while the open circles are the observed y values are shown as the open circles along the vertical axis. Although no pairs exist, the (x, y) pairs that would have fallen in the shaded region would be considered IQIs. The mean estimates are plotted as $(\hat{\mu}_X,\ \hat{\mu}_Y) = (21.75, 6.75)$.

to have come from the same individuals. Accordingly, under the assumption that X and Y result from a bivariate distribution, the shaded area of Figure 1.1 depicts the area over which $Y > X$. Thus any(x, y) pair in this region provides evidence for an IQI in the population. The challenge of course is finding a suitable match for the observed value in each pair, or to specify the proper bivariate distribution on $(X \quad Y)^T$ since the within individual correlation is not identifiable. Although extreme, the next section discusses two possible solutions, which illustrate some important properties of treatment heterogeneity.

1.1.3. Solution by Boundaries

Suppose $(X \quad Y)^T$ follows a bivariate normal distribution with $\sigma_X = \sigma_Y$. Then by taking $\rho_{XY} = -1$ the missing values may be imputed by

$$x_{i,impute} = \hat{\mu}_X + \hat{\mu}_Y - y_i \quad \text{and} \quad y_{i,impute} = \hat{\mu}_Y + \hat{\mu}_X - x_i. \tag{1.2}$$

Table 1.2 (a) displays the imputed values in bold using (1.2). Note that the sample averages do not change under this strategy. Furthermore, a t-test for equal means is still significant $(p - value = 0.006)$. However, the amount of treatment heterogeneity may seem quite large under (1.2) as seen by the large variance of the 16 $x - y$ values, which is equal to 1026. As expected, the high degree of treatment heterogeneity results in a rather large number of IQIs in the data. In fact, 5 out of the 16 individuals (approximately 31%) respond better to R than to T despite the significant effect of T on average. Each of the IQIs are identified as a shaded row, wherein both $X - Y < 0$ and $I(Y > X) = 1$. Even though taking $\rho_{XY} = -1$ is likely extreme, this imputation strategy shows that a test for equality of means, despite being rejected, fails to address questions related to treatment heterogeneity.

Table 1.2 (b) gives the results of an imputation strategy from the opposite extreme by taking $\rho_{XY} = 1$ resulting in

$$x_{i,impute} = \hat{\mu}_X - \hat{\mu}_Y + y_i \quad \text{and} \quad y_{i,impute} = \hat{\mu}_Y - \hat{\mu}_X + x_i. \tag{1.3}$$

Again, the sample averages remain unchanged and a test for equal means is rejected $(p - value = 0.006)$. However, the analysis indicates that no treatment heterogeneity is present as indicated by a zero variance in the 16 reported $x - y$ values. That is, there is a constant effect from T with respect to R, so that every individual receives the same benefit, namely $\hat{\mu}_X, -\hat{\mu}_Y = 15$, from T rather than R. Clearly, there are no IQIs in this imputed data set.

Table 1.2: Assessing treatment heterogeneity by boundaries
Panels (a) and (b) use (1.2) and (1.3), respectively, to impute the missing values found in Table 1.1. In panel (a) both the number of IQIs and the variance in the sample becomes quite large, while in panel (b) both quantities are zero.

(a) $\rho_{XY} = -1$				
Observed and Imputed Responses			Individual Response Comparisons	
Patient	Soy (X)	Ref. (Y)	*X-Y*	*I(Y>X)*
1	17	**12**	**6**	**0**
2	41	**-13**	**54**	**0**
3	35	**-7**	**42**	**0**
4	-4	**33**	**-37**	**1**
5	22	**7**	**16**	**0**
6	48	**-20**	**68**	**0**
7	14	**15**	**-1**	**1**
8	1	**28**	**-27**	**1**
9	**28**	1	**27**	**0**
10	**7**	22	**-16**	**1**
11	**34**	-5	**39**	**0**
12	**18**	11	**7**	**0**
13	**-3**	31	**-34**	**1**
14	**29**	0	**29**	**0**
15	**25**	4	**21**	**0**
16	**39**	-10	**49**	**0**
Average	**21.75**	**6.75**	**15**	**.31**
Variance	**253**	**258**	**1027**	**-**

(b) $\rho_{XY} = 1$				
Observed and Imputed Responses			Individual Response Comparisons	
Patient	Soy (X)	Ref. (Y)	*X-Y*	*I(Y>X)*
1	17	**2**	**15**	**0**
2	41	**26**	**15**	**0**
3	35	**20**	**15**	**0**
4	-4	**-19**	**15**	**0**
5	22	**7**	**15**	**0**
6	48	**33**	**15**	**0**
7	14	**-1**	**15**	**0**
8	1	**-14**	**15**	**0**
9	**16**	1	**15**	**0**
10	**37**	22	**15**	**0**
11	**10**	-5	**15**	**0**
12	**26**	11	**15**	**0**
13	**46**	31	**15**	**0**
14	**15**	0	**15**	**0**
15	**19**	4	**15**	**0**
16	**5**	-10	**15**	**0**
Average	**21.75**	**6.75**	**15**	**0.00**
Variance	**250**	**250**	**0**	**-**

1.1.3. Soy Trial Summary

Although simplified, the soy treatment example illustrates some important points that should be considered in the study of treatment heterogeneity. First, just because there is a statistically significant average effect of T with respect to the average effect of R, a high proportion IQIs may yet be present in a population. This example shows that even though $\mu_X > \mu_Y$ $(p - value = 0.044)$, 30% of the population may exhibit an IQI. Second, to eliminate the presence of treatment heterogeneity in a population, ρ_{XY} must equal 1, which in many circumstances may be just as naïve as considering $\rho_{XY} = -1$. Finally, the study of treatment

heterogeneity at the individual level may be conducted by a study of the variance of individual effects. Note that in this example, it is implied that an individual effect is defined as $X - Y$ (the response of an individual treated with T relative to the individuals' response treated with R). That is, the $Var(X - Y)$ is directly related to the proportion of IQIs in the population. Thus a study of treatment heterogeneity may be conducted by a study of the $Var(X - Y)$.

Admittedly, in this example it may seem reasonable to assume that ρ_{XY} is rather large, since treatment T is simply an enhanced version of treatment R. However, there are reported cases in which the response to treatment T relative to R may have an opposite effect on some individuals. For example, Lader (2011) states that as many as 5% of patients treated with a sedative will become aggressive or even violent after treatment. It may be supposed that such patients would have been calmer had they not taken the sedative, or had taken some other treatment.

Other applications include the measurement of variables related to the side effects of a treatment. For example, it has been reported that soy products may increase the risk of breast cancer or initiate nutrient deficiencies in infants (Fang, Tseng, Daly, 2005; Setchell et al, 1998). Treatment heterogeneity on side effect variables may be evaluated using the techniques presented herein. However, for simplicity the methods presented here are applied to only primary outcome variable(s) of interest. The next section briefly discusses the importance of analyzing IQIs in an even more general setting.

1.2. Additional Applications to IQI

Evaluating the plausible heterogeneity in treatment response, and even more so the proportion of a population with an IQI, has other applications as well. Consider the context of

claims for weight loss products which can often be quite extravagant. The US Federal Trade Commission (FTC) states that "No [weight loss] product will work for everyone," and therefore claims implying that a "product causes substantial weight loss for all users" is a likely sign of fraud (FTC statement). Is there evidence a company could provide to the FTC to show that in their randomized clinical trial (RCT) showing a positive mean effect, the plausible proportion of people who will have an effect less than a threshold τ is negligible? Alternatively, is there evidence that the FTC could muster to show a company that their claim of a universal positive effect is almost certainly untrue despite there being a positive mean effect? The results described herein may help clarify the issues involved when answering these questions.

As a final example, one can imagine applications in legal settings (see, for example, Marchant 2001, 2010). Imagine that a plaintiff (e.g., a consumer) sues a defendant (e.g., a distributor of a drug, food, or pharmaceutical) claiming that use of defendant's product caused a stroke secondary to markedly elevated blood pressure (BP) as a result of using the product. Imagine further that defense experts present evidence that well-designed RCTs show an *average* effect of the product on BP to be less than or equal to zero. Plaintiff's experts reply that there is great interindividual variability in response and even though the average response is less than or equal to zero, some people will be hypersensitive hyper-responders with extreme BP increases. What evidence can the court bring to bear on the question of how probable it is that plaintiff was such a hyper-responder? The first question which must precede this is what evidence is there that hyper-responders in the opposite direction even exist and with what frequency? Again, the techniques presented in this dissertation can provide a plausible range of answers.

1.3. Dissertation Overview

Chapter 2 provides a review of current literature in the area of treatment heterogeneity at both subset and individual levels. The main topics include potential outcomes, subject-treatment interaction, crossover designs, subset and qualitative interaction, and density overlap or the proportion of similar response (PSR) based on two outcome or response variables.

Chapter 3 outlines a new framework from which to study treatment heterogeneity at the individual level. Under a bivariate normal model new connections are made between treatment heterogeneity and the PSR, which show the conditions under which the PSR may be used to represent the boundary for the maximum proportion of IQIs found within a population. Other previously unspecified relationships between the PSR and the IQI are also characterized. Additionally, the PSR is shown to have a conditional counterpart, and the relationship between the PSR and the IQI is shown to hold over subsets defined over observable covariates. This chapter also provides a new parameterization of the variance of individual effects defined by the difference between X and Y using the potential outcomes framework. Based on a known covariate, the variance is shown to decompose into two components; one based on subset affiliation and the other based on individual differences within subsets. Finally, the soy treatment example is revisited using the entire data set and additional insight is gleaned from the data about potential IQIs found within the population and subpopulations.

Chapter 4 explores the use of the PSR/IQI relationship outside the normal model by utilizing a bivariate skew normal distribution. Since the PSR is largely based on the normal probability model, this chapter provides a new structure for evaluating and computing the PSR. Chapter 5 provides a natural extension to chapter 3 by developing the PSR/IQI relationship in a normal model for bivariate continuous outcomes, so that responses measuring both the efficacy

and safety of a treatment may be studied simultaneously. A new development of the bivariate PSR and the bivariate IQI is presented here based on a bivariate response. Finally, a summary and some possible new directions are given in Chapter 6.

Chapter 2 - Literature Review

2.1. Potential Outcomes

A problem in the study of treatment heterogeneity at the individual level, as mentioned in Chapter 1, is the inability to observe an individual's outcome to both treatments T and R simultaneously. In fact, the unobserved response due to a treatment not given is a form of missing data (Rubin, 1983). As illustrated in the soy treatment example given in Section 1.2, one strategy of handling the missing data is to use imputation methods. Potential outcomes provide the structure for a counterfactual argument, which is the identification of an outcome that would have been observed if conditions, such as treatment assignment, would have happened another way. In the potential outcomes framework, the outcome to every treatment assignment is considered whether observed or not. Throughout this proposal X is used to measure an individual's response to treatment T and Y is used to measure an individual's response to the reference treatment R. Thus, using both X and Y for the simultaneous measurement of T and R on the same individual has been called counterfactual (Dawid, 2000). However, since both X and Y are potentially observable before the treatment is assigned, some prefer to use the term potential outcomes (Rubin, 1974).

Even though the observation of both X and Y is not possible, using potential outcomes allows for the description of concepts such as the observation of both X and Y allows for a true effect of T with respect to R on an individual in a two treatment randomized controlled trial. Neyman (1923) introduced the use of this approach to define the "best estimate" for a mean as the average of the 'potential' outcomes, and not just the average of the 'observed' outcomes.

Rubin (1974, 1978) is credited with developing the potential outcomes framework, which is often called Rubin's causal model (Holland, 1986) in honor of his contribution, and because of its impact on causal inference. Only a part of the potential outcome framework will be given here as it relates to the inference of nonestimable parameters defined in the potential outcomes framework. Given (X_j, Y_j) are potential outcome variables in a population, $D_j = X_j - Y_j$ is defined as the "true" individual effect of T opposed to R for the j^{th} individual. Thus, a set of N subjects and their respective differences taken from a population of potential outcomes may be given as

$$\begin{pmatrix} X_1 & - & Y_1 & = & D_1 \\ \vdots & & \vdots & & \vdots \\ X_N & - & Y_N & = & D_N \end{pmatrix}. \tag{2.1}$$

Since only X or Y, not both, can be measured on each individual after the treatment assignment has been made, D is unobservable, and so the actual data will come from the values on the right of

$$\begin{pmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_N & Y_N \end{pmatrix} - \ Treatment\ Assignment\ \rightarrow \begin{pmatrix} X_1 \\ Y_2 \\ Y_3 \\ \vdots \\ X_{N-1} \\ Y_N \end{pmatrix}. \tag{2.2}$$

The fact that not all potential outcomes can be observed has been referred to as the fundamental problem of causal inference (Holland, 1986). The values of (X_j, Y_j) may be thought of as coming from an infinite bivariate population. Furthermore, the distribution of D may be partially defined by the parameters $E(D) = \mu_D = \mu_X - \mu_Y$ and $Var(D) = Var(X - Y) = \sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY}$. The 'best estimate' of μ_D as defined by Neyman is not available, as made clear by (2.2). However, given random treatment assignment, an unbiased estimate of μ_D can be obtained from estimates of both μ_X and μ_Y calculated from the observed data ($\bar{x}$ and $\bar{y}$) on the

right side of (2.2) (Rubin, 1974). Likewise estimates for both σ_X^2 and σ_Y^2 can be calculated from data on the right side of (2.2). Again similar to the soy treatment example, σ_D^2 cannot be estimated since an important component of σ_D^2, ρ_{XY}, is not estimable. Thus, despite the fact that D is unobservable, the potential outcomes framework provides the basis for characterizing the presence and degree of variability in effects across individuals. As noted by Gadbury (2004), treatment heterogeneity within the population is a function of σ_D^2, so that treatment heterogeneity exists when $\sigma_D^2 > 0$ and is nonexistent when $\sigma_D^2 = 0$. Despite the importance of σ_D^2, it cannot be directly estimated due to the afore mentioned fundamental problem of causal inference. So, even though an individual effect can be defined using potential outcomes, the challenge of working with individual effects are that "strong and largely untestable assumptions" about σ_D^2 need to be made (Cox, 1992, p. 296). Note again that these constraints are not unlike the constraints imposed on the imputation models given in $(1.1\ a, b)$. One assumption made throughout this proposal is that individuals' outcomes are not affected by the treatment assigned to other individuals, an assumption called non-interference (Cox, 1958).

Alternatively, X_j and Y_j may be thought of as coming from a finite population of size N. Then $\overline{D} = \overline{X} - \overline{Y}$ may defined as the finite population average. Given random assignment, the observed average difference equal to $\bar{d} = \bar{x} - \bar{y}$, where $\bar{x} = \frac{2}{N}\sum_{i \in T} x_j$ and $\bar{y} = \frac{2}{N}\sum_{i \in C} y_j$, is an unbiased estimate of $\overline{D}$. Similarly, finite population standard deviations given by S_X and S_Y can be estimated using observed data. If the finite population is a random sample of the super-population of D with parameters μ_D and σ_D^2, then *if the potential outcomes were observable*, $\overline{D}$ and $S_D^2 = S_X^2 + S_Y^2 - 2S_X S_Y R_{XY}$, where R_{XY} is the finite population correlation between X and Y, are unbiased and consistent estimators of μ_D and σ_D^2, respectively. Using only observed data, R_{XY}

is not estimable in a finite population for the same reason that ρ_{XY} is not estimable using an infinite bivariate population model.

2.2. Subject-Treatment Interaction

The fact that individuals may interact with particular drug formulations was acknowledged by Hwang et al. (1978), a phenomenon they called subject-by-product interaction in bioequivalence studies, and more recently called subject-by-formulation interaction (e.g. Endrenyi and Tothfolusi, 1998). More generally, Cox (1992) used the term treatment-by-patient interaction and Gadbury et al. (2001) used the term subject-treatment (S-T) interaction to capture the idea that treatment heterogeneity exists at the individual level.

Gadbury et al. (2001) defined a 'true' individual effect as $D = X - Y$ and used this to delineate assumptions about σ_D^2. They show, given (*X, Y*) originate from an infinite bivariate normal distribution defined as

$$\begin{bmatrix} X_j \\ Y_j \end{bmatrix} \sim iid\ N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\right), \tag{2.3}$$

where ρ_{XY} is the correlation between *X* and *Y*, that $\sigma_D^2 = Var(D) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY}$ can be bound by taking $\rho_{XY} = \pm 1$, and estimating all other parameters from the observed data given in (2.2). Furthermore, they show the probability of an individual receiving a harmful effect, or a negative effect, from T, is given by $P(D < 0) = \Phi\left(\frac{-\mu_D}{\sigma_D}\right)$, and bounded by

$$\Phi\left(\frac{-\mu_D}{\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y}}\right) \leq P(D < 0) \leq \Phi\left(\frac{-\mu_D}{\sqrt{\sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y}}\right). \tag{2.4}$$

Note that the upper bound is achieved when $\rho_{XY} = -1$ and the lower bound is achieved when $\rho_{XY} = 1$. When $\rho_{XY} = 1$ and $\sigma_X = \sigma_Y$, a condition which indicates a constant individual effect (Holland, 1986), then $P(D < 0) = 0$. Gadbury and Iyer (2000) provide maximum likelihood

estimates for the parameters $\mu_D, \sigma_X,$ and σ_Y so that large sample confidence intervals can be placed on lower and upper bounds for $P(D < 0)$ using estimates from the observed data.

In addition, Gadbury et al. (2001) showed that using a continuous covariate not affected by the treatment, say *Z*, that augments the potential outcomes (*X, Y*) can be used to reduce the overall variability of individual effects σ_D^2, under the assumption that the distribution of *D* given $Z = z_0$ is normal with conditional mean

$$\mu_{D|Z=z_0} = \mu_Y - \mu_X + (\beta_{XZ} - \beta_{YZ})(z_0 - \mu_Z) \tag{2.5}$$

and conditional variance

$$\sigma_{D|Z}^2 = \sigma_{X|Z}^2 + \sigma_{Y|Z}^2 - 2\sigma_{X|Z}\sigma_{Y|Z}\rho_{XY|Z}. \tag{2.6}$$

β_{XZ} and β_{YZ} in (2.5) are the slope coefficients between *Z* and *X* and *Z* and *Y,* respectively, and $\rho_{XY|Z}$ in (2.6) is the partial correlation of *X* and *Y* given *Z*. The conditional variances, $\sigma_{X|Z}^2$ and $\sigma_{Y|Z}^2$, are allowed to be different across the two treatment groups but are assumed to not depend on the value of *Z*. Coupled with (2.5) and (2.6), Gadbury et al. (2001) showed that

$$\sigma_D^2 = \left(\sigma_{X|Z} - \sigma_{Y|Z}\right)^2 + 2\sigma_{X|Z}\sigma_{Y|Z}\left(1 - \rho_{XY|Z}\right) + (\beta_{XZ} - \beta_{YZ})^2\sigma_Z^2.$$

Therefore, if evidence showed that $\beta_{XZ} \neq \beta_{YZ}$, $\sigma_{D|Z}^2$ may be less than σ_D^2 making it possible to reduce the bounds on $P(D < 0)$ over $Z = z_0$. Thus similar to (2.4) the probability of an IQI for particular values of z_0 may be bounded by

$$\Phi\left(\frac{-\mu_{D|Z=z_0}}{\sqrt{\sigma_{X|Z}^2+\sigma_{Y|Z}^2-2\sigma_{X|Z}\sigma_{Y|Z}}}\right) \leq P(D < 0)_{Z=z_0} \leq \Phi\left(\frac{-\mu_{D|Z=z_0}}{\sqrt{\sigma_{X|Z}^2+\sigma_{Y|Z}^2+2\sigma_{X|Z}\sigma_{Y|Z}}}\right), \tag{2.7}$$

by letting the partial correlation $\rho_{XY|Z}$ be 1 and -1, respectively. Confidence intervals for the bounds on $P(D < 0)_{Z=z_0}$ given in (2.7) can be derived using bootstrap samples from the

observed data or using asymptotic properties of maximum likelihood estimators (cf. Gadbury et al., 2001).

2.3. Cross-over Designs

Perhaps the most straightforward design for studying IQI is a cross-over design. We illustrate the concepts for a two period two treatment cross-over design, assuming no carry over effects but that period effects may vary across individuals. Since each individual is crossed over from one treatment to another after a washout period, an individual treatment effect may seem to be observable. That is, the observed differences are $x_j - y_j$, for $j = 1,2, \dots, n$ and where the two values $x_j,\ y_j$, are observed in different periods. Table 2.1 illustrates an example where $n = 8$. Again, the patients given treatment T are recorded with X and the patients given treatment R are

Table 2.1:
In this example 8 patients are crossed over between T and R. The question marks illustrate that the evaluation of a 'true' individual effect is not possible without assumptions. The naïve effect assumes no effect due to the time variable.

Observed 2×2 Crossover Design							
Observed Responses					Treatment Heterogeneity		
Patient	Time 1		Time 2		True Effect		Naïve Effect
	X	Y	X	Y	$x_{j1} - y_{j1}$	$x_{j2} - y_{j2}$	
1	x_{11}	NA	NA	y_{12}	?	?	$x_{11} - y_{12}$
2	x_{21}	NA	NA	y_{22}	?	?	$x_{21} - y_{22}$
3	x_{31}	NA	NA	y_{32}	?	?	$x_{31} - y_{32}$
4	x_{41}	NA	NA	y_{42}	?	?	$x_{41} - y_{42}$
5	NA	y_{51}	x_{52}	NA	?	?	$x_{52} - y_{51}$
6	NA	y_{61}	x_{62}	NA	?	?	$x_{62} - y_{61}$
7	NA	y_{71}	x_{72}	NA	?	?	$x_{72} - y_{71}$
8	NA	y_{81}	x_{82}	NA	?	?	$x_{82} - y_{81}$

recorded with Y. Note that in Time 1 patients 1 through 4 are given T and patients 5 through 8 are given R and the missing data are recorded with NA. For Time 2 the treatments are reversed. A true check for an IQI would compare the difference between X and Y for each individual within each time. However, since half of the data are missing a direct evaluation of the IQI is not possible as indicated by the question marks. A straightforward, albeit naïve, estimate of the an IQI may be obtained from the last column by comparing the response to T and R for each individual, but across time periods. The proportion of values that are negative provide an estimate of the proportion of IQI present in the population.

Note that the observed intra-individual difference in the outcome measurement between the two periods in the last column of Table 2.1 inseparably includes the actual effect due to the treatment and the effect due to the period in which the outcome to treatment was observed. Hence, this column is labeled as the naïve effect since these two contributors cannot be separated at the individual level without introducing additional assumptions. This can be further illustrated using the potential outcomes structure introduced in Section 2.1. Potential outcome variables are (X_1, Y_1) at time 1 and (X_2, Y_2) at time 2. There are two "true" individual treatment effects given by the variables, $D_1 = X_1 - Y_1$ at time 1 and $D_2 = X_2 - Y_2$ at time 2. In some applications it may be D_1 that is the effect of most interest. Another effect that may be of interest, and the one considered here, is the average over the two time periods, denoted as $D = (D_1 + D_2)/2$. Once again, $E(D) = \mu_D$ is usually of interest but here we consider the available information to estimate $Var(D) = Var(X - Y) = \sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY}$.

Assume there is a population model with density $f(x_1, y_1, x_2, y_2)$. The joint density of the 'full data' sample of *n* subjects (i.e., of the sample of potential outcomes) is $\prod_{i=1}^{n} f(x_{i1}, y_{i1}, x_{i2}, y_{i2})$. Assume that n_1 subjects are assigned to the sequence, *TR*, where *TR*

implies treatment *T* at time 1 and treatment *R* at time 2, and n_2 subjects to the reverse sequence of treatments, *RT*, with $n_1 + n_2 = n$. The density of observed outcome variables is of the form, $\prod_{i \in TC} f(x_{i1}, y_{i2}) \prod_{j \in CT} f(x_{j2}, y_{j1})$. A large body of literature exists on estimating mean treatment effects, mean period effects, carry-over effects, etc. (e.g., Senn 2006a; Yang and Stufken 2008). Mixed-effects models fit to data from a cross-over design with a random subject effect may even compute a "subject-treatment interaction variance" (e.g., Hauck et al. 2000; Endrenyi and Tothfalusi 1999). However, this variance computed from observed data may not be the same as the variance, σ_D^2 above, without certain assumptions. Insight into this is given by the correlation matrix of the potential outcomes

$$CORR(X_1, Y_1, X_2, Y_2) = \begin{bmatrix} 1 & \rho_{X_1Y_1} & \rho_{X_1X_2} & \rho_{X_1Y_2} \\ \rho_{X_1Y_1} & 1 & \rho_{Y_1X_2} & \rho_{Y_1Y_2} \\ \rho_{X_1X_2} & \rho_{Y_1X_2} & 1 & \rho_{X_2Y_2} \\ \rho_{X_1Y_2} & \rho_{Y_1Y_2} & \rho_{X_2Y_2} & 1 \end{bmatrix}.$$

Note the observed data only allows for direct estimation of $\rho_{X_1Y_2}$ and $\rho_{X_2Y_1}$. Thus, in the absence of knowledge (or assumptions) regarding the other four correlations, neither σ_D^2 nor the level of IQI can be directly evaluated using observed data.

If the *Balaam* design (Balaam 1968) is used, where some subjects remain on the same treatment over the two periods (i.e., *TR*, *RT*, *TT*, and *RR* sequences), then $\rho_{X_1X_2}$ and $\rho_{Y_1Y_2}$ can be estimated so that four of the six correlations are directly estimable from observed data, although direct estimation of σ_D^2 is still not possible. However, assuming for instance that the correlations $\rho_{X_1Y_1} = \rho_{X_1Y_2}$ and $\rho_{X_2Y_2} = \rho_{X_2Y_1}$, then σ_D^2 becomes estimable making direct estimation of the proportion of IQIs possible, with distributional assumptions on *D*. Without assumptions regarding the two nonestimable correlations, mathematical bounds for σ_D^2 can be estimated in

observed data - making use of the positive definiteness requirement of the four dimensional correlation matrix for the full data model.

Required assumptions for direct estimation of σ_D^2 may be more plausible in certain applications than assumptions that are required without the multiple period feature of the design. As mentioned by others, repeated measures cross-over designs have advantages over single period designs for estimating subject-treatment interaction and its consequences (e.g., Senn 2001). More methodological development is needed to define the required assumptions and resulting estimators from different types of cross-over designs, and potential outcomes may be the best structure to use when doing this.

Consequently, cross-over designs can potentially provide more information regarding treatment heterogeneity than designs where only a single outcome is observed per individual. However, cross-over designs are not always practical to implement in many applications. For example, the designs presented here assume no carry over effects (cf. Brown, 1980). Other conditions often required for crossover designs may be whether the treatment is reversible, whether the disease is chronic, whether a four period crossover is practical, and/or whether the risk of carry-over can be reasonably achieved (cf. Senn, 2001). Cross-over designs are not further considered here.

2.4. Qualitative Interactions

Interaction between subsets, or simply subset interaction (SI), is a form of treatment heterogeneity. This type of treatment heterogeneity has been an important element in the discovery of "what treatment is most appropriate for what kinds of patients," which, according to Simon (1982, p. 473) is "the overriding objective of clinical research." Thus, the importance of methods such as those introduced by Byar and Corle (1977) for selecting covariates, on which to

base subsets, becomes apparent. Generally, SI occurs when the effects of T and/or R change based on the subset identifiable by an observable covariate (Milliken and Johnson, 1984, p. 113). Previous to Peto (1982) interaction effects were defined only in terms of the magnitude of the change over subsets, irrespective of whether the effects of T were greater than the effects of R across all subsets. Peto's delineation of SI into two categories, which he termed quantitative if the effects of T remain larger than the effects of R across all subsets and qualitative if the most effective treatment changes across subsets, made a significant contribution to the study SI. Quantitative SI may be thought of as a less extreme form of SI and qualitative SI, a more severe form. Gail and Simon (1985) used the expressions non-crossover interaction and crossover interaction as opposed to quantitative and qualitative interaction, respectively, and developed tests for each based on large sample theory. The important contributions of Peto and Gail and Simon made possible tests to determine whether a treatment T, which is beneficial on average as opposed to R, may be harmful to some patients. In addition to Gail and Simon, Silvapulle (2001) developed a test for QI based on finite sample theory, and Li and Chan (2006) proposed a potentially more powerful test than Gail and Simon's test for detecting QIs, which they called the extended range test.

Qualitative and quantitative interactions, including their respective tests, can be applied to populations that include any number of subsets. However, to illustrate the difference between qualitative and quantitative interactions, suppose there is a population with only two subsets s_1 and s_2. Consider a population in which each individual within a subset is treated with either treatment T or control R, and the true mean outcomes to each treatment, within the subsets, are given by $\mu_{T(s_i)}$ and $\mu_{R(s_i)}$, for $i = 1, 2$. These means, as defined here, refer to the mean response if all subjects in a subset were to receive T or R, and so could represent finite subpopulation

means. The true mean differences are given by $\delta_i = \mu_{T(s_i)} - \mu_{R(s_i)}$. If $\delta_1 = \delta_2$, then there is no SI, and the population treatment effect is homogenous with respect to the subsets. Populations in which both $\delta_1 \geq 0$ and $\delta_2 \geq 0$, or both $\delta_1 \leq 0$ and $\delta_2 \leq 0$ when $\delta_1 \neq \delta_2$ are considered to have a quantitative interaction. A qualitative interaction (QI) introduced in Chapter 1 is present when either $\delta_1 > 0$ and $\delta_2 < 0$, or $\delta_1 < 0$ and $\delta_2 > 0$ is true. This was illustrated in a figure by Gail and Simon (1985, p 362) that is depicted as Figure 2.1 here. QIs exist when the point (δ_1, δ_2) lies in quadrants II and IV of Figure 2.1 and quantitative interactions exist when the δ_1 and δ_2 pairs are in the I and III quadrants of Figure 2.1, with the exception of the line where $\delta_1 = \delta_2$.

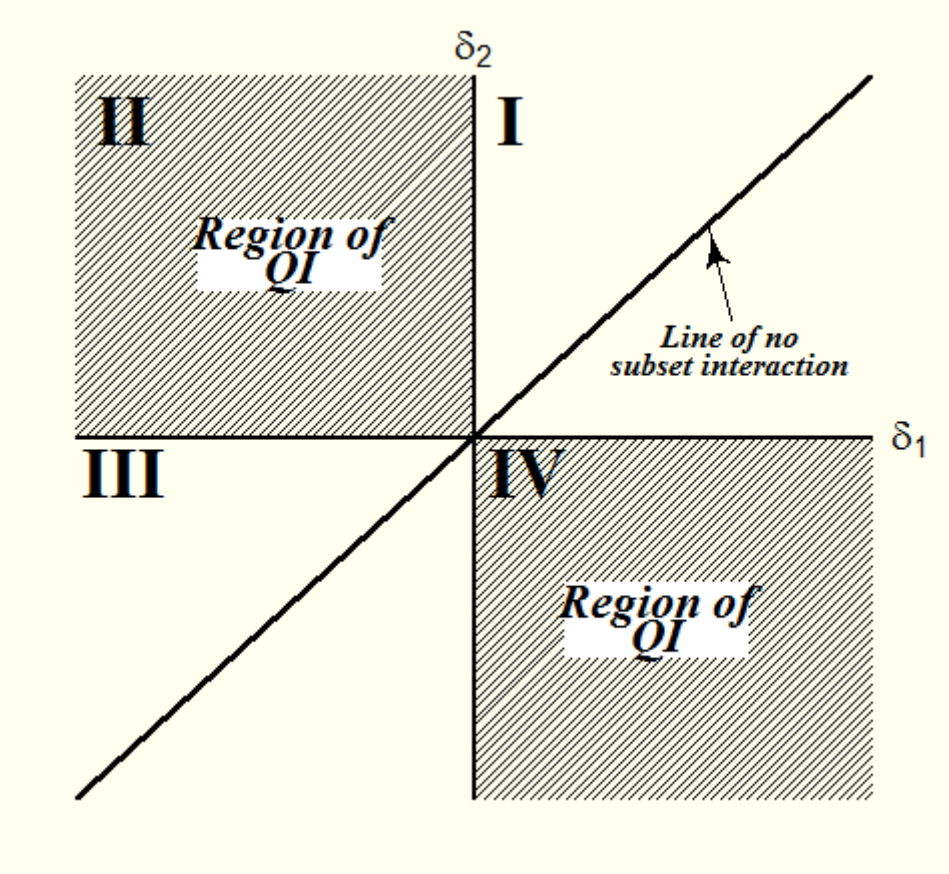


Figure 2.1: Qualitative vs. Quantitative Interaction
δ_1 and δ_2 represent the mean difference between two treatments in subsets 1 and 2, respectively. Subset qualitative interaction takes place when δ_1 and δ_2 occur in quadrants II and IV. Subset quantitative interaction takes place when δ_1 and δ_2 occur in quadrants I and III, except where $\delta_1 = \delta_2$, where no interaction occurs between subsets.

2.5. The Proportion of Similar Responses

The proportion of a similar responses (PSR), sometimes referred to as the overlapping coefficient (Inman and Bradley, 1989), measures the area of overlap between the probability density functions (pdf's) of X and Y. Although the term PSR was not used until 1996 (Rom and Hwang, 1996) the characterization of distributional similarities (or differences) as the overlap of their respective density curves has been used for more than a century. According to Inman and Bradley (1989), Pearson (1895) used a measurement of non-overlap, which he called "ariel deviations," previous to the development of the Pearson Chi-square goodness-of-fit test. A paper published by Tilton (1937) made important connections between the density overlap measurement and the idea of 'sigma scores' for normal distributions. Tilton defined a sigma score as the difference in the means of the overlapping distributions relative to their common standard deviation. In this sense, there is a deterministic relationship between the percentage of overlap and the sigma score. Cohen (1969) later used the sigma score to determine the 'effect size', or the difference between means in a population with respect to a common standard deviation.

Rom and Hwang (1996) suggested that the density overlap of X and Y may be related to individual treatment heterogeneity, and introduced the PSR terminology. They observed that for two normal distributions, the greater the difference between the variances of X and Y, the more individual treatment heterogeneity may be present in the population. The connection between the PSR and treatment heterogeneity is that, when treatment variances are unequal, the density overlap must be less than 100%.

As an illustration, the entire shaded area in Figure 2.2 graphically portrays the PSR of X and Y when $\mu_X > \mu_Y$ and $Var(Y) > Var(X)$. The measurement of overlap is given by

$$\text{PSR} = \int min\big(f_X(u), f_Y(u)\big)\, du, \tag{2.8}$$

where $f_X(x)$ and $f_Y(x)$ are the pdf's of the outcome variables X and Y to treatments T and R, respectively. The calculation of the PSR depends on values of x such that $f_X(x) = f_Y(x)$. If the two distributions are equal, $f_X(x) = f_Y(x)$ for all x and the PSR is equal to 1. If $f_X(x) \neq f_Y(x)$ for any x, the PSR is equal to 0. Given that both X and Y follow normal distributions, Inman and Bradley (1989) show that for the most general case under the given constraints, which is where $\sigma_X^2 \neq \sigma_Y^2$, there are two points, a lower point and an upper point denoted by x_L and x_U, where $f_X(x) = f_Y(x)$, so that $f_X(x_L) = f_Y(x_L)$ and $f_X(x_U) = f_Y(x_U)$. Otherwise, $f_X(x) \neq f_Y(x)$. Furthermore, they show x_L and x_U result from

$$\frac{\mu_Y\sigma_X^2 - \mu_X\sigma_Y^2 \pm \sigma_Y\sigma_X\left[(\mu_Y - \mu_X)^2 + (\sigma_X^2 - \sigma_Y^2)\ln\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)\right]^{\frac{1}{2}}}{\sigma_X^2 - \sigma_Y^2}. \tag{2.9}$$

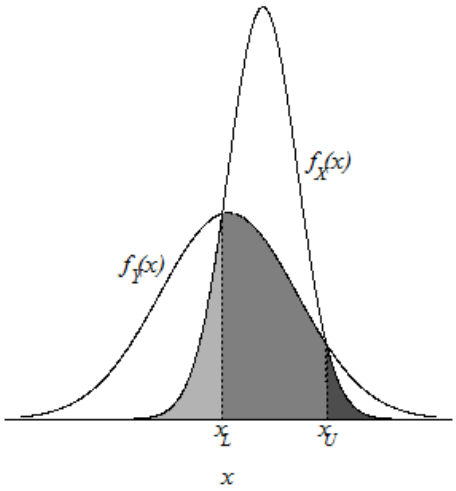


Figure 2.2: An illustration of the PSR with unequal variances
The PSR is the shaded area displayed as the overlap between the pdf's of two normally distributed outcome variables X and Y, where $\boldsymbol{\mu_X > \mu_Y}$, which can be calculated using $(\mathbf{2.8})$. Since $\boldsymbol{\sigma_X^2 < \sigma_Y^2}$, $\boldsymbol{f_X(x) = f_Y(x)}$ in exactly two places $\boldsymbol{x_L}$ and $\boldsymbol{x_U}$, as referenced by the dotted lines below the crossing points of the densities, where $\boldsymbol{x_L}$ and $\boldsymbol{x_U}$ can be calculated using $(\mathbf{2.9})$

Figure 2.2 illustrates the case for $\mu_X > \mu_Y$ and $\sigma_X^2 < \sigma_Y^2$, indicating treatment T is beneficial on average. The dashed vertical lines in Figure 2.2 show the points of equality directly over the points x_L and x_U. The PSR is equal to the shaded area represented by the overlap between $f_X(x)$ and $f_Y(x)$ and, in addition to (2.8), under the distributional assumptions that *X* and *Y* follow normal distributions, the PSR may be calculated by adding the three probabilities in equation (2.10) given here

$$\text{PSR} = P(X \leq x_L) + P(x_L \leq Y \leq x_U) + P(X \geq x_U), \tag{2.10}$$

and illustrated in Figure 2.2 by adding the three shaded regions separated by the dotted lines. When $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, there is exactly one point of equality denoted by $x_E = \frac{\mu_X + \mu_Y}{2}$, where $f_X(x_E) = f_Y(x_E)$ and the

$$\text{PSR} = 2 \times \Phi\left(\frac{-|\mu_X - \mu_Y|}{2\sigma}\right). \tag{2.11}$$

Figure 2.3 illustrates this case.

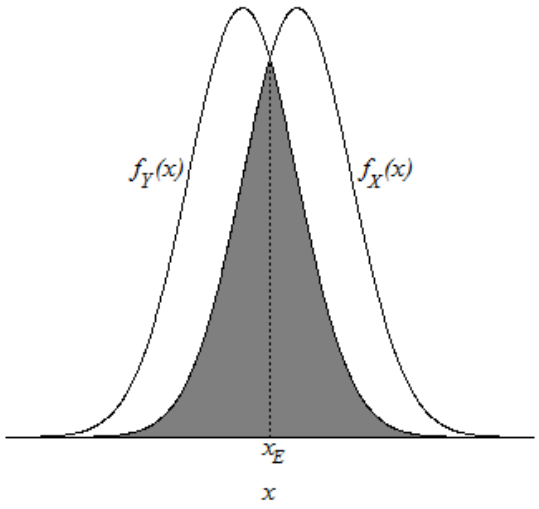


Figure 2.3: An illustration of the PSR with equal variances
The PSR is the shaded area displayed as the overlap between the *pdf's* of the normally distributed outcome variables *X* and *Y*, where $\mu_X > \mu_Y$. Since $\sigma_X^2 = \sigma_Y^2$, $f_X(x_E) = f_Y(x_E)$, and otherwise $f_X(x) \neq f_Y(x)$. The PSR can be calculated using (2.11) and x_E can be found by taking the average of μ_X and μ_Y.

Stine and Heyse (2001) propose a kernel density estimate of the PSR as an alternative to using normal densities. The kernel density estimates for $f_X(x)$ and $f_Y(x)$ take on the form

$$\hat{f}_{X_K}(x) = \frac{1}{n_x}\sum_{i=1}^{n_x}\frac{1}{h_x}K\left(\frac{x - X_i}{h_x}\right) \text{ and } \hat{f}_{Y_K}(x) = \frac{1}{n_y}\sum_{i=1}^{n_y}\frac{1}{h_y}K\left(\frac{x - Y_i}{h_y}\right), \tag{2.12}$$

respectively, where h is the bandwidth, n is the sample size, and $K(\cdot)$ is the kernel. Subscripts are used to indicate that these values may be unique to the particular density of *X* or *Y*. The PSR is then approximated numerically by

$$\widehat{\text{PSR}} = \int min\left(\hat{f}_{X_K}(x), \hat{f}_{Y_K}(x)\right) dx. \tag{2.13}$$

Typically, the bandwidth is the important parameter in a kernel density estimate (cf. Sheather, 2004), and simulations performed by Stine and Heyse show this holds with kernel density estimates of the PSR. Intuitively, the area of overlap shrinks as the bandwidth shrinks. However, Stine and Heyse show that when the data are normal or mixed normal, the normal reference bandwidth given as $h_N = 1.06\sigma n^{-1/5}$, where σ is estimated from the data, produces kernel estimates of the PSR comparable to PSR estimates generated from maximum likelihood estimation that assumes normal distributions. Thus they suggest using h_N or a multiple of h_N when the data are considered to be close to normal.

Since the overlap of the density curves provides a natural way to think about individual effects, the PSR has been used as a proxy for the proportion of IQIs in a population given treatments T and R (see Rom and Hwang, 1996; Stine and Heyse, 2001). Gastwirth (1975) noted that the maximum values for the PSR and the $P(X < Y) = P(D < 0)$, which are 1 and 0.5, respectively, have the same interpretation. Additionally, the overlap seems to suggest that as the PSR increases, the potential for a value from $f_X(x)$ to be less than a value from $f_Y(x)$ also

increases. As such, the PSR appears to be a sensible measurement to study IQIs. Furthermore, they enjoy the added benefit of using either parametric or nonparametric kernel density estimation of the two densities (e.g. Stine and Heyse, 2001).

However, in an assessment of the PSR, Senn (2006, pp. 3944-3945) points out, "If every patient benefits by having his or her outcome improved by the same amount [under treatment T] compared to what it would have been [under treatment R], then 100 percent of the patients have benefited" (brackets added to provide context). Thus Senn identifies what is clear using the soy treatment example (see Table 1.1 (b)), which is, if D is a constant, then the number of IQIs in a population would always be equal to 0, even when the PSR > 0. Furthermore, Gastwirth (1975) questioned the utility of the PSR because values of x from either $f_X(x)$ or $f_Y(x)$ that are in the interval $x_L < x < x_U$ are free to move within the interval so long as x_L and x_U remain unchanged. Inman and Bradley (1989, p.3871) further identified the PSR's weakness by stating that "The magnitude of the overlap in itself does not indicate where the common probability mass is located." This concern was echoed by Stine and Heyse (p. 232) in which they concede that the PSR "Does not consider where differences occur. If these locations are important, one might be better served by measuring the overlap differently." This is an important point since treatment heterogeneity is a function of the variance of $X - Y$. Even under the normality assumption, the PSR is limited as a measurement treatment heterogeneity since the PSR has no inherit mechanism for evaluating the correlation between X and Y. It uses only the marginal densities. Thus, despite the apparent usefulness of the PSR, its utility cannot be fully realized for studying individual treatment heterogeneity or, more particularly IQIs, until a clear probabilistic meaning can be given for the area of the overlap.

Chapter 3 - The Probability of an IQI and the PSR

The bulk of the ideas and methods presented in the rest of this dissertation constitute new research. Those that are not new are only given to provide context. The main ideas and results presented in Chapter 3 have been packaged in a paper that is being revised for *The American Statistician* (Poulson, Gadbury, and Allison, 2010). Concepts discussed in Chapters 1 and 2 are brought together under the potential outcomes framework in Chapters 3, 4, and 5. Table 3.1 provides a quick reference to some terms, their definitions, and their abbreviations that have been given in previous chapters, and also others that will yet be developed and used throughout the remainder of this dissertation.

Table 3.1: Important terms, abbreviations, and definitions

Term	Abbreviation	Definition
Treatment Heterogeneity	n/a	When the effect of T with respect to R changes over either subsets or individuals in a population
Subset Interaction	SI	When the average effects of T with respect to R changes based on the group.
Subset Qualitative Interaction	QI	When the most effective average response to treatment T or R depends on the subset.
Subset Quantitative Interaction	Non-QI	When there is SI, but T remains the most effective treatment over all subsets.
Subject-Treatment Interaction	S-T Interaction	When the effect of T with respect to R changes based on the individual.
Individual Qualitative Interaction	IQI	When the most effective response to treatment T or R depends on the individual.
Individual Quantitative Interaction	Non-IQI	When there is S-T Interaction, but T remains the most effective treatment over all individuals.
Probability of an Individual Qualitative Interaction	PIQI	The probability that a selected individual in the population or subpopulation will respond opposite to the response suggested by the average effects.

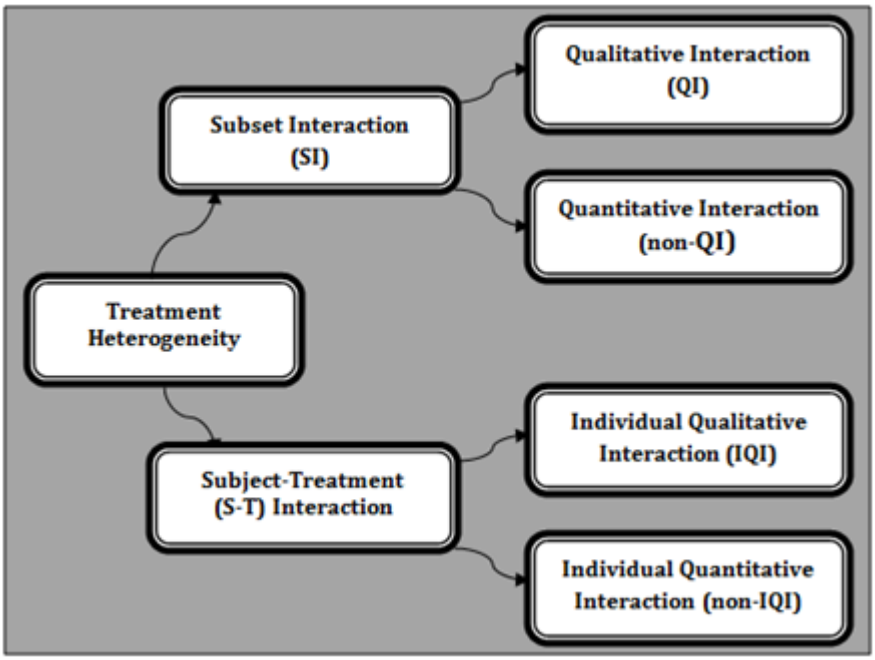
3.1. Subset and Individual Qualitative Interaction

The study of treatment heterogeneity at the individual level is synonymous with S-T interaction, which is characterized by σ_D^2. As outlined in Sections 1.1 and 2.2, S-T interaction exists when $\sigma_D^2 > 0$, and is non-existent with $\sigma_D^2 = 0$. However, a component-wise delineation of S-T interaction helps to formulate the components of σ_D^2 as the proportion explained by SI and the proportion explained by S-T interaction. Recall from Section 2.4 that SI may include QI and/or quantitative interaction (non-QI), with QI being the more severe type of SI. Figure 2.1 was used to provide a visual illustration of the differences between QI and non-QI. Recall from Section 2.4 that, given a population of individual units and $i = 1,2, \dots, g$ groups, there are subsets denoted $s_1, s_2, \dots, s_g$ in which each individual in a subset is given either treatment T or R. Thus, each subset is comprised of two means $\mu_{T(s_i)}$ and $\mu_{R(s_i)}$, and the differences are represented by $\delta_i = \mu_{T(s_i)} - \mu_{R(s_i)}$. Given two subsets, $i = 1,2$, QI exists when the point (δ_1, δ_2) lies in quadrants II and IV of Figure 2.1 and non-QI exists when the δ_1 and δ_2 pairs are in the I and III quadrants of Figure 2.1, with the exception of the line where $\delta_1 = \delta_2$.

S-T interaction exists when the effects of T with respect to R change based on the individual. Similar to SI, S-T interaction may include both individual qualitative interaction (IQI) and/or individual quantitative interaction (non-IQI), where IQI is the more severe form of S-T interaction. Describing these types of interactions at the level of individuals is analogous to the description of SI except each subset contains only one individual . Thus, the mean outcomes to treatments T and R given in Section 2.4 for subset i given as $\mu_{T(s_i)}$ and $\mu_{R(s_i)}$ are now replaced by potential outcomes for the i^{th} individual, (X_i, Y_i), where X_i is the i^{th}outcome to treatment T and Y_i the i^{th} outcome to treatment R. These individual effects are unobservable because either X or Y, not both, can be measured on each individual after the treatment

assignment. Utilizing the potential outcome framework, the S-T interaction can be defined similarly to SI by letting $D_i = X_i - Y_i$ replace δ_i. Thus, similar to QI, IQI exists when any two individuals in the population have treatment effects in opposite directions. Formally, IQI is present when if for any i and i' $D_i > 0$ and $D_{i'} < 0$ or $D_i < 0$ and $D_{i'} > 0$. Furthermore, non-IQI exists when $D_i \neq D_{i'}$, but either $D_i > 0$ or $D_i < 0$ over all i. Again, by restricting $i = 1, 2$ as in Section 2.4, Figure 2.1 can be used to display the components of S-T interaction. By letting $D_1 = X_1 - Y_1$ and $D_2 = X_2 - Y_2$ be the individual treatment effects for two individuals, the IQI is illustrated by the same quadrants for a QI in Figure 2.1 with the axes labels replaced by D_1 and D_2. Quadrants and I and III (except for the "constant effect" line where $D_1 = D_2$) indicate S-T interaction but would not be an IQI. Now that each component of treatment heterogeneity has been defined, Figure 3.1 illustrates how treatment heterogeneity may be partitioned within a population.

Figure 3.1: Delineation of treatment heterogeneity



3.2. IQI Bounds Based on the PIQI

For subsets, QIs are of particular interest because treatments T and R have effects in opposite directions across subsets. Likewise, IQI tends to be of particular interest with respect to individuals since T with respect to R have opposite effects across individuals. For clarity of illustration, and without loss of generality, it will be assumed that the average outcome from T is larger than the average outcome from R so that $\mu_X > \mu_Y$. Furthermore, let both X and Y follow a bivariate normal distribution and let $\sigma_X^2 = \sigma^2$, and $\sigma_Y^2 = k^2\sigma^2$ for the remainder of this chapter. Thus, $\mu_D = \mu_X - \mu_Y > 0$ and $\sigma_D^2 = \sigma^2(1 + k^2 - 2k\rho_{XY})$, since $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY} = \sigma^2 + k^2\sigma^2 - 2\sigma k\sigma\rho_{XY}$. Consequently, if the bivariate normal model given in (2.3) is assumed, the proportion of the population experiencing a negative effect of treatment T with respect to R defined by Gadbury (2004) can now be defined as the probability of an individual qualitative interaction (PIQI) given by

$$\text{PIQI} = P(D < 0) = \Phi\left(\frac{-\mu_D}{\sigma_D}\right). \tag{3.1}$$

However, since X and Y are potential outcomes, the PIQI cannot be calculated due to ρ_{XY} being inestimable. The impact of choosing a suitable value for ρ_{XY} may be illustrated graphically by a return to the soy treatment example introduced in Chapter 1. Figure 3.2 panel (a) is duplication of Figure 1.1. Recall that the $(X \quad Y)^T$ pairs that fall in the shaded region represent individuals that exhibit an IQI. However, despite the 16 observed values, the only value that can be plotted is the mean pair $(\hat{\mu}_X,\ \hat{\mu}_Y) = (21.75, 6.75)$ indicated with the solid 'bullet.' Thus without assumptions there would be no way of estimating the PIQI despite its form given in (3.1). Specifying (2.3), assuming $(\mu_X,\ \mu_Y) = (21.75, 6.75)$ and $\sigma_X^2 = \sigma_Y^2 = s_p^2 \cong 250$, one may think

of the PIQI as random draws from this bivariate distribution. That is, the probability of a random $(X \quad Y)^T$ pair falling in the shaded region constitutes the PIQI for this model specification.

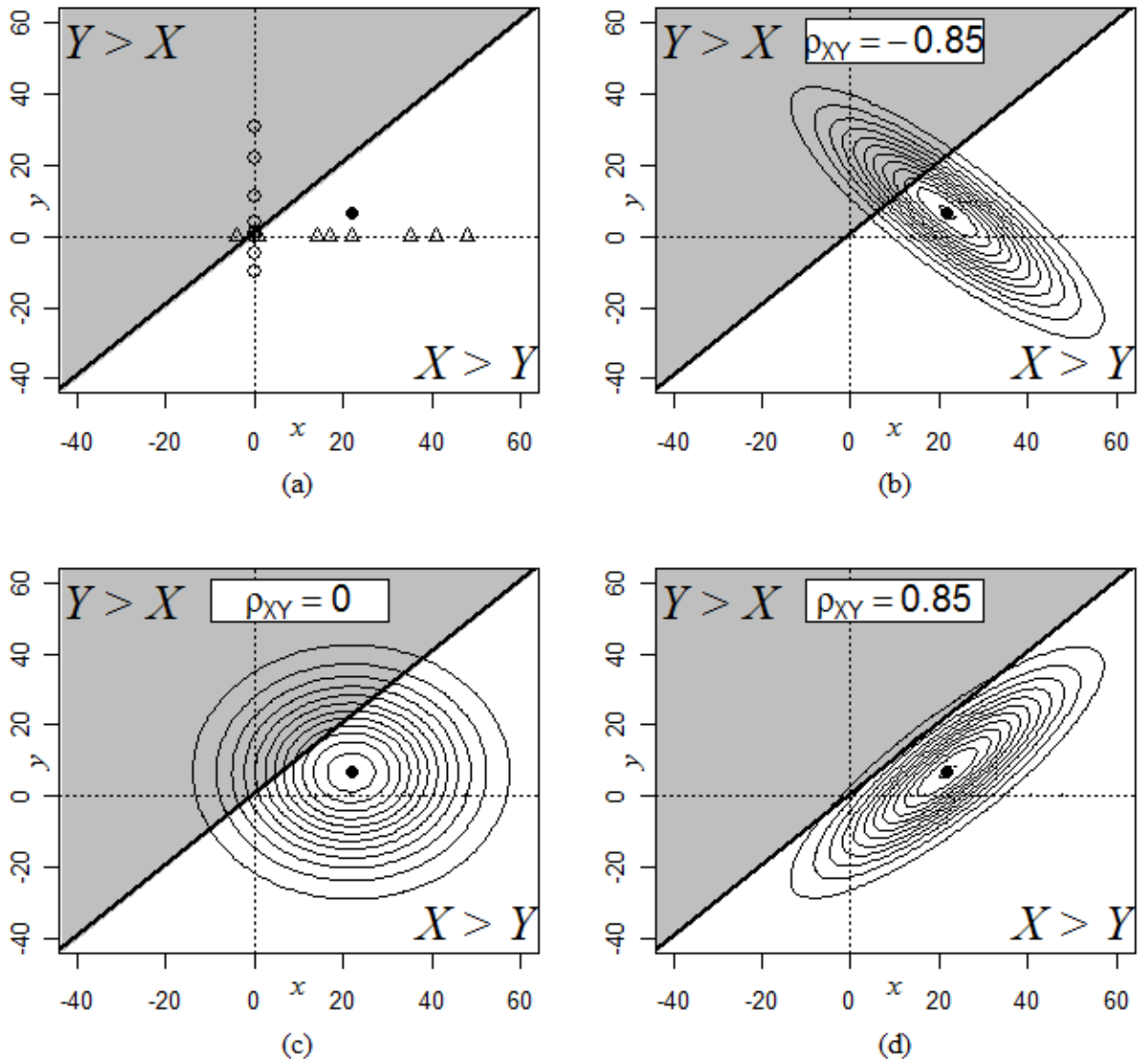


Figure 3.2: A PIQI illustration using soy data
This figure illustrates the development of the PIQI from the perspective of bivariate normal distribution. The PIQI is the probability of an $(X, Y)^T$ being drawn from the bivariate distributions of $(X \quad Y)^T$. The shape of the distribution of $(X \quad Y)^T$, and therefore the PIQI, is dependent upon the unidentifiable value of ρ_{XY}.

However, the PIQI is dependent on ρ_{XY}. If X and Y have a strong inverse relationship such as $\rho_{XY} = -0.85$, then the PIQI may be rather large as can be seen by the proportion of the

bivariate density in the shaded region shown in Figure 3.2 panel (b). Even if $\rho_{XY} = 0$, panel (c) shows that the PIQI may be rather larger than expected based on the area of the $(X \quad Y)^T$ density that falls in the shaded region. In fact, even under $\rho_{XY} = 0.85$ as panel (d) shows, the PIQI > 0 ($\text{PIQI}^{\rho_{XY}=0.85} \cong 0.042$ computed from (3.1)).

The value of ρ_{XY} that maximizes the PIQI is $\rho_{XY} = -1$ under (2.3), which means that $(X \quad Y)^T$ is a degenerate bivariate distribution and the (X, Y) pairs fall on the line defined a slope of -1 that passes through the point $(\mu_X, \ \mu_Y) = (21.75, 6.75)$. Likewise, the PIQI is minimized when $\rho_{XY} = 1$. Consequently, similar to (2.7), minimum and maximum bounds for the PIQI can be expressed by

$$\text{PIQI}^{min} = P(D < 0)_{\rho_{XY}=1} = \Phi\left(\frac{\mu_Y - \mu_X}{\sigma\sqrt{1 + k^2 - 2k}}\right) \tag{3.2}$$

and

$$\text{PIQI}^{max} = P(D < 0)_{\rho_{XY}=-1} = \Phi\left(\frac{\mu_Y - \mu_X}{\sigma\sqrt{1 + k^2 + 2k}}\right), \tag{3.3}$$

respectively. As shown in (3.2) and (3.3) the calculation of the PIQI is a function of ρ_{XY}, as expected, and as ρ_{XY} becomes larger the PIQI gets smaller. Figure 3.3 approaches the PIQI from the univariate distribution of D rather than from the bivariate distribution of X and Y shown in Figure 3.2. Figure 3.3 panel (a) displays the marginal distributions of X and Y for the case where $\mu_X = 1, \mu_Y = 0,$ and $\sigma_X^2 = \sigma_Y^2 = 1$. Their respective overlap is represented by the shaded area. Two potential distributions of $D = X - Y$ are given in panels (b) and (c) of Figure 3.3 for the different values of ρ_{XY} equal to -1 and 0.5, respectively. These two graphs demonstrate how the PIQI gets smaller as ρ_{XY} gets larger from the perspective of the distribution of D. Note that panel (b) gives $\text{PIQI} = \text{PIQI}^{max} \cong 0.31$ and represents the maximum value of $P(D < 0)$. Not shown is the case for $\rho_{XY} = 1$, since in this particular case $k = 1$ (i. e., $\sigma_X^2 = \sigma_Y^2$) so that the distribution of

D is degenerate and the PIQI $= 0$. Panel (d) of Figure 3.3 shows that the PIQI, depicted with the dashed line, is a strictly decreasing function of ρ_{XY}. For reference, the solid line in panel (d) is the value of the ½ PSR shown in panel (a), which is not a function of ρ_{XY}.

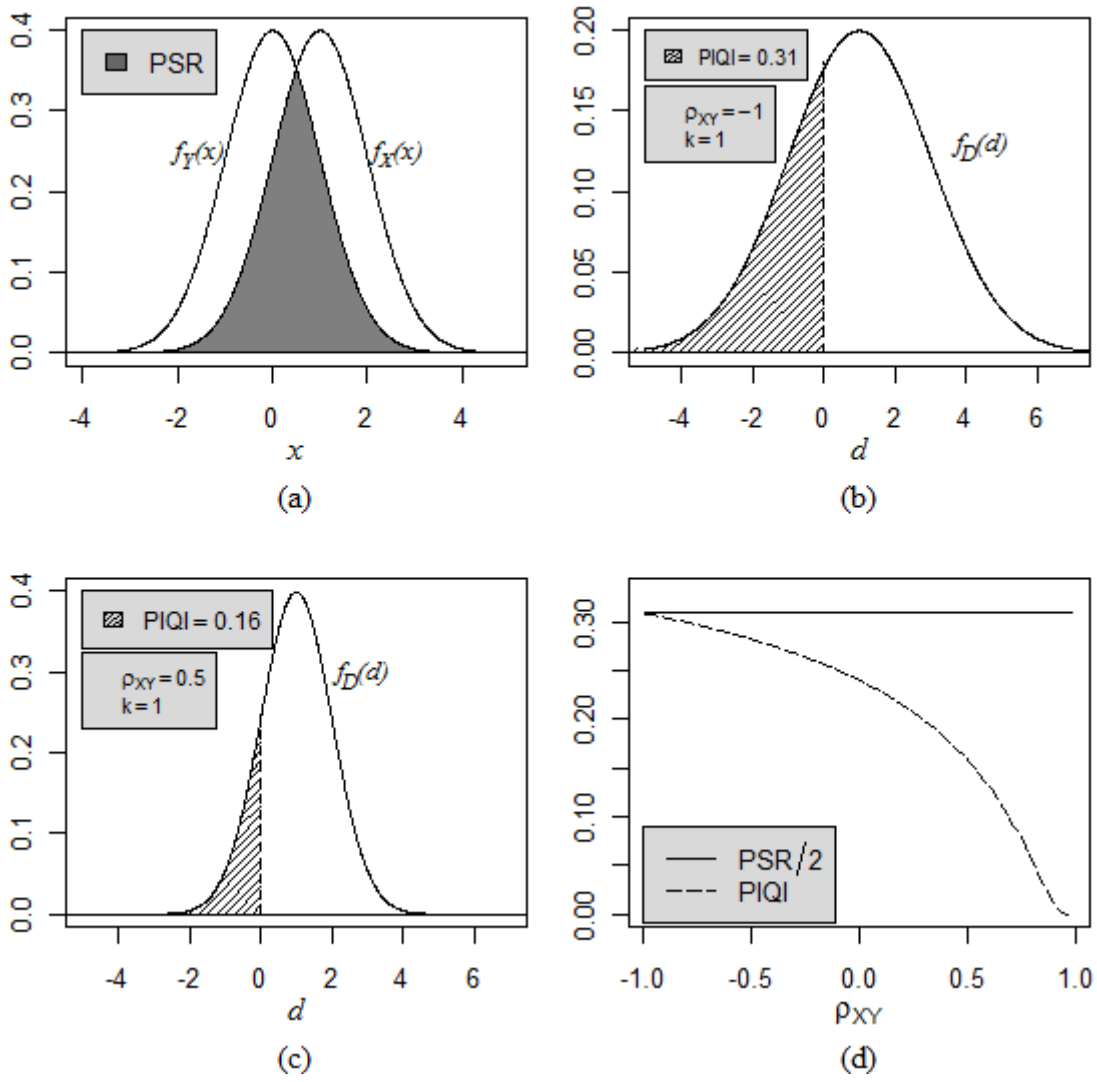


Figure 3.3: The PSR and the PIQI as areas in a density
Panel (a) gives the marginal densities of the normally distributed outcome variables X and Y where $\mu_x = 1,\ \mu_Y = 0, \sigma_X^2 = \sigma_Y^2 = 1$and $\rho_{XY} = 0$. The PSR is the shaded area. Since $D = X - Y$ is unobservable, panels (b) and (c) illustrate potential densities of D under different correlations. Panel (d) illustrates that the PIQI is a strictly decreasing function of ρ_{XY}.

3.3. PSR Reparameterization

The PSR was introduced in Section 2.5 and first illustrated in Figure 2.2 where $\sigma_Y^2 > \sigma_X^2$, which now implies that $k > 1$. As stated in Section 2.5, the PSR calculated from (2.8) has been used as a method for assessing S-T interaction. Figure 3.2 illustrates why there is controversy over the PSR's interpretation as a measure for treatment heterogeneity. Panel (d) of Figure 3.3 illustrates that the PIQI is a strictly decreasing function of ρ_{XY}, and hence, σ_D^2. This should be expected since the PIQI is one measurement of treatment heterogeneity. Figure 3.3 (d) also shows that ½ PSR, depicted with the solid line, is not a function of interindividual correlation ρ_{XY}. So even though the PIQI decreased from 0.31 to 0.16 from panels (b) to (c), ½ PSR remains a constant 0.31. However, some connections between the PSR and the PIQI will show that the PSR is still a useful quantity in the area of treatment heterogeneity.

In order to establish a connection between the PSR and the PIQI some important features of the PSR given in Section 2.5 as (2.9) are reparameterized. The reparameterization criteria was given in the previous section as $\mu_X > \mu_Y$, $\sigma_X^2 = \sigma^2$, and $\sigma_Y^2 = k^2\sigma^2$. Thus, under the normal model, when $k \neq 1$ there will be exactly two finite points of equality, x_L, x_U with $x_L < x_U$, where $f_X(x_L) = f_Y(x_L)$ and $f_X(x_U) = f_Y(x_U)$. As shown in the appendix (see A.3.1), under the reparameterization, both x_L and x_U now result from

$$\frac{(\mu_Y - k^2\mu_X) \pm \sqrt{k^2(\mu_X^2 + \mu_Y^2 - 2\mu_X\mu_Y) - k^2 2\sigma^2 \ln(k)(1-k^2)}}{(1-k^2)}, \qquad (3.4)$$

which provides the same solutions as (2.9). Again, under normality the PSR can be calculated by adding the three probabilities in (2.10), as illustrated in Figure 2.2 by summing over the three areas separated by the dotted lines.

As a working example, suppose Figure 2.2 was generated from $\mu_X = 1,\ \mu_Y = 0, \sigma^2 = 1$, and $k = 2$. Then based on (3.3), $x_L = -0.18$ and $x_U = 2.85$, and from (2.10) the PSR is equal to 0.61. Furthermore, to see how the PSR responds to changes over k, let $k = 1$ as in Figure 2.3. Then $f_X(x) = f_Y(x)$ at a single value determined by $x_E = \frac{\mu_Y + \mu_X}{2}$. The calculation of the PSR is then simplified to

$$\text{PSR} = P(X \leq x_C) + P(Y \geq x_C) = 2 \times \Phi\left(\frac{\mu_Y - \mu_X}{2\sigma}\right). \tag{3.5}$$

Leaving all other parameters unchanged $x_E = 0.5$ and the PSR $= 0.617$. So given $\mu_X = 1,\ \mu_Y = 0, \sigma^2 = 1$, and $k = 1$, the entire shaded area in Figure 2.3 would be equal to 0.617.

3.4. Connections Between the PSR and the PIQI

❖ **Proposition 3.1**

Given the distributions assigned earlier (bivariate normality of *X* and *Y* under (2.3)) and the reparameterization in Section 3.2, when $\sigma_X^2 = \sigma_Y^2$ (i. e., $k = 1$)

$$\text{PIQI}^{max} = \frac{\text{PSR}}{2}. \tag{3.6}$$

Proof:

From (3.3)

$$\begin{aligned} \text{PIQI}^{max} &= \Phi\left(\frac{\mu_Y - \mu_X}{\sigma\sqrt{1 + k^2 + 2k}}\right) \\ &= \Phi\left(\frac{\mu_Y - \mu_X}{2\sigma}\right) \\ &= \frac{\text{PSR}}{2}. \ \blacksquare \end{aligned}$$

Proposition 3.1 follows intuitively from (2.11) since the calculation of the PIQI^{max} is calculated under the assumption that $\mu_X \geq \mu_Y$, and the PSR makes no distinction on which mean is greater. Furthermore, proposition 3.1 defines not only the constraints under which the PSR can be used to assign probabilities, but also gives probabilistic meaning to the PSR with respect

to individual effects; an issue heretofore not resolved. Thus, using Figure 3.3, which displays the PSR $= 0.617$ given $\mu_X = 1,\ \mu_Y = 0,\ \sigma^2 = 1,$ and $k = 1$ in panel (a), additionally shows the PIQI^{max} = ½PSR $= \frac{0.617}{2} = 0.308$ displayed in panel (b).

The following results are needed for proposition 2, which will establish a more general relationship between the PSR and the PIQI. The results are given for $k > 1$ under (2.3), but similar results can be obtained for $k < 1$.

Result 3.1: Given $\rho_{XY} = \pm 1$, the sample space for the (x, y) pairs is restricted to the line $y = \beta_0 + \beta_{XY}x$, with probability one, where $\beta_0 = \mu_Y - \beta_{XY}\mu_X$ and $\beta_{XY} = \pm k$.

Proof:

$$\beta_1 = \frac{\rho_{XY}\sigma_X\sigma_Y}{\sigma_X^2} = \frac{\rho_{XY}k\sigma^2}{\sigma^2} = \pm k$$

and

$$\beta_0 = \mu_Y - \beta_1\mu_X$$

by definition. ■

Result 3.2: Given $\rho_{XY} = -1$ and setting x equal to y and denoting their common value by x_{-1} then,

$$x_{-1} = x_{\rho_{XY}=-1} = \frac{\mu_Y + k\mu_X}{1 + k}$$

and

$$x_L < x_{-1} < \mu_X.$$

Proof of x_{-1} follows directly from result 1 by substituting x_{-1} for both x and y and solving for x_{-1}. Proof of $x_L < x_{-1} < \mu_X$ is given in the appendix (see A.3.2).

Result 3.3: PIQI$^{max} = P(D < 0)_{\rho_{XY}=-1} = P(X \leq x_{-1}) = P(Y \geq x_{-1})$.

Proof:

$$P(D<0)_{\rho_{XY}=-1} = \Phi\left(\frac{0-(\mu_X-\mu_Y)}{\sigma\sqrt{1+k^2+2k}}\right) = \Phi\left(\frac{\mu_Y-\mu_X}{\sigma\sqrt{(1+k)^2}}\right)$$

$$= \Phi\left(\frac{\mu_Y-\mu_X}{\sigma(1+k)}\right)$$

$$P(X \le x_{-1}) = \Phi\left(\frac{x_{-1}-\mu_X}{\sigma_X}\right) = \Phi\left(\frac{\frac{\mu_Y+k\mu_X}{(1+k)}-\mu_X}{\sigma}\right) = \Phi\left(\frac{\mu_Y+k\mu_X-(1+k)\mu_X}{\sigma(1+k)}\right)$$

$$= \Phi\left(\frac{\mu_Y-\mu_X}{\sigma(1+k)}\right)$$

$$P(Y \ge x_{-1}) = 1-\Phi\left(\frac{x_{-1}-\mu_Y}{\sigma_Y}\right) = 1-\Phi\left(\frac{\frac{\mu_Y+k\mu_X}{(1+k)}-\mu_Y}{\sigma}\right) = 1-\Phi\left(\frac{\mu_Y+\mu_X-(1+k)\mu_Y}{\sigma(1+k)}\right)$$

$$= 1-\Phi\left(\frac{\mu_X-\mu_Y}{\sigma(1+k)}\right) = \Phi\left(\frac{-(\mu_X-\mu_Y)}{\sigma(1+k)}\right)$$

$$= \Phi\left(\frac{\mu_Y-\mu_X}{\sigma(1+k)}\right). \quad \blacksquare$$

A graphical illustration of results 3.1 through 3.3 is given in Figure 3.4. In particular note the representation of x_{-1} from result 3.2, and the equality from result 3 in which the $P(X \le x_{-1})$ given as the lighter region to the left of x_{-1} is equal to the $P(Y \ge x_{-1})$ displayed as the darker shaded region on the right of x_{-1}.

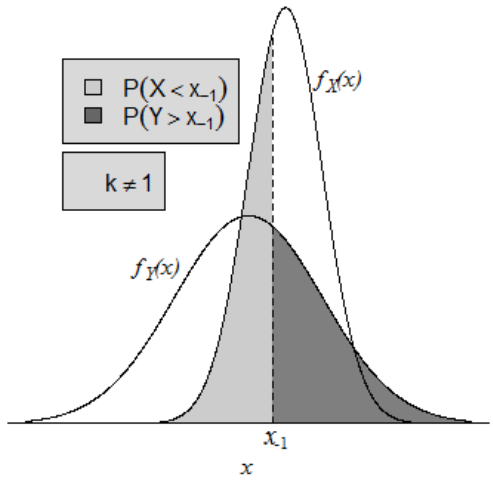


Figure 3.4: Application of result 3.3
The *pdf's* of the normally distributed outcome variables *X* and *Y* are displayed, where $\mu_X > \mu_Y$ and $k > 1$ so that $\sigma_X^2 < \sigma_Y^2$. The point x_{-1} is calculated from result 2, under the constraint that $\rho_{XY} = -1$, which results in the equality of the two shaded regions as given in result 3 in which $P(X \le x_{-1}) = P(Y \ge x_{-1})$.

Result 3.4: Given (2.3) and $k > 1$,

$$P(x_L \le X < x_{-1}) > P(x_L \le Y < x_{-1}) \text{ and } P(Y \ge x_U) > P(X \ge x_U)$$

Proof:

Since $k > 1$, $f_X(x) > f_Y(x)$ for $x_L < x < \mu_X$, from result 3.3 $x_L < x < x_{-1} < \mu_X$. Therefore, $f_X(x) > f_Y(x)$ for $x_L < x < x_{-1}$, so $P(x_L \le X < x_{-1}) \ge P(x_L \le Y < x_{-1})$.

Since $k > 1$, $f_X(x) < f_Y(x)$ for $x_U < x < \infty$. Therefore, $P(Y \ge x_U) \ge P(X \ge x_U)$. ■

❖ **Proposition 3.2**

Given (2.3), proposition 1, results 3.1 through 3.4, and $k \neq 1$

$$\text{PIQI}^{max} > \frac{\text{PSR}}{2}. \tag{3.7}$$

Proof: Let $k > 1$, then

$$\begin{aligned}
2 \times \text{PIQI}^{max} &= P(X \le x_{-1}) + P(Y \ge x_{-1}) \\
&= P(X \le x_L) + P(x_L \le X < x_{-1}) + P(x_L \le Y < x_U) - P(x_L \le Y < x_{-1}) \\
&\quad + P(Y \ge x_U) + P(X \ge x_U) - P(X \ge x_U) \\
&= P(X \le x_L) + P(x_L \le Y < x_U) + P(X \ge x_U) + P(x_L \le X < x_{-1}) \\
&\quad - P(x_L \le Y < x_{-1}) + P(Y \ge x_U) - P(X \ge x_U) \\
&= \text{PSR} + P(x_L \le X < x_{-1}) - P(x_L \le Y < x_{-1}) + P(Y \ge x_U) - P(X \ge x_U)
\end{aligned}$$

Then, by result 3.4 and since $k > 1$

$$\text{PIQI}^{max} > \frac{\text{PSR}}{2}. \ ■$$

Proof for $k < 1$ is similar. Thus, from propositions 3.1 and 3.2, $\text{PIQI}^{max} \ge \frac{\text{PSR}}{2}$ when $k \in (0, \infty)$.

Figure 3.5 illustrates result 3.4. Note that $k > 1$ so that $\sigma_Y > \sigma_X$ as in Figure 3.4. The lighter shade is equal to the PSR, while the darker region is the area in excess of $2 \times \text{PIQI}^{max}$ after subtracting the PSR. Figure 3.6 illustrates propositions 3.1 and 3.2 from an alternative perspective. Limiting k over the range $0 \le k \le 3$, the dashed line in Figure 3.6 represents the PIQI^{max} and the solid line represents ½ PSR. From this perspective it is seen that the PIQI^{max} is an upper bound for ½ PSR.

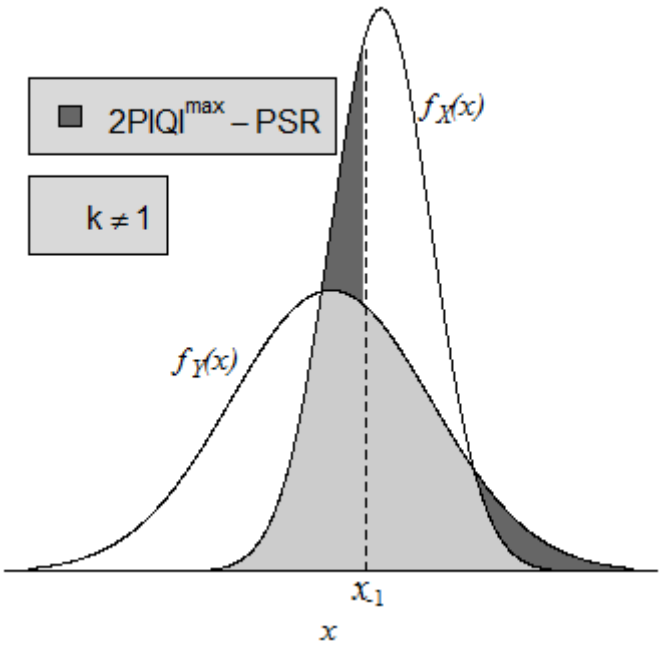


Figure 3.5: Illustration of proposition 3.2
The lighter shade is equal to the PSR and the darker shade is equal $2 \times \text{PIQI}^{max} - \text{PSR}$. As k gets close to 1, the darker shade will go to zero. Otherwise, the darker shade will be greater than zero.

A return to the working example further illustrates the concepts of propositions 3.1 and 3.2 and Figures 3.4 through 3.6, where $\mu_X = 1,\ \mu_Y = 0, \sigma^2 = 1$, and $k\ = 2$. Recall that $x_{-1} = 0.67$, PSR $= 0.61$, and $\text{PIQI}^{max} = 0.37$. As a consequence of result 3.3 is that $\text{PIQI}^{max} = P(X < x_{-1}) = P(Y > x_{-1}) = 0.37$. Note then that the darker area in Figure 3.4 is equal to $2 \times \text{PIQI}^{max} - \text{PSR} \cong 0.74 - 0.61 = 0.13$. Furthermore, it can be seen that the difference between the $\text{PIQI}^{max} = 0.37$ and ½ PSR $= 0.305$ at $k = 2$ in Figure 3.5 is equal to $\frac{PSR}{2} = \frac{0.13}{2} = 0.065$. Note this difference gets smaller as k gets closer to 1 and disappears entirely as illustrated with the dotted line over $k = 1$ in Figure 3.6.

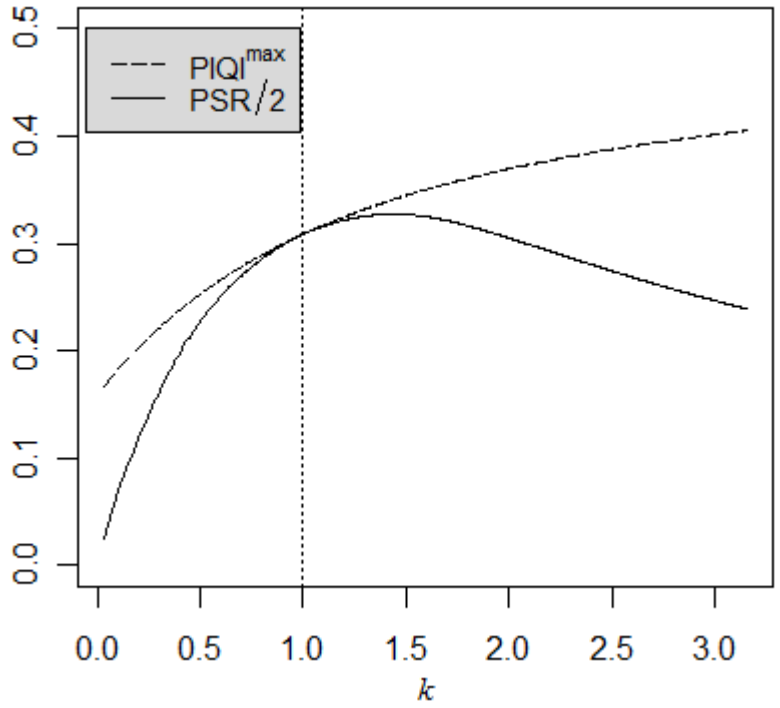


Figure 3.6: The PIQI and the PSR as functions of *k*
This figure illustrates propositions 3.1 and 3.2 in which the PIQI^{max} shown as the dashed line is an upper bound for ½PSR shown as the solid line over k from 0 to 3. The only point of equality is at $k = 1$ (see proposition 3.1) as shown with the vertical dotted line. In this illustration $\mu_X = 1, \mu_Y = 0$, and $\sigma^2 = 1$.

3.5. Additional PSR/IQI Connections

3.5.1. Expansion of the PSR as an Upper Bound

Made clear from Figure 3.6, the PIQI^{max} is an upper bound for ½ PSR. This occurs when $k = 1$. But what is the relationship between the PIQI and the PSR more generally? For example, note from Figure 3.3 panel (a) that the ½ PSR is an upper bound for the PIQI. Furthermore, since the PSR is observable, while the PIQI requires a value for ρ_{XY}, additional connections between the PSR and the PIQI other than the PIQI^{max} may prove useful. Presented here are additional circumstances under which ½ PSR may be employed as a measurement for the PIQI, other than the case where $\sigma_X^2 = \sigma_Y^2$ and $\rho_{XY} = -1$.

As seen in Figure 3.6, the PIQI and the PSR are both functions of k. Also seen in Figure 3.3 panel (d) the PIQI is a function of ρ_{XY}, while the PSR is not (refer to (3.1) and (2.10)). So for a fixed k there exists a ρ_{XY} for which $\text{PIQI}_{k,\rho_{XY}} = \text{PSR}_k$. Since ρ_{XY} would be constrained by k, this value would be denoted by $\rho_{XY(k)}$. Furthermore, since $\text{PIQI}_{k,\rho_{XY}}$ is a strictly decreasing function of ρ_{XY}, if it were determined, say *a priori*, that $-1 < \rho_{XY(k)} \leq \rho_{XY}$, then

$$\text{PIQI}^{UB}_{\rho_{XY(k)}} = ½\ \text{PSR}_k \tag{3.8}$$

would be an upper bound to the PIQI. In which case, ½ PSR may be used to calculate and display the upper bound for the PIQI even though $k \neq 1$. The calculation of $\rho_{XY(k)}$ is the focus of proposition 3.3 given next.

❖ **Proposition 3.3**

Given (2.3), $\rho_{XY} = \rho_{XY(k)} > -1$, and $k \geq 1$, if (3.8) holds, then

$$\rho_{XY(k)} = \frac{\left[\dfrac{\mu_Y - \mu_X}{\sigma \times \Phi^{-1}\left(\dfrac{\text{PSR}}{2}\right)}\right]^2 - 1 - k^2}{-2k} \tag{3.9}$$

Proof:

$$\text{PIQI}^{UB}_{\rho_{XY(k)}} = \frac{\text{PSR}}{2} = \Phi\left(\frac{\mu_Y - \mu_X}{\sigma\sqrt{1 + k^2 - 2k\rho_{XY(k)}}}\right)$$

Thus,

$$\Phi^{-1}\left(\frac{\text{PSR}}{2}\right) = \frac{\mu_Y - \mu_X}{\sigma\sqrt{1 + k^2 - 2k\rho_{XY(k)}}} \tag{3.10}$$

Then, solving for $\rho_{XY(k)}$ in (3.10) yields the result. ■

Of course, when $k = 1,\ \rho_{XY} = -1$ and $\text{PIQI}^{UB}_{\rho_{XY(k=1)}} = \text{PIQI}^{max} = ½\ \text{PSR}_{k=1}$.

Caution should be used when specifying any value of $\rho_{XY} > -1$, since the potential outcomes framework makes it clear that no observable data will be available for supporting

$\rho_{XY(k)} > -1$. However, given *a priori* knowledge of ρ_{XY}, certain values of $\rho_{XY(k)}$ may be reasonable. For example, since ρ_{XY} is the correlation between *X* and *Y* on the same individual, $\rho_{XY} \geq 0$ may be plausible in some situations. In our working example, where $\mu_X = 1$, $\mu_Y = 0$, $\sigma^2 = 1$, and $k = 2$, $\rho_{XY(k)} = 0.289$. Thus, given *a priori* that $\rho_{XY(k)} \geq 0.289$, $\text{PIQI}_{0.289}^{UB} =$ ½ PSR $= \frac{0.61}{2} = 0.305$ may be used to specify and display the upper bound on the PIQI as opposed to $\text{PIQI}^{max} = 0.37$. Due to the seemingly large value of $\rho_{XY(k)} = 0.289$ in this example is that $\rho_{XY(k)}$ is a function of $\mu_D = \mu_X - \mu_Y$. Here $\mu_D = 1$ since $\mu_X = 1$ and $\mu_Y = 0$, but if $\mu_X = 2$ and $\mu_Y = 0$, then $\rho_{XY(k)} = -0.53$. Working with different values for ρ_{XY} has the same appeal here as it does with the imputation model introduced in Chapter 1 with the soy treatment example. Generating imputations for given $\rho_{XY} = 0.289$ would have the same result as suggesting that ½ the PSR is the upper bound for the PIQI. Thus an estimate for the proportion of IQIs in the population would be approximately equal to ½ PSR.

3.5.2. Utilization of the PIQI Lower Bound

An argument can be made that the lower bound for the PIQI, which has been denoted and defined as PIQI^{min} in (3.2) is at least as important as the PIQI^{max} since the PIQI^{max} offers only the "potential" for the existence of IQI within the population. Whereas, if the $\text{PIQI}^{min} > 0$, then IQI exists. Although this work does not draw any connections between the PIQI^{min} and the PSR, an important result about the PIQI^{min} exists shown by way of propositions 3.4 and 3.5 given here. First, note that similar to results 3.1 and 3.2 that fixing $\rho_{XY} = 1$ setting x equal to y denoting their common value by x_1, then

$$x_1 = x_{\rho_{XY}=1} = \frac{\mu_Y - k\mu_X}{1 - k}$$

❖ **Proposition 3.4**

Given (2.3) and $k \in (0, \infty)$

$$\text{PIQI}^{min} = P(D < 0)_{\rho_{XY}=1} = P(X > x_1) = P(Y > x_1)$$

Proof: For $k > 1$

$$P(D < 0)_{\rho_{XY}=1} = \Phi\left(\frac{0 - (\mu_X - \mu_Y)}{\sigma\sqrt{1 + k^2 - 2k}}\right) = \Phi\left(\frac{\mu_Y - \mu_X}{\sigma\sqrt{(1-k)^2}}\right)$$

$$= \Phi\left(-\frac{\mu_Y - \mu_X}{\sigma(1-k)}\right)$$

$$P(X > x_1) = 1 - \Phi\left(\frac{x_1 - \mu_X}{\sigma_X}\right) = 1 - \Phi\left(\frac{\frac{\mu_Y - k\mu_X}{(1-k)} - \mu_x}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\mu_Y - k\mu_X - (1-k)\mu_X}{\sigma(1-k)}\right) = 1 - \Phi\left(\frac{\mu_Y - \mu_X}{\sigma(1-k)}\right)$$

$$= \Phi\left(-\frac{\mu_Y - \mu_X}{\sigma(1-k)}\right)$$

$$P(Y > x_1) = 1 - \Phi\left(\frac{x_1 - \mu_Y}{\sigma_Y}\right) = 1 - \Phi\left(\frac{\frac{\mu_Y - k\mu_X}{(1-k)} - \mu_Y}{k\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\mu_Y - k\mu_X - (1-k)\mu_Y}{k\sigma(1-k)}\right)$$

$$= 1 - \Phi\left(\frac{k(\mu_Y - \mu_X)}{k\sigma(1-k)}\right) = 1 - \Phi\left(\frac{\mu_Y - \mu_X}{\sigma(1-k)}\right)$$

$$= \Phi\left(-\frac{\mu_Y - \mu_X}{\sigma(1-k)}\right).$$

Note that since $k > 1$, $\sigma(1-k) < 1$, which gives the results. ■

A similar result holds for $k < 1$.

❖ **Proposition 3.5**

Given $(\mathbf{2.3})$ and $\boldsymbol{k \neq 1}$,

$$\text{PIQI}^{min} > 0$$

Proof:

$\text{PIQI}^{min} = 0$ if and only if $\sigma_D^2 = 0$. When $k \neq 1$, $\sigma_D^2 > 0$. ■

The importance of these results are that when $Var(X) \neq Var(Y)$, $\text{PIQI}^{min} > 0$ and can be calculated using either (3.2) or proposition 3.4. Figure 3.7 illustrates result 3.5 and propositions 3.4 and 3.5 by way of our working example in which $\mu_X = 1,\ \mu_Y = 0, \sigma^2 = 1$, and $k = 2$. The dashed vertical line over 2 in panel (a) indicates $x_1 = 2$. Consequently, $\text{PIQI}^{min} = P(D < 0)_{\rho_{XY}=1} = P(X > 2) = P(Y > 2) \cong 0.16$, which can be seen from the shaded areas of the densities for *X*, *Y*, and *D* shown in panels (b), (c), and (d), respectively of Figure 3.7. Note that despite the different shapes, the areas are equal. The interpretation is that a minimum of 16 percent of the individuals in the population will do better on the control than on the treatment even though $\mu_X = 1 > \mu_Y = 0$. Alternatively expressed, the proportion of individuals that will exhibit an IQI in the population is 0.16. Combining this information with the calculation of the $\text{PIQI}^{max} = 0.37$ from Section 3.4, the bounds for the PIQI are (0.16, 0.37), which can be calculated from(3.2) and (3.3) .

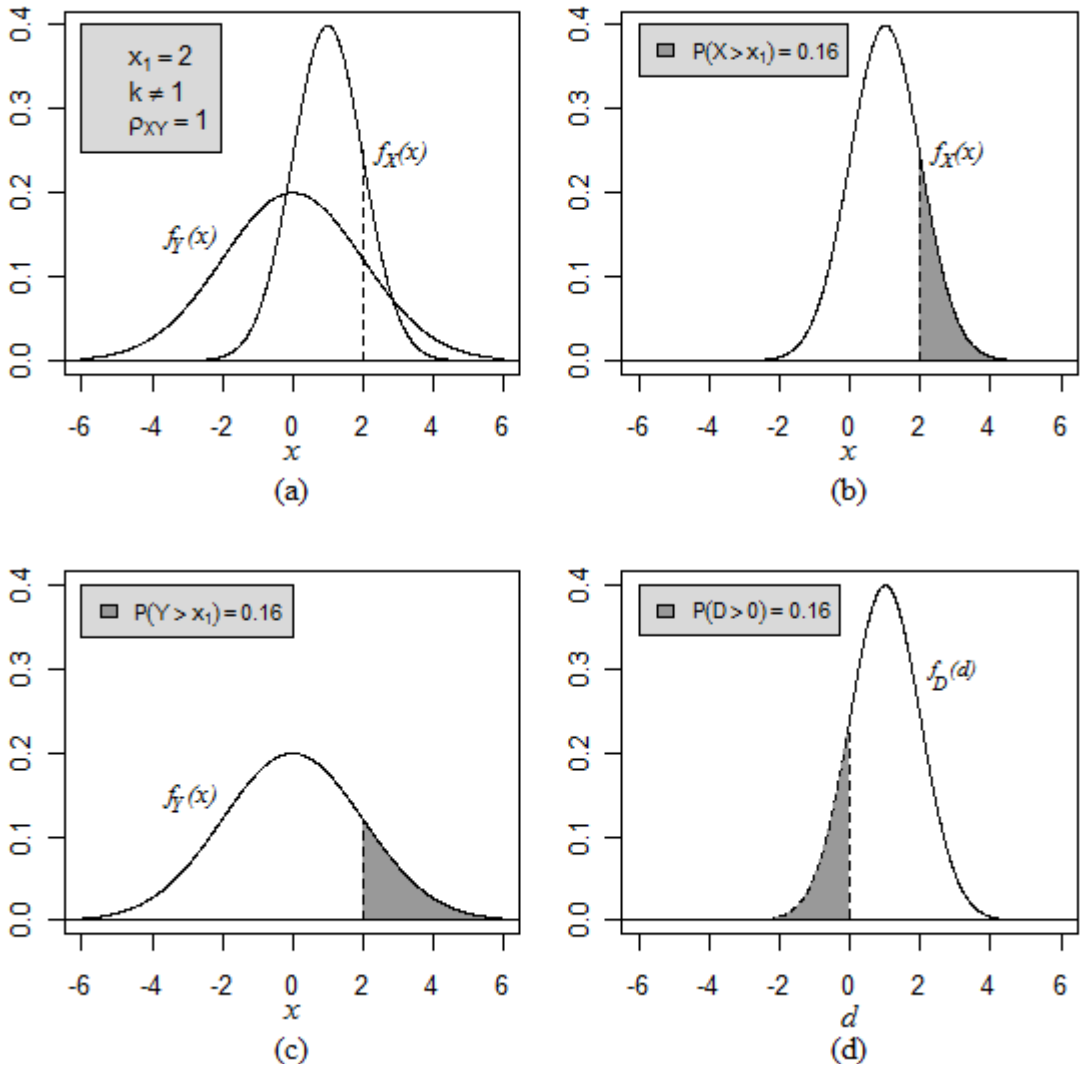


Figure 3.7: Illustration of minimum PIQI
This figure illustrates an example of result 3.5 and propositions 3.4 and 3.5 in the particular case of where $\mu_X = 1$, $\mu_Y = 0$, and $k = 2$. The densities of the normal outcome variables *X* and *Y* are displayed in panel (a). The vertical dashed line is over the value $x_1 = 2$ found using result 3.5. Panels (b), (c), and (d) show the marginal densities of *X*, *Y*, and *D*, respectively. The shaded areas in each of these graphs are equal, illustrating proposition 3.4.

3.6. Connections between SI and S-T Interaction

So far it has been shown that $Var(D) = \sigma_D^2$ contributes to the PIQI, and the components of S-T interaction as either IQI or non-IQI have been identified. Furthermore, the conditions

under which the PIQI^{max} may be displayed as the PSR have been given. However, information about SI, and its components, which include QI and non-QI, has not been considered. Although it is true that the amount of S-T interaction does not change across the entire population, some S-T interaction may be explained by SI. In fact, since the study of SI presupposes a grouping variable and some degree of homogeneity of treatment response within groups, with SI then explained by differences in treatment response across groups, σ_D^2 may be reduced within subsets. As such, in this section the degree to which SI contributes to σ_D^2 in hopes of reducing the PIQI within subsets is investigated.

3.6.1. Subset and Individual Qualitative Interaction

Recall from Section 3.1 in which for $i = 1,2, \dots, g$, the number of subsets, $\delta_i = \mu_{T(s_i)} - \mu_{R(s_i)}$ represents the average difference between the treatment groups in subset i, which can be estimated from observed data and, thus, some evaluation of QI made from the data by detecting situations where certain δ_i, $\delta_{i'}, i \neq i'$, differ in sign. An IQI cannot be directly evaluated using observed data because the individual effects, D_i, $i = 1, \dots, N$, are not observable in the data. However, proposition 3.6, presented next, and the observations that follow establish some additional connections between QI and IQI.

❖ **Proposition 6**

If there is QI in the population, then there is IQI in the population.

Proof:

Assume a population with subsets s_i for $i = 1, 2, \dots, g$. If QI is present, then there is a $\delta_i > 0$ and $\delta_{i'} < 0$ for some i and i'. Because means are defined as averages with respect to a population distribution, there must be at least one j^{th} individual in group i such that $D_{ij} > 0$ and at least one j' individual in group i' such that $D_{i'j'} < 0$.

Corollary 3.1 If there is no IQI in the population, then there is no QI in the population.

The proof follows from proposition 6.

Additional Observations:

1. The subsets may be thought of as delineated by a grouping variable, say *Z*. Thus a SI is a consequence of a treatment by *Z* interaction, an interaction that can be estimated in observed data with the appropriate design. Within levels of *Z* (i.e., within subsets), there may be an S-T interaction that cannot be directly evaluated in observed data.

2. If there is no QI in a population, there still may be IQI within subsets of the population.

3. If there is a constant individual treatment effect (Holland, 1986) within each subset, then an IQI is explained by subset interaction in the form of a QI and/or non-QI. Simple diagnostics are available to evaluate the bounds on treatment heterogeneity within subsets (including the constant effect assumption), and these are described in a later sub-section.

3.6.2. The Conditional PSR

So that the relationship established above between the PSR and the PIQI may be utilized within subsets, a new quantity termed the conditional PSR is introduced here, which is defined using the conditional distributions of *X* and *Y* given the observed covariate $Z = z_0$ so that

$$\text{PSR}_{z_0} = \int min\left(f_{X|Z=z_0}(u), f_{Y|Z=z_0}(u)\right) du. \tag{3.11}$$

To simplify notation, let $Z = z_0$ be denoted simply as z_0 for the remainder of this chapter. The conditional PSR_{z_0} has all the advantages of the marginal PSR, but at a particular value of z_0, thus displaying the overlap of conditional distribution at that value of Z. As with the PSR, the relationship between PSR_{z_0} and PIQI_{z_0} (introduced in the next section) depends on whether $\sigma^2_{X|Z} = \sigma^2_{Y|Z}$, which may be reparameterized as $\sigma^2_{X|Z} = k^2_Z \sigma^2_{X|Z}$, where the conditional

k^2 is given by k_Z^2. Note that when Z is continuous, for the constraint that $k_Z^2 = 1$ requires that $(1 - \rho_{XZ}^2) = k^2(1 - \rho_{YZ}^2)$, since $\sigma_{X|Z}^2 = \sigma^2(1 - \rho_{XZ}^2)$ and $\sigma_{Y|Z}^2 = k^2\sigma^2(1 - \rho_{YZ}^2)$. Thus, specification of $k_{Z_0}^2 = 1$ requires that

$$(1 - \rho_{XZ}^2) \leq k^2 \leq \left(\frac{1}{1 - \rho_{YZ}^2}\right).$$

This constraint on k means that for the conditional variances to be equal $\left(i.e.\, \sigma_{X|Z}^2 = \sigma_{Y|Z}^2\right)$ there is a constraint on how different the marginal variances (σ_X^2 and σ_Y^2) can be, and that difference is a function of the correlation parameters ρ_{XZ} and ρ_{YZ}.

3.6.3. The Conditional PIQI and Continuous Covariates

The development and definition of the conditional PIQI depends on the type of covariate; continuous or categorical. First, a continuous covariate *Z* that augments potential outcomes (X, Y), as in Gadbury, Iyer, and Allison (2001), is considered in this section, where the distribution of *D* given z_0 is assumed to be normal with conditional mean

$$\mu_{D|z_0} = \mu_Y - \mu_X + (\beta_{XZ} - \beta_{YZ})(z_0 - \mu_Z) \tag{3.12}$$

and conditional variance,

$$\sigma_{D|Z}^2 = \sigma_{X|Z}^2 + \sigma_{Y|Z}^2 - 2\sigma_{X|Z}\sigma_{Y|Z}\rho_{XY|Z}. \tag{3.13}$$

β_{XZ} and β_{YZ} in (3.12) are the slope coefficients between *Z* and *X* and Z and *Y*, respectively, and $\rho_{XY|Z}$ in (3.13) is the partial correlation of *X* and *Y* given *Z*. The conditional variances, $\sigma_{X|Z}^2$ and $\sigma_{Y|Z}^2$, are allowed to be different across the two treatment groups but are assumed to not depend on the value of Z. The PIQI_{z_0} is the conditional probability of IQI formed within the subset (or subpopulation) defined by the covariate *Z*. Similar to equations (3.2) and (3.3)

$$\Phi\left(\frac{-\mu_{D|z_0}}{\sqrt{\sigma^2_{X|Z}+\sigma^2_{Y|Z}-2\sigma_{X|Z}\sigma_{Y|Z}}}\right) \leq \text{PIQI}_{z_0} \leq \Phi\left(\frac{-\mu_{D|z_0}}{\sqrt{\sigma^2_{X|Z}+\sigma^2_{Y|Z}+2\sigma_{X|Z}\sigma_{Y|Z}}}\right) \tag{3.14}$$

by letting the partial correlation $\rho_{XY|Z}$ be 1 and -1, respectively. Since, $\mu_{D|z_0}$ could be greater than μ_D when $\beta_{XZ} \neq \beta_{YZ}$, and both $\sigma^2_{X|Z} \leq \sigma^2_X$ and $\sigma^2_{Y|Z} \leq \sigma^2_Y$, it is possible to identify subsets of the population for which $\text{PIQI}^{max}_{z_0} < \text{PIQI}^{max}$. In fact, if σ^2_D is completely determined by SI, then σ^2_D within subsets denoted by $\sigma^2_{D|Z}$ will be 0 so that the $\text{PIQI}^{max}_{z_0}$ will be 0. This fact follows from observations 2 and 3 in Section 3.6.1.

❖ **Proposition 3.7**

Given $(\mathbf{2.3})$ the conditional distributions $\boldsymbol{f_{X|z_0}}$ and $\boldsymbol{f_{Y|z_0}}$, and a finite value of $\boldsymbol{z_0}$, if $\boldsymbol{k_Z \neq 1}$, then

$$\text{PIQI}^{max}_{z_0} \geq \frac{\text{PSR}_{z_0}}{2}. \tag{3.15}$$

The proof follows directly from the proofs of propositions 3.1 and 3.2 replacing the marginal means and variances of *X*, *Y*, and *Z* with their conditional counterparts. Proposition 3.7 shows that the relationship between the $\text{PIQI}^{max}_{z_0}$ and the PSR_{z_0} holds over subsets of the population. Gadbury et al. (2001) showed that,

$$\begin{aligned}\sigma^2_D &= \left(\sigma_{X|Z}-\sigma_{Y|Z}\right)^2 + 2\sigma_{X|Z}\sigma_{Y|Z}\left(1-\rho_{XY|Z}\right) + (\beta_{XZ}-\beta_{YZ})^2\sigma^2_Z \\ &= {\sigma_{D|Z}}^2 + (\beta_{XZ}-\beta_{YZ})^2\sigma^2_Z \end{aligned} \tag{3.16}$$

so σ^2_D is comprised of two components; those that can be attributed to SI and those that can be attributed to S-T interaction, within subsets. If $\beta_{XZ} = \beta_{YZ}$, σ^2_D is entirely a function of S-T interaction, then $\text{PIQI}^{max}_{z_0} = \text{PIQI}^{max}$. However, if $(\beta_{XZ}-\beta_{YZ})^2 > 0$, then both $\text{PIQI}^{max}_{z_0} < \text{PIQI}^{max}$ and $\text{PSR}_{z_0} < \text{PSR}$ for selected values of z_0. The quantity, $(\beta_{XZ}-\beta_{YZ})^2\sigma^2_Z$, can be

estimated using the observed data, which serves as a diagnostic tool for detecting the presence and the degree of SI within a population as suggested by observation 3 in Section 3.6.

Although the component wise delineation of σ_D^2 when Z is continuous is the product of previous work, these results provide the context for the next section, which provides a delineation of σ_D^2 when Z is categorical.

3.6.4. The Conditional PIQI and Categorical Covariates

The study of subset interaction presupposes a grouping variable and some degree of homogeneity of treatment response within groups, with QI then explained by differences in treatment response across groups. One reason for subset analysis then is to identify "which treatment is best for which kinds of patients," (Byar and Corle 1997, p. 455). Standard methods seek to find such subsets through an investigation of interaction effects (Byar and Corle 1977; Simon 1982) or a direct test for a qualitative interaction (Gail and Simon 1985; Silvapulle 2001; Li and Chan 2006). In each case the interaction is detectable by changes in the mean response across subsets. This section presents the PIQI as a way to not only help identify such subsets, but also help to assess remaining treatment heterogeneity and its possible consequences within subsets. Using potential outcomes, the reduction in σ_D^2 given the grouping variable is shown to be proportional to the sum of squares used to test for a subset-treatment interaction. Thus an effective grouping variable not only identifies sensitive subsets, but also reduces the PIQI within at least one subset.

Suppose Z is a categorical covariate with g levels. Only balanced designs are considered so that there are n units per group for a total of ng experimental units. Prior to grouping by Z, individual treatment effects are rewritten as, $D_j = X_j - Y_j$, where $X_j = \mu_X + \varepsilon_{Xj}$, $Y_j = \mu_Y + \varepsilon_{Yj}$,

and $j = 1,2,\ldots,ng$, and μ_X and μ_Y are the population averages for X and Y, respectively, as before. So,

$$D_j = \mu_X - \mu_Y + \varepsilon_{Xj} - \varepsilon_{Yj}, \tag{3.17a}$$

where

$$\begin{bmatrix}\varepsilon_{Xj}\\ \varepsilon_{Yj}\end{bmatrix} \sim iid\ N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y\\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2\end{bmatrix}\right). \tag{3.17b}$$

The variance parameters in(3.17b), are those used in section 2 for σ_D^2. Now assume that a random sample of potential outcomes (X_j, Y_j), $j = 1,2,\ldots,ng$, are generated from the above population with parameters denoted above in (3.17a, b). We refer to this sample as a finite population because all potential outcomes are not observable post treatment assignment (i.e., only X or Y is observable for an individual). Let R_{XY} be the finite population correlation of X and Y and $e_{Xj} - e_{Yj}$ be the residuals from a model fitted to potential outcomes data of the form in $(3.17a)$. Then the finite population variance of D may be defined by

$$S_D^2 = \frac{SSE}{ng-1} = \frac{\sum_{j=1}^{ng}(e_{Xj}-e_{Yj})^2}{ng-1}$$

$$= \frac{\sum_{j=1}^{ng}(X_j-\bar{X}.)^2}{ng-1} + \frac{\sum_{j=1}^{ng}(Y_j-\bar{Y}.)^2}{ng-1} - 2R_{XY}\frac{\sqrt{\sum_{j=1}^{ng}(X_j-\bar{X}.)^2\sum_{j=1}^{ng}(Y_j-\bar{Y}.)^2}}{ng-1},$$

where $\bar{X}.$ and $\bar{Y}.$ are given by $\frac{1}{ng}\sum_{i=1}^{ng} X_j$ and $\frac{1}{ng}\sum_{i=1}^{ng} Y_j$, respectively. If the potential outcomes were observable, S_D^2 would be an unbiased and consistent estimator for σ_D^2 discussed in section 2. For a potential outcomes model that includes Z, let $X_{ij} = \mu_X + \tau_{Xi} + \varepsilon_{Xij}$ and $Y_{ij} = \mu_Y + \tau_{Yi} + \varepsilon_{Yij}$, where $i = 1,2,\ldots,g,\ j = 1,2,\ldots,n$, and τ_{X_i} and τ_{Y_i} represent the effect of subset i on outcomes X and Y so that $D_{ij} = X_{ij} - Y_{ij}$ now gives

$$D_{ij} = \mu_X - \mu_Y + \tau_{Xi} - \tau_{Yi} + \varepsilon_{Xij} - \varepsilon_{Yij}, \tag{3.18a}$$

where

$$\begin{bmatrix}\varepsilon_{Xij}\\ \varepsilon_{Yij}\end{bmatrix} \sim iid\ N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{X|Z} & \rho_{XY|Z}\,\sigma_{X|Z}\,\sigma_{Y|Z}\\ \rho_{XY|Z}\,\sigma_{X|Z}\,\sigma_{Y|Z} & \sigma^2_{Y|Z}\end{bmatrix}\right), \tag{3.18b}$$

and $\rho_{XY|Z}$ is the partial correlation of X and Y given Z and $\sigma^2_{X|Z}$ and $\sigma^2_{Y|Z}$ are conditional variances, assumed for now to be the same across levels of Z but allowed to be different across treatment groups within levels of Z. The residual sum of squares for this "full" model that includes Z, fit to the D_{ij}, are given by

$$\text{SSE}_Z = \sum_{i=1}^{g}\sum_{j=1}^{n}\left(e_{Xij} - e_{Yij}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n} e^2_{Xij} + \sum_{i=1}^{g}\sum_{j=1}^{n} e^2_{Yij} - 2\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}\ e_{Yij}$$

where $e_{Xij} = X_{ij} - \bar{X}_{i.}$ and $e_{Yij} = Y_{ij} - \bar{Y}_{i.}$ and where $\bar{X}_{i.}$ is the mean of potential outcome values of X in the i^{th} group and similarly for $\bar{Y}_{i.}$. The expression can then be used to generate

$$\begin{aligned} S^2_{D|Z} &= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n} e^2_{Xij}}{g(n-1)} + \frac{\sum_{i=1}^{g}\sum_{j=1}^{n} e^2_{Xij}}{g(n-1)} - \frac{2\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}e_{Yij}}{g(n-1)} \\ &= S^2_{X|Z} + S^2_{Y|Z} - 2R_{XY|Z}S_{X|Z}S_{Y|Z}, \end{aligned} \tag{3.19}$$

the conditional finite population variance (see appendix A.3.3) . A comparison of the SSE_Z and the SSE for the model given in $(3.17a, b)$ provides for the following proposition as shown in A.3.4 of the appendix:

❖ **Proposition 3.8**

Given $(\mathbf{3.18}\ \boldsymbol{a, b})$

$$\frac{ng-1}{g(n-1)}S^2_D = S^2_{D|Z} + \frac{n\sum_{i=1}^{g}\left(\bar{D}_{z_i} - \bar{D}\right)^2}{g(n-1)}, \tag{3.20}$$

where $\bar{D}_{z_i}$ is the finite population average of the n true individual treatment effects within the i^{th} group of Z , given by $\bar{D}_{z_i} = \frac{1}{n}\sum_{j=1}^{n} D_{ij}$, and, $\bar{D} = \frac{1}{ng}\sum_{i=1}^{g}\sum_{j=1}^{n} D_{ij}$. (3.20) shows that the components of S^2_D (after scaling for a degrees of freedom adjustment) include both S-T

interaction within subsets specified by $S^2_{D|Z}$ and SI as a function of $\sum_{i=1}^{n}\left(\overline{D}_{z_i} - \overline{D}\right)^2$. Thus when $\sum_{i=1}^{n}\left(\overline{D}_{z_i} - \overline{D}\right)^2 > 0$, both the mean treatment effect and the PIQI within subsets varies across the subsets. In the extreme case that $S^2_{D|Z} = 0$, the finite population S-T interaction is completely explained by the interaction across subsets, which indicates a constant individual effect of treatment T relative to treatment R within subsets. In the other extreme that $S^2_{D|Z} = S^2_D$, then Z is not useful for predicting subsets of individuals who may respond successfully to one treatment over the other (i.e., Z does not explain any S-T interaction that may be present in a population). A large SI implies, as expected, that the remaining amount of S-T interaction within groups (i.e., remaining unexplainable variance) is reduced by the grouping variable.

Although D_{ij} cannot be observed, a post treatment assignment estimate for the second term in (3.20), $\frac{n\sum_{i=1}^{g}\left(\overline{D}_{z_i}-\overline{D}\right)^2}{g(n-1)}$, is available using observable data. For convenience, assume the first $\frac{n}{2}$ units within each subset were assigned to treatment T and the second $\frac{n}{2}$ units in each subset were assigned treatment R. Observable statistics are $\bar{x}_{i.} = \frac{2}{n}\sum_{j=1}^{n/2} x_{ij}$, $\bar{y}_{i.} = \frac{2}{n}\sum_{j=\frac{n}{2}+1}^{n} y_{ij}$, $\bar{x}_{..} = \frac{2}{ng}\sum_{i=1}^{g}\sum_{j=1}^{n/2} x_{ij}$, and $\bar{y}_{..} = \frac{2}{ng}\sum_{i=1}^{g}\sum_{j=\frac{n}{2}+1}^{n} y_{ij}$. An estimate for $\frac{n\sum_{i=1}^{g}\left(\overline{D}_{z_i}-\overline{D}\right)^2}{g(n-1)}$ is

$$\frac{\frac{n}{2}\sum_{i=1}^{g}[(\bar{x}_{i.} - \bar{y}_{i.}) - (\bar{x}_{..} - \bar{y}_{..})]^2}{g\left(\frac{n}{2} - 1\right)}, \tag{3.21}$$

which is a scalar of the usual sum of squares for the SI term in a $2 \times g$ factorial analysis of variance computation with $\frac{n}{2}$ observations for each treatment group combination (see appendix A.3.5). Consequently, an F-test for the contribution of the interaction term as given in a standard

ANOVA table for such a model may not only be used to diagnose the degree of SI, but also provides evidence that $S_{D|Z}^2$ may be less than S_D^2, and hence, $\sigma_{D|Z}^2 < \sigma_D^2$.

$S_{D|Z}^2$, the first term in (3.20), may be used to evaluate the PIQI within groups. Given Z is an effective grouping variable, at least one group will have a smaller PIQI than the PIQI across all levels of Z. However, the counterfactual aspect of potential outcomes still prohibits direct evaluation of this term, but it is bounded by two quantities that can be evaluated using observed data. Bounds for the individual effect variance within subsets are given by

$$\left(S_{X|Z} - S_{Y|Z}\right)^2 \leq S_{D|Z}^2 \leq \left(S_{X|Z} + S_{Y|Z}\right)^2$$

as a result of taking $R_{XY|Z} = \pm 1$, but $S_{X|Z}$ and $S_{Y|Z}$ cannot be calculated from observed data. However, estimates

$$s_{x|z} = \sqrt{\frac{\sum_{i=1}^{g}\sum_{j=1}^{n/2}\left(x_{ij} - \bar{x}_{i.}\right)^2}{g\left(\frac{n}{2} - 1\right)}} \tag{3.22a}$$

and

$$s_{y|z} = \sqrt{\frac{\sum_{i=1}^{g}\sum_{j=\frac{n}{2}+1}^{n}\left(y_{ij} - \bar{y}_{i.}\right)^2}{g\left(\frac{n}{2} - 1\right)}} \tag{3.22b}$$

can be calculated from observed data (again assuming that the first $\frac{n}{2}$ observed units were assigned treatment T and the last $\frac{n}{2}$ observed units were assigned treatment R).

Estimating bounds for the PIQI requires distributional assumptions on the individual treatment effects. If $D_{ij} \sim N\left(\mu_{D_{z_i}}, \sigma_{D|Z}^2\right), j = 1, \ldots, n$, then bounds for the PIQI at a given $Z = z_i$ are given as

$$\text{PIQI}_{Z=z_i}^{min} = \Phi\left(\frac{-\mu_{D_{z_i}}}{|\sigma_{X|Z} - \sigma_{Y|Z}|}\right) \leq \text{PIQI}_{z_i} \leq \Phi\left(\frac{-\mu_{D_{z_i}}}{\sigma_{X|Z} + \sigma_{Y|Z}}\right) = \text{PIQI}_{Z=z_i}^{max}. \tag{3.23}$$

Estimates for the parameters in these bounds, $\mu_{D_{z_i}}$, $\sigma^2_{X|Z}$, and $\sigma^2_{Y|Z}$, are given by $\bar{x}_{i.} - \bar{y}_{i.}$, $s^2_{x|z}$, and $s^2_{y|z}$, respectively. We also note that one can estimate $\sigma_{X|Z}$ and $\sigma_{Y|Z}$ separately within each group with $s_{x|z_i}$ and $s_{y|z_i}$, respectively, as opposed to "pooling" across groups as given by $(3.22a, b)$. This approach is equivalent to conducting a separate analysis within each subset. As a result, it is possible that the bounds for PIQI_{z_i} vary widely across subsets, with some subsets exhibiting the plausibility of more S-T interaction than others. A significant degree of SI is analogous to a significant F-test in a one-way ANOVA. Once the significance is discovered, a search among pair-wise comparisons will reveal which means are statistically different. Although in a search for IQIs, a large SI does not necessarily mean a high degree of IQI in the population will be explained by SI, but only the potential for this to be the case. A check for IQI across subsets will reveal the degree of IQI explained by SI. Furthermore, if the SI interaction is not significant, then it would not be expected that the amount of IQI within subsets would be significantly reduced.

3.7. Soy Treatment Trial Follow-up

The soy treatment study published by Allison et al. (2003) was used in Chapter 1 as an introduction to the idea of analyzing IQIs within a population as an addition to standard tests. Although the full data were not given, it was reported that significant reductions in both average cholesterol levels and weight were detected for the soy treatment groups over the control groups. Recall that the estimated proportion of IQI based on the 16 individuals used in the example was 0.31, which is now referred to as the PIQI^{max}. Furthermore, if k is equal to 1, then ½PSR is equal to 0.31 as well. Otherwise, $\text{PIQI}^{max} >$ ½PSR . Otherwise, there exists a $\rho_{XY(k)}$ for which ½PSR $= \text{PIQI}^{\text{UB}}_{\rho_{XY(k)}}$, where $\rho_{XY(k)}$ may be determined from the data. An important part of the

work presented in this dissertation is to bring to an investigators attention the potential for IQI within a population represented by the PIQI^{max}, as expressed in the reduced soy treatment example given in Section 1.1. Thus a follow up on the IQI analysis will be presented here based on all 73 individuals and both the PIQI^{max} and PIQI^{min}. The second part of an IQI analysis will also be presented, which involves the determination of how much of the IQI may be explained by a QI with an identifiable subset based on both either a continuous or a categorical covariate. That is the subset portion of an IQI analysis will be carried out.

3.7.1. Soy-Treatment Example

The variables considered here are cholesterol and weight, taken at baseline and again at weeks 4, 8, and 12. Participant gender was also recorded. Analysis will be conducted on the 37 participants in the soy-treatment group and the 36 participants of the control group who completed 12 weeks of the study. The two outcome variables considered are change in the 12 week cholesterol level and change in the 12 week weight measurement. The PIQI and the PSR are illustrated both without and with the use of a covariate. Normal distributions are assumed for both the soy-treatment outcome variables, denoted by *X*, and the control treatment outcome variable, denoted by *Y*.

3.7.2. IQI Analysis

Results are given for the change in the cholesterol level from baseline to week 12, so positive values are reductions in cholesterol. The analysis results for change in weight were similar and thus, are not provided here. The sample statistics are given in Table 3.2. In particular note $\bar{x} = 26.05, \bar{y} = 6.72, s_x = 30.36$, and $s_y = 16.44$ result in a significant t-test for the null hypothesis $\mu_X = \mu_Y$. So, on average, the soy treatment is significantly better than the control

treatment in reducing mean cholesterol levels. However, a test for equal means does not directly address the question of possible IQIs in the population. Thus also given in Table 3.2 is the estimated The estimated $\text{PIQI}^{max} = 0.34$. A test for $\sigma_X^2 = \sigma_Y^2$ was rejected, so using $\hat{k}^2 = 0.29$ $\widehat{\text{PIQI}}^{max} = 0.34 > ½\,\widehat{\text{PSR}} = 0.30$ also shown in Table 3.2. Thus, despite the superior performance of the soy-treatment on average, it is 'possible' that over 30% of the individuals could have improved cholesterol levels on the control using only the information in the data.

Table 3.2: Marginal sample results for soy-treatment trial

Sample Results Based on Marginal Distributions					
$\bar{x}$	$\bar{y}$	s_x	s_y	$\widehat{\text{PIQI}}^{max}$	$\widehat{\text{PSR}}$
26.1	6.72	30.36	16.44	0.34	0.60

3.7.3. Subset IQI Analysis: Continuous Covariate

The next question is whether there may be subpopulations defined by a covariate over which the estimate of the proportion of IQIs in the population may be reduced. This question will first be explored by defining populations defined by baseline cholesterol levels *Z*, which is a continuous covariate. The estimated means of the conditional distributions of cholesterol reduction depends on treatment T or R by generating linear models for both *X* and *Y* with *Z*. The sample results, including the estimated regression equations, are given in Table 3.3 and illustrated in Figure 3.8. Most importantly, note that since $\hat{\beta}_{XZ} = 0.35 \neq \hat{\beta}_{YZ} = 0.16$. Thus at least some of σ_D^2 is estimated to be explained by SI so that the reduction in conditional variance $\sigma_{D|Z}^2$ suggests that there will be a reduction in the PIQI over a range of *Z*.

Table 3.3: Conditional sample results based on baseline cholesterol level

Sample Results Based on Conditional Distributions					
$\bar{z}$	s_z	$\hat{x}_{z_0}$	$\hat{y}_{z_0}$	$s_{x\|z}$	$s_{y\|z}$
203.9	36.98	$-45.3+0.35z_0$	$-25.7+0.16z_0$	27.46	15.36

The impact of conditioning on baseline cholesterol on the PIQI and the PSR within subpopulations defined over values of Z defined by z_0 are given in Table 3.4. The results from row (a) provide a more comprehensive treatment of the IQI analysis without conditioning. The estimated standard error of $\widehat{\text{PIQI}}^{max}$ given in parentheses in Table 3.4 is a bootstrap estimate using 1000 bootstrap samples. The estimated standard error suggests that the PIQI^{max} may reasonably range from 0.25 to 0.43. Although the $\widehat{\text{PIQI}}^{min} = 0.082$ estimates that at least 8% of the individuals would do better on the control, the estimated standard error of 0.071 indicates that the PIQI^{min} may not be different from zero. The corresponding $\frac{1}{2}\text{PSR} = 0.30$ estimate is also reported with its corresponding standard error estimate equal to 0.0385. Figure 3.9 panel (a) provides a graphical display of the PSR, again illustrated as the shaded area.

Table 3.4 line (b) shows that by conditioning on Z, $\widehat{\text{PIQI}}^{min}$ is reduced to 0.055, when $z_0 = \bar{z} = 203.9$ and $\widehat{\text{PIQI}}^{max} = 0.326$. This improvement is only due to the reduction in $s_{x|z} = 27.46$ and $s_{y|z} = 15.36$ as opposed to $s_x = 30.36$ and $s_y = 16.44$, respectively. The corresponding reduction in the PSR is illustrated in Figure 3.9 panel (b) in which $\frac{1}{2}\text{PSR}_{204} = 0.294$. Reference to rows (c) and (d) of Table 3.4 along with Figures 3.8 and 3.9 show similar reductions in both the $\text{PIQI}_{z_0}^{min/max}$ and $\frac{1}{2}\text{PSR}_{z_0}$.

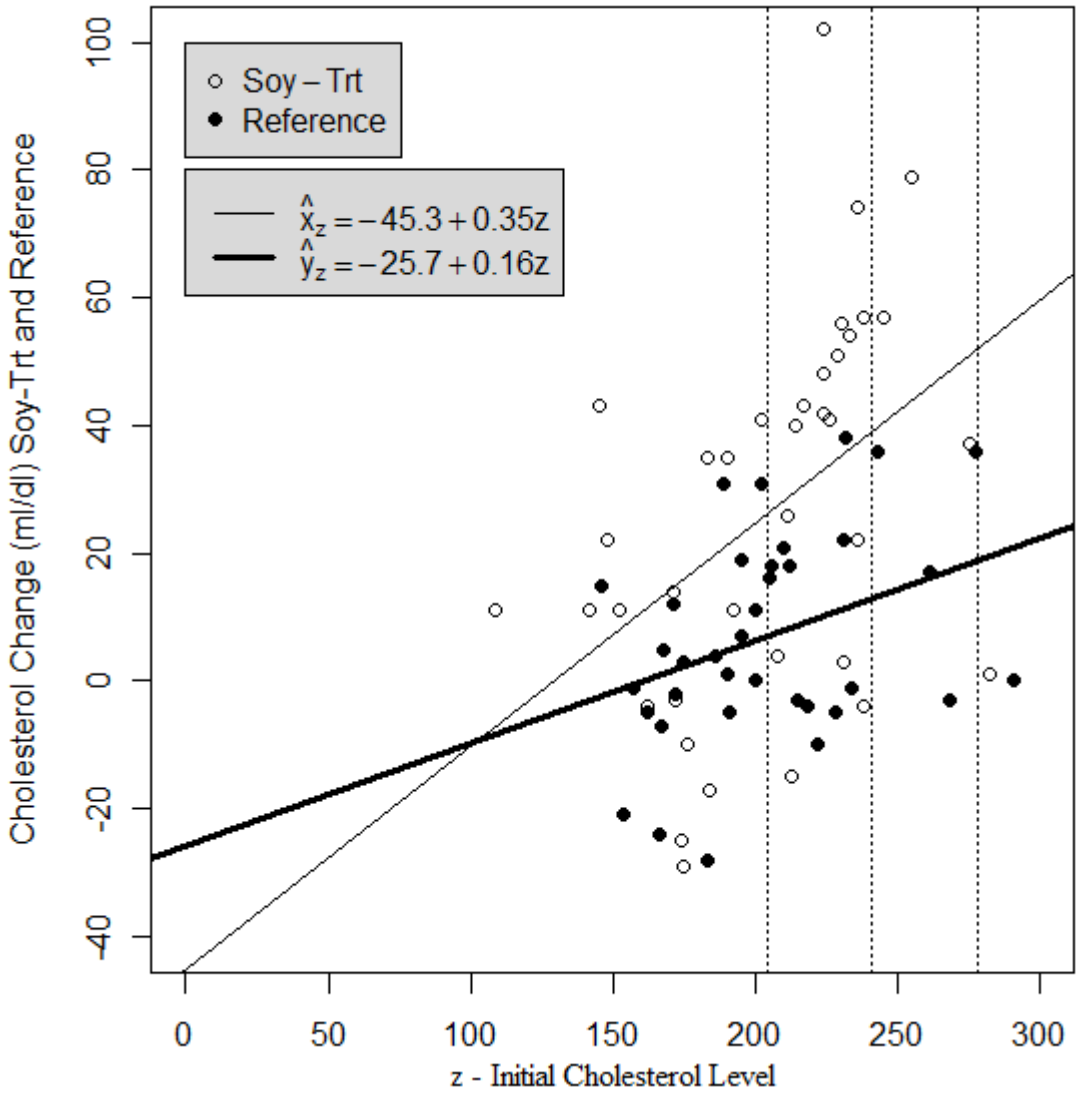


Figure 3.8: Estimated regression equations for soy treatment
Soy (X) and Reference (Y) treatments plotted against initial cholesterol level (Z), and their respective estimated regression equations given in Table 3.2. The vertical dotted lines are at the conditional PSR locations located at 204, 241, and 278, respectively, as shown row 2 of Table 2.4.

Table 3.4: Soy treatment PSR and PIQI results from conditioning on baseline cholesterol
Estimated marginal and conditional results for the PSR and the PIQI and their respective standard error estimates, shown in parentheses, between the change in cholesterol for soy treatment (*X*) and reference (*Y*). Conditional results are based on the baseline cholesterol level (*Z*). The corresponding graphs are displayed in Figure 3.8.

Graph/Row	z_0	½ PSR*	PIQI^{min*}	PIQI^{max*}
a	n/a	0.300 (0.0385)	0.082 (0.071)	0.340 (0.045)
b	204[1]	0.294 (0.039)	0.055 (0.067)	0.326 (0.050)
c	241[2]	0.249 (0.054)	0.015 (0.054)	0.296 (0.066)
d	278[3]	0.204 (0.073)	0.003 (0.057)	0.217 (0.087)

Footnotes: *The standard error estimates are based on 1000 bootstrap sample
1. $z_0 = \bar{z}$
2. $z_0 = \bar{z} + s_z$
3. $z_0 = \bar{z} + 2s_z$

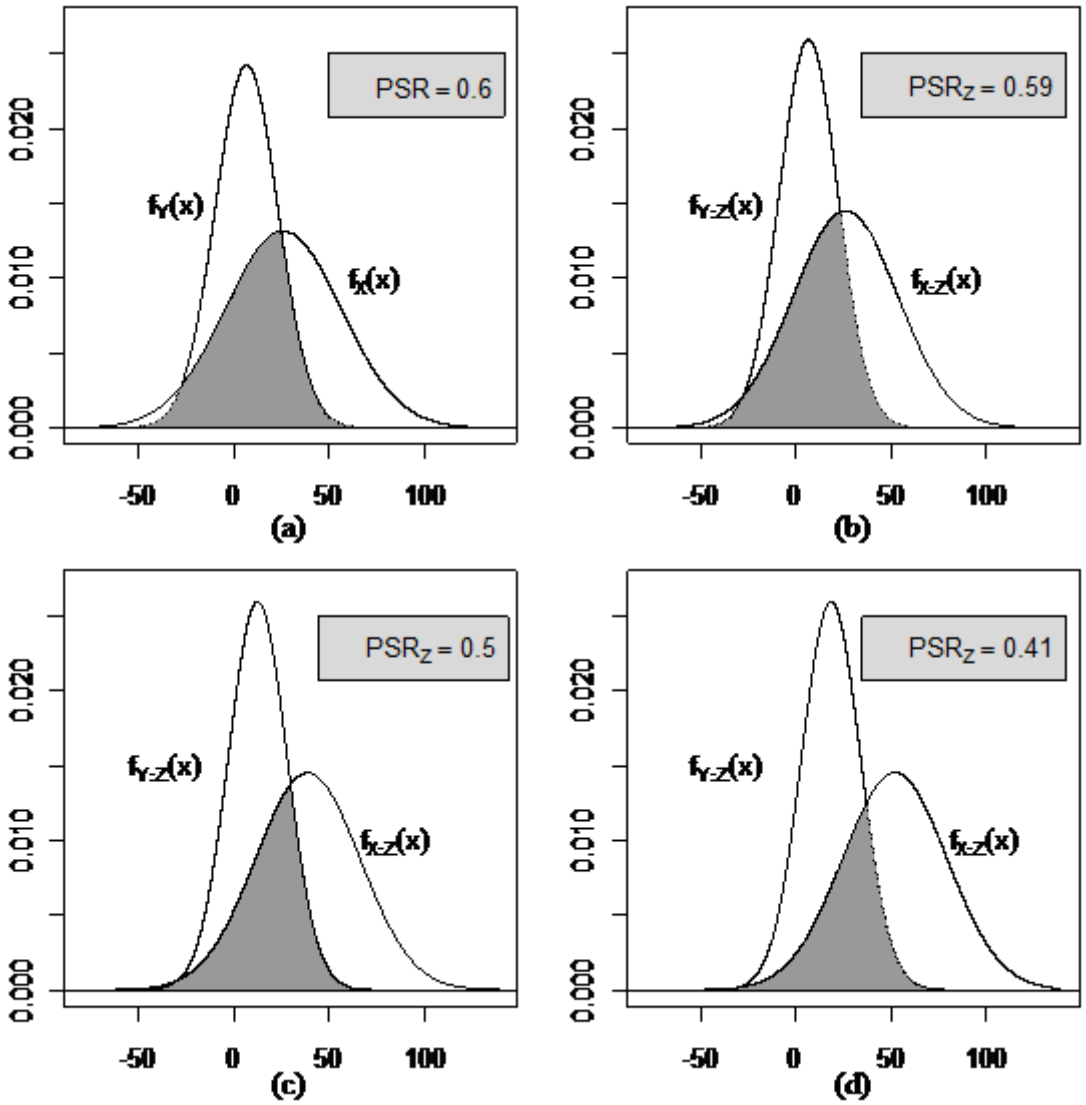


Figure 3.9: Soy treatment PSR and conditional PSR illustrations
Each panel displays the PSR, equal to the shaded area, as an overlap between the estimated densities and the estimated conditional densities of X and Y, where X is the cholesterol change for the soy-treatment group and Y is the cholesterol change for the control group. Panel (a) is the overlap of the marginal densities, and panels (b), (c), and (d) are conditioned on the baseline cholesterol level, and correspond to values of z_0 in Table 3.4, lines (b), (c), and (d).

3.7.4. Subset IQI Analysis: Categorical Covariate

The gender variable may be used to determine whether both the PSR and the $\text{PIQI}^{min/max}$ may be reduced within subsets of the population defined over gender. Due to a non significant interaction term between treatment and gender, the sums of squares due to the interaction is rather small indicating a small estimated reduction in the conditional variance $\sigma^2_{D|Gender}$ over σ^2_D. As such, a large difference in either the PSR or the PIQI would not be expected within any of the subsets. Nevertheless, Table 3.5 provides some final estimates of the PIQI^{min} and the PIQI^{max}. Although the PIQI seems to be smaller for men than women, the standard errors for these estimates (not reported) suggest that there may not be any difference.

Table 3.5: Soy treatment PIQI results from conditioning on gender

Group	PIQI^{min}	PIQI^{max}
Overall	0.082	0.340
Men	0.081	0.306
Women	0.154	0.356

Chapter 4 - IQI and the Skew Normal Distribution

4.1. Introduction and Overview

Since the results in the Chapter 3 require normality for both *X* and *Y*, this chapter studies the relationship between the PIQI and the PSR when either *X* or *Y*, or both, is not normally distributed. Gastwirth (1975, p. 33) used Figure 4.1 to introduce the PSR depicted as the area with diagonal lines, which he referred to as the 'overlap.' Figure 4.1 also serves as a reminder that such distributions, and thus such overlaps, occur naturally. This chapter investigates under what constraints, if any, the PSR can be used as a proxy for the PIQI without the restriction of bivariate normality. In particular, the main objective of this chapter is to investigate whether $\text{PIQI}^{max} \geq$ ½ PSR (i.e. proposition 2 given in Chapter 3) remains true when the bivariate normal constraint is relaxed. Given $\text{PIQI}^{max} \geq$ ½ PSR does not hold generally, the secondary objective is to explore the constraints under which $\text{PIQI}^{max} \geq$ ½ PSR does hold true.

Figure 4.1: A skewed distribution introduced by Gastwirth
This figure was originally published by Gastwirth (1975) to both exhibit and critique the PSR depicted as the shaded in area. Such distributions occur frequently so questions about the 'overlap' naturally result.

Because there are many bivariate distributions outside the normal, the first priority of this section will be to establish the conditions under which bivariate distribution the relationship between IQI and the PSR may be studied. The potential outcome framework and the need to establish a basis for analyzing the correlation between X and Y makes the selection of the bivariate distribution not only important, but challenging. Some of the reasons that the bivariate skew normal distribution makes a good choice include: a natural extension of the bivariate normal distribution, the correlation coefficient ρ_{XY} is the same correlation operator as that of the normal distribution, closed form expressions exist for both the PIQI and the PSR, and the variety of shapes that may be represented for both the joint and marginal skew normal distributions. Three other approaches were investigated; simulating the bivariate population, using a mixture of normal distributions, and copulas, but are not considered further here. Section 4.1 introduces the skew normal distribution. Sections 4.2 and 4.3 develop the PIQI and the PSR under the skew normal model, respectively. Non-parametric kernel density estimates of the PSR are investigated in Section 4.4. Section 4.5 shows that PIQI$^{max} \geq$ ½ PSR does not hold generally under the skew normal model, but shows the conditions under which this relationship is maintained. Finally, Section 4.6 addresses challenges related to establishing a useful relationship between an IQI and the PSR from the observed data, and Section 4.7 addresses future work under the skew normal model.

4.2. An Introduction to the Skew Normal Distribution

The joint skew normal distribution proposed in Azzalini and Valle (1996) and further developed by Azzalini and Capitanio (1999) offers some distinct advantages in this research including: the specification of a linear correlation parameter and the quantification of the PIQI in

closed form. However, before introducing the bivariate skew normal distribution, the next section offers a brief introduction to the univariate skew normal.

4.2.1. The Univariate Skew Normal Distribution

The skew normal distribution originally introduced by O'Hagan and Leonhard (1976) and more fully developed by Azzalini (1985) provides a natural extension to the normal distribution, since it can be formed by simply adding a shape parameter to the normal distribution, denoted by α herein. A skew normal random variable Z may then be denoted by $Z \sim SN(\mu, \sigma, \alpha)$. The pdf of Z is expressed as

$$f_Z(z) = \frac{2}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\alpha\frac{z-\mu}{\sigma}\right), \tag{4.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively. In Figure 4.2, μ and σ are fixed at 0 and 1, respectively. When $\alpha = 0$, Z is normally distributed (in this case standard normal since $\mu = 0$ and $\sigma = 1$) as demonstrated with the solid line. Also illustrated in Figure 4.2, setting α equal to ± 3 shows when $\alpha < 0$, a skew normal distribution is skewed to the left and is skewed to the right when $\alpha > 0$.

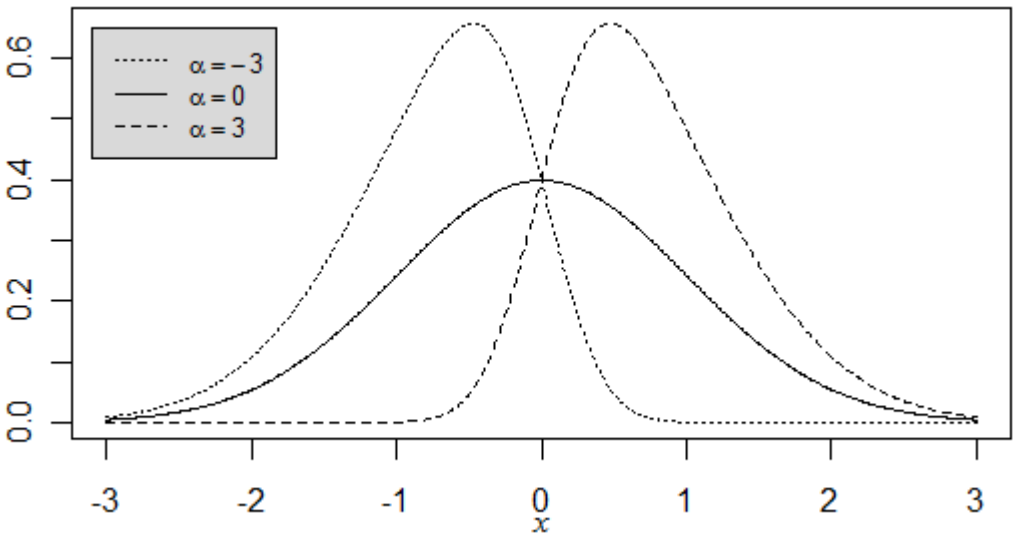


Figure 4.2: Introduction to the shape parameter in a skew normal distribution
Each of the three densities of Z, denoted by $f_Z(z)$, are generated from a univariate skew normal distribution with common location and scale parameters equal to $\mu = 0$ and $\sigma = 1$, respectively. The solid density curve is generated by setting $\alpha = 0$, which generates a normal density curve (standard normal in this case). Setting $\alpha = -3$ produced the dotted density curve and $\alpha = 3$ produced the dashed density curve, demonstrating that when $\alpha < 0$ the skew normal distribution is skewed to the left and that when $\alpha > 0$ the skew normal distribution is skewed to the right.

Furthermore, when $\alpha = 0$, $E(Z) = \mu$ and $Var(Z) = \sigma^2$. Thus like the normal distribution, μ is the location parameter, and σ is the scale parameter. However, unlike the normal distribution the mean and variance of Z are influenced by the shape parameter in the following way:

$$E(Z) = \mu + \sqrt{\frac{2\sigma^2\alpha^2}{\pi(1+\alpha^2)}} \qquad (4.2)$$

and

$$Var(Z) = \sigma^2\left(1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}\right). \tag{4.3}$$

Consequently, since the skew normal distribution generalizes the normal distribution, μ and σ^2 are not necessarily the mean and variance of *Z*, respectively.

Figure 4.3 illustrates the flexibility of the skew normal distribution, as well as its application to the research presented here, by emulating Gastwirth's hand drawn graph shown above as Figure 4.1. The skew normal densities of *X* and *Y* in Figure 4.3 are generated from $\mu_X = 4,\ \mu_Y = 0$, with a common $\sigma = 5$ and $\alpha = 3$, which gives $E(X) \cong 7.8$, $E(Y) \cong 3.8$, and a common $Var(X) = Var(Y) \cong 10.7$. The shaded area depicts the PSR approximately equal to 0.514. Although the PSR may be calculated from the marginal skew normal densities utilizing (2.8) given in Chapter 2, the PIQI is developed from the joint skew normal distribution of *X* and *Y* introduced in the next section.

4.2.2. The Bivariate Skew Normal Distribution

Complete coverage of the joint skew normal distribution is given in Azzalini and Valle (1996) and Azzalini and Capitanio (1999). Similar to the univariate skew normal distribution, a bivariate skew normal distribution of a random vector $\boldsymbol{Z}^T = (X \quad Y)^T$ is an extension of the bivariate normal distribution. Z is generated by adding a vector of shape parameters $\boldsymbol{\alpha}_J^T = \left(\alpha_{X_J}\ \alpha_{Y_J}\right)^T$ to a bivariate normal distribution given in (2.3) in Chapter 2, where α_{X_J} and α_{Y_J} denote the shape parameters of the joint distribution of *X* and *Y*, respectively.

Thus $\boldsymbol{Z} \sim SN_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha_J})$ denotes the bivariate skew normal random vector where

$$\boldsymbol{\mu}^T = (\mu_X \quad \mu_Y)^T \tag{4.4}$$

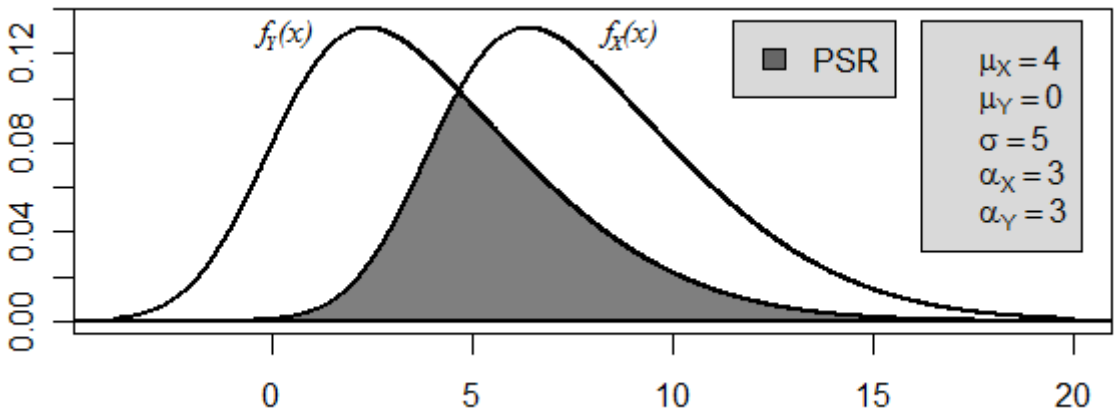


Figure 4.3: Replication of Gastwirth's overlap using skew normal densities
This figure is a recreation of the hand drawn distributions and the resulting PSR published by Gastwirth (1975) and shown above in Figure 1. Both $f_X(x)$ and $f_Y(x)$ are skew normal densities generated according to the specifications given in the legend resulting in $E(X) \cong 7.8$, $E(Y) \cong 3.8$, and a common $Var(X) = Var(Y) \cong 10.7$. The shaded region depicts the PSR $\cong 0.514$.

and

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \sigma^2 k\rho_{XY} \\ \sigma^2 k\rho_{XY} & \sigma^2 k^2 \end{pmatrix}. \tag{4.5}$$

The bivariate skew normal pdf is defined as

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = 2\phi_2(\boldsymbol{z}; \boldsymbol{\mu}, \Sigma)\Phi\big(\alpha^T\Sigma^{-1}(\boldsymbol{z} - \boldsymbol{\mu})\big), \tag{4.6}$$

where $\phi_2(\cdot)$ is the bivariate normal pdf and $\Phi(\cdot)$ is the univariate standard normal cdf as given in (4.1) above. Similar to the univariate case, when $\boldsymbol{\alpha}_J^T = (0 \quad 0)^T$, $\boldsymbol{Z}$ becomes bivariate normal. Perhaps the most important aspect of (4.5) is the retention of the correlation parameter ρ_{XY} as a measurement of the linear association between *X* and *Y*. Note also that the parameterization of the scale parameters in (4.4) is the same as that under bivariate normality, namely $\sigma_X^2 = \sigma^2$ so that $\sigma_Y^2 = \sigma^2 k^2$. Figure 4.4 illustrates the potential impact of adding $\boldsymbol{\alpha}_J^T \neq (0 \quad 0)^T$ to a

bivariate normal distribution. Panels (a) and (b) display density and contour plots, respectively, for a bivariate normal distribution with $\boldsymbol{\mu}^T = (3 \quad 0)^T$, $\boldsymbol{\sigma}^T = (2 \quad 2)^T$, and $\rho_{XY} = -0.5$. Panels (c) and (d) are the density and contour plots of a bivariate skew normal distribution formed as result of adding $\boldsymbol{\alpha}_J^T = (-3 \quad 3)^T$ to the bivariate normal distribution displayed in panels (a) and (b). The shape of $f_{\boldsymbol{Z}}(\boldsymbol{z})$ in panels (c) and (d) emphasize the complexity of working with the bivariate skew normal distribution. In fact, the next result confirms that adding the shape parameter $\alpha_J^T \neq (0 \quad 0)^T$ to a bivariate normal distribution generates a dependency between the marginal distributions of *X* and *Y* through α_J^T even when there is no correlation.

Result 4.1: Given $(X \quad Y)^T \sim SN_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$, where $\boldsymbol{\mu}$ and Σ are defined as in (4.4) and (4.5), respectively, $\boldsymbol{\alpha}^T = (\alpha_X \quad \alpha_Y)^T \neq (0 \quad 0)^T$, and $\rho_{XY} = 0$, then the marginal distributions of *X* and *Y* are not independent.

Proof:

Without loss of generality let $\mu = 0$ and $\sigma = 1$. Then, by (4.1) and (4.6)

$$f_X(x; \alpha_X) \cdot f_Y(y; \alpha_Y) = 4\,\phi(x)\phi(y)\Phi(x\alpha_X)\,\Phi(y\alpha_Y)$$

and

$$\begin{aligned} f_{XY}(x, y; \alpha_X, \alpha_Y, \rho_{XY} = 0) &= 2\phi_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \Phi\left([\alpha_X \quad \alpha_Y]^T \begin{bmatrix} x \\ y \end{bmatrix}\right) \\ &= 2\phi(x)\phi(y)\Phi(x\alpha_X + y\alpha_Y) \end{aligned}$$

Thus,

$$f_X(x; \alpha_X) \cdot f_Y(y; \alpha_Y) \neq f_{XY}(x, y; \alpha_X, \alpha_Y, \rho_{XY} = 0). \blacksquare$$

Due to not only the dependent structure, but also the skewed nature of a bivariate skew normal distribution, the calculation of the PIQI is more complex when $\alpha_J^T \neq (0 \quad 0)^T$ than when $\alpha_J^T = (0 \quad 0)^T$, which returns $(X \quad Y)^T$ to a bivariate normal distribution. Section 4.2 develops the framework for expressing and computing the PIQI under the $\alpha_J^T \neq (0 \quad 0)^T$ constraint.

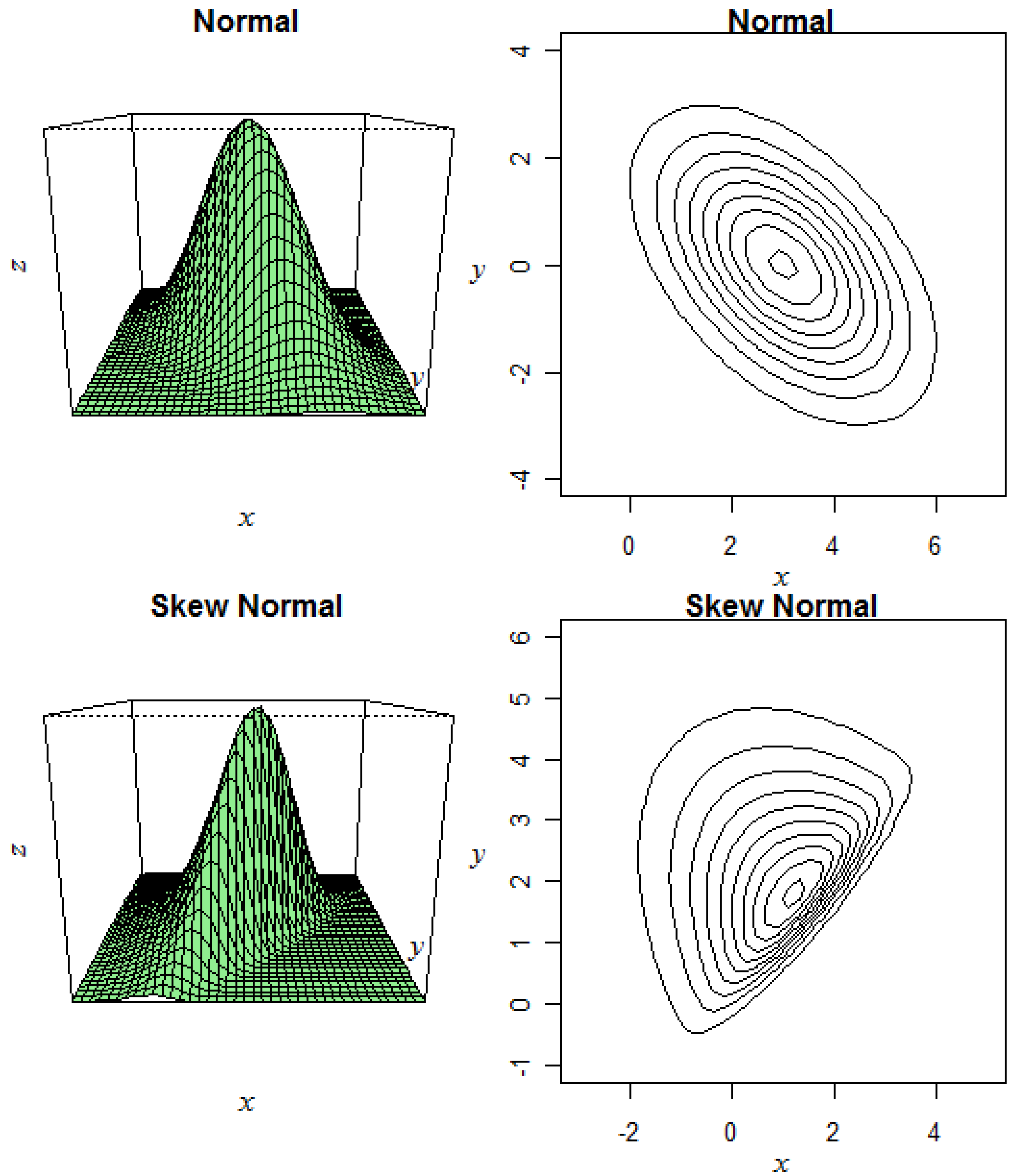


Figure 4.4: Introduction to the bivariate skew normal distribution
Panels (a) and (b) depict the bivariate density and contour plot, respectively, for a bivariate normal distribution with location parameter vector $\boldsymbol{\mu}^T = (\mu_X \quad \mu_Y)^T = (3 \quad 0)^T$, scale parameter vector $\boldsymbol{\sigma}^T = (\sigma_X \quad \sigma_Y)^T = (2 \quad 2)^T$, and correlation $\rho_{XY} = -0.5$. The skew in the joint distribution of X and Y displayed in panels (c) and (d) are the result of adding a shape parameter vector $\boldsymbol{\alpha}^T = (\alpha_X \quad \alpha_Y)^T = (-3 \quad 3)^T$ to a bivariate normal distribution displayed in panels (a) and (b). The skew present in panels (c) and (d) create some new challenges in the calculation of the PIQI.

4.3. Assessing IQI given a Bivariate Skew Normal Distribution

In Chapter 3 the PIQI was defined in the context of the bivariate normal distribution. At times the notation $\text{PIQI}^{\rho_{XY}}$ was used to emphasize the PIQI's dependency on ρ_{XY} under the normality constraint. Although the bivariate distribution may be altered, the definition $\text{PIQI}^{\rho_{XY}} = P(D < 0)_{\rho_{XY}}$ holds irrespective of the distribution of D. However, the calculation of the $\text{PIQI}^{\rho_{XY}}$ and its dependency on ρ_{XY} may change over different bivariate distributions. As shown in Chapter 3, when the distribution of X and Y is bivariate normal, the distribution of $D = X - Y$ is normal giving rise to (3.1). In this chapter, since the bivariate distribution is no longer constrained by normality, the distribution of D needs to be identified before the $\text{PIQI}^{\rho_{XY}}$ can be calculated and its dependency on ρ_{XY} can be delineated. The next section addresses these issues in the context of the bivariate skew normal distribution.

4.3.1. The Distribution of D

Azzalini and Capitanio (1999) provide general results for the distribution of a linear combination of $(X \quad Y)^T$ when $(X \quad Y)^T$ follows a bivariate skew normal distribution. Application of these results makes the development of the distribution of $D = X - Y$ possible. As given above in Section 4.2.1, a bivariate skew normal distribution of $(X \quad Y)^T$ is denoted as $(X \quad Y)^T \sim SN_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha}_J)$ where both $\boldsymbol{\mu}$ and Σ are given in (4.4) and (4.5), respectively. From Azzalini and Capitanio (1999), a linear combination of $(X \quad Y)^T$ defined by a 2×1 matrix A is

$$A^T \begin{pmatrix} X \\ Y \end{pmatrix} \sim SN(A^T \boldsymbol{\mu}, \Sigma_{\mathrm{A}}, \alpha_A) \qquad (4.7a)$$

where

$$\Sigma_{\mathrm{A}} = A^T \Sigma \mathrm{A} \qquad (4.7\mathrm{b})$$

$$\alpha_A = \frac{\sqrt{\Sigma_A}\Sigma_A^{-1}B_A^T\boldsymbol{\alpha}}{\sqrt{1+\boldsymbol{\alpha}^T(P - B_A^T\Sigma_A^{-1}B_A^T)\boldsymbol{\alpha}}} \tag{4.7c}$$

$$B = \left(\sqrt{\Sigma_\sigma}\right)^{-1}\Sigma A \tag{4.7d}$$

$$\Sigma_\sigma = \begin{pmatrix}\sigma_X & 0\\ 0 & \sigma_Y\end{pmatrix} \tag{4.7e}$$

$$P = \begin{pmatrix}1 & \rho_{XY}\\ \rho_{XY} & 1\end{pmatrix}. \tag{4.7f}$$

Note that Σ_σ is a diagonal matrix composed of the square root of the diagonal elements of the variance-covariance matrix of $(X \quad Y)^T$, and P is the correlation matrix for a bivariate standard normal random vector.

The distribution of $D = X - Y$ can then be generated by taking $A^T = (1 \quad -1)^T$. Consequently, D is distributed as a univariate skew normal random variable with $\mu_D = \mu_X - \mu_Y$, $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY} = \sigma^2(1 + k^2 - 2k\rho_{XY})$, and α_D equal to

$$\frac{\dfrac{\alpha_{X_J}(1-k\rho_{XY}) - \alpha_{Y_J}(k-\rho_{XY})}{\sqrt{1+k^2-2k\rho_{XY}}}}{\sqrt{1+\alpha_{X_J}^2\left[1-\dfrac{(1-k\rho_{XY})^2}{1+k^2-2k\rho_{XY}}\right] + 2\alpha_{X_J}\alpha_{Y_J}\left[\rho_{XY} - \dfrac{(1-k\rho_{XY})(\rho_{XY}-k)}{1+k^2-2k\rho_{XY}}\right] + \alpha_{Y_J}^2\left[1-\dfrac{(\rho_{XY}-k)^2}{1+k^2-2k\rho_{XY}}\right]}}, \tag{4.8}$$

as shown in the appendix (see A.4.1). Thus α_D is a function of both ρ_{XY} and k, and therefore, $\text{PIQI}^{\rho_{XY}}$ is a function of ρ_{XY} not only through σ_D^2 but also through α_D. The $\text{PIQI}^{\rho_{XY}}$ then may be strictly defined as

$$\text{PIQI}^{\rho_{XY}} = P(D<0)_{\rho_{XY}} = \int_{-\infty}^{0}\frac{2}{\sigma_D}\phi\left(\frac{d-\mu_D}{\sigma_D}\right)\Phi\left(\alpha_D\frac{d-\mu_D}{\sigma_D}\right)dd, \tag{4.9}$$

where $d = x - y$. Similar to the normal cdf, there is not a closed form solution for the skew normal cdf. However, the '*psn'* function in the '*sn'* package in R (www.r − project.org)

generates a numerical calculation of (4.9) used in the remainder of this work (see appendix B.4.1).

4.3.2. The Shape Parameter for the Distribution of D and the PIQI

As highlighted in the previous section, when *D* follows a skew normal distribution, α_D plays an important role in the calculation of the PIQI$^{\rho_{XY}}$. This section delineates the relationship between the PIQI$^{\rho_{XY}}$ and α_D. Since α_D is a function of both ρ_{XY} and k, restricting $k = 1$ makes the relationship between α_D and the PIQI$^{\rho_{XY}}$ more transparent. As shown in the appendix (see A.4.2), given $k = 1$, the derivation of α_D is

$$\alpha_D(k=1) = \frac{\left(\alpha_{X_J} - \alpha_{Y_J}\right)\left(\frac{1-\rho_{XY}}{2}\right)^{\frac{1}{2}}}{\sqrt{1 + \left(\frac{1+\rho_{XY}}{2}\right)\left(\alpha_{X_J} + \alpha_{Y_J}\right)^2}}. \tag{4.10}$$

The role of the joint shape parameters α_{X_J} and α_{Y_J} in (4.10) is more apparent than in(4.8). In fact, since the denominator of $\alpha_D(k=1) > 0$ over all $-1 \leq \rho_{XY} \leq 1$, $\alpha_{X_J} \in \mathbb{R}$, and $\alpha_{Y_J} \in \mathbb{R}$, the direction of the skew in the distribution of D $\left(i.e.\ \text{the sign on } \alpha_D(k=1)\right)$ is completely determined by the relationship between α_{X_J} and α_{Y_J}. That is $\alpha_D(k=1) \leq 0$ when $\alpha_{X_J} \leq \alpha_{Y_J}$. Otherwise, $\alpha_D(k=1) > 0$. Furthermore, when $\alpha_{X_J} = \alpha_{Y_J}$, $\alpha_D(k=1) = 0$ so that *D* is normally distributed. In this case, both the calculation and the behavior of the PIQI$^{\rho_{XY}}$ are the same as that of Section 3. That is the PIQI$^{\rho_{XY}}$can be calculated from (3.1), wherein the PIQI$^{\rho_{XY}} = \Phi\left(\frac{-\mu_D}{\sigma_D}\right)$, and the PIQI^{min} and PIQI^{max} are achieved at ρ_{XY} equal to 1 and -1, respectively as shown in (3.2) and (3.3).

A graphical illustration of (3.1), (3.2), and (3.3) is given in panel (a) of Figure 4.5 where, since $k = 1$ and $\alpha_{X_J} = \alpha_{Y_J} = 1$, *D* is normally distributed. Thus the PIQI$^{\rho_{XY}}$ is a function

of ρ_{XY} through σ_D^2 only, and $\sigma_D^2 = \sigma^2(1 + k^2 - 2k\rho_{XY})$ is a strictly decreasing function of ρ_{XY}. Therefore, from(3.1) $\text{PIQI}^{\rho_{XY}}$ is a strictly decreasing function of ρ_{XY} and illustrated by the solid line. Panel (c) of Figure 4.5 provides an alternative perspective of this phenomenon. The effect of ρ_{XY} on the $\text{PIQI}^{\rho_{XY}}$ is illustrated using the three pdf's of *D* in panel (c) where $f_D(d; \rho_{XY} = -1)$ is displayed by the solid line, $f_D(d; \rho_{XY} = -0.5)$ is displayed as the dashed line, and $f_D(d; \rho_{XY} = 0)$ is displayed as the dotted line. Since σ_D^2 is decreasing as ρ_{XY} increases and $\mu_D = E(D)$ is unchanged by ρ_{XY}, $\text{PIQI}^{\rho_{XY}} = P(D < 0)_{\rho_{XY}}$ is shrinking with increases in ρ_{XY}. Graphically this means that the shaded area under the density curve decreases as ρ_{XY} increases. So in Figure 4.5 panel (c), the largest amount of shaded area occurs under the density represented by the solid line, which is generated under $\rho_{XY} = -1$.

Figure 4.5 panel (b) is also a display of $\text{PIQI}^{\rho_{XY}}$, but in this case $\alpha_{X_J} = 2 > \alpha_{Y_J} = 1$ so that $\alpha_D(k = 1; \rho_{XY}) > 0$ over all $-1 \leq \rho_{XY} \leq 1$, and thus the distribution of *D* is skew normal. As illustrated with the solid line in panel (b), the effect is that $\text{PIQI}^{max} \cong \text{PIQI}^{\rho_{XY}=0}$ instead of $\rho_{XY} = -1$. This example serves as an illustration and a proof of the next result.

Result 4.2: When $(X \quad Y)^T$ is not bivariate normal, the PIQI^{max} may not occur at $\rho_{XY} = -1$. Figure 4.5 panel (d) illustrates why the PIQI^{max} may not be equal to $\text{PIQI}^{\rho_{XY}=-1}$ when *D* is distributed as skew normal. As opposed to the normal case when $\alpha_D = 0$, when $\alpha_D \neq 0$, the $\text{PIQI}^{\rho_{XY}}$ is a function of ρ_{XY} through both σ_D^2 and α_D. For both panels (b) and (d), $\alpha_D > 0$ since $\alpha_{X_J} > \alpha_{Y_J}$, and therefore, $E(D) > \mu_D = 2$. For example, when $\rho_{XY} = -1$, $\alpha_D(k = 1, \rho_{XY} = -1) = 1$ and $E(D) \cong 4.3$. And when $\rho_{XY} = 0$, $\alpha_D(k = 1, \rho_{XY} = 0) \cong 0.3$ and $E(D) \cong 2.7$. The shift in $E(D)$ coupled with the shorter left tail and longer right tail of $f_D(d; \alpha_D, \rho_{XY} = -1)$, represented by the solid line, than that of $f_D(d; \alpha_D, \rho_{XY} = 0)$, represented

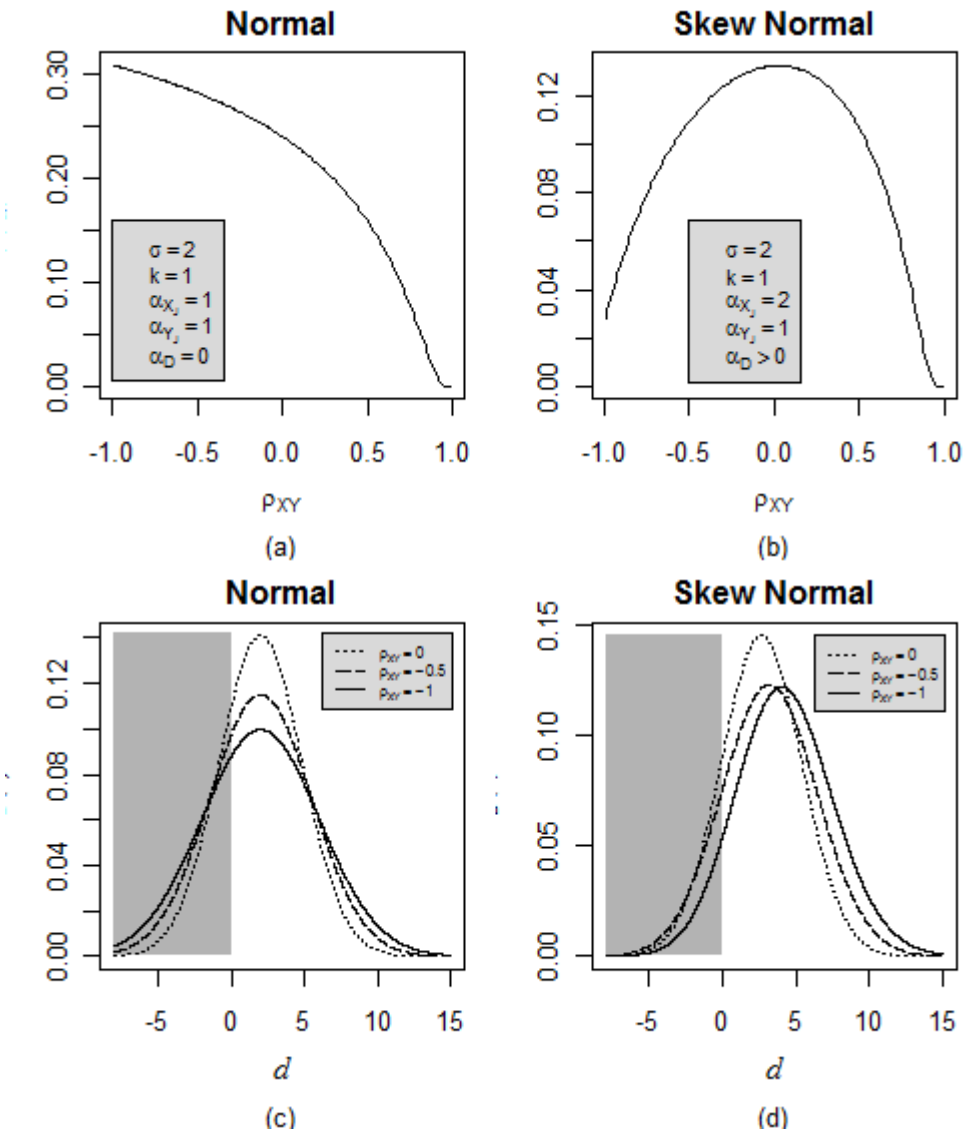


Figure 4.5: An illustration of the shape parameter on the distribution of *D*
Each of these graphs are generated from $\mu_X = 2$, $\mu_Y = 0$, $\sigma = 2$, and $k = 1$. In panels (a) and (c) $\alpha_{X_J} = \alpha_{Y_J} = 1$ so that $\alpha_D = 0$ resulting in *D* being normally distributed. In panels (b) and (d) $\alpha_{X_J} = 2$ and $\alpha_{Y_J} = 1$ so that $\alpha_D > 0$ resulting in *D* being distributed as a skew normal. Panel (a) illustrates that the PIQI$^{\rho_{XY}}$ is a strictly decreasing function of ρ_{XY} when *D* is normally distributed, as in Section 3, so that PIQI^{min} = PIQI$^{\rho_{XY}=1} \leq$ PIQI$^{\rho_{XY}} \leq$ PIQI$^{\rho_{XY}=-1}$ = PIQI^{max}. Panel (b) shows that when *D* is not normally distributed the PIQI$^{\rho_{XY}}$ may not be maximized at $\rho_{XY} = -1$, since for this example PIQI^{max} = PIQI$^{\rho_{XY}\cong 0}$. Panel (c) shows that when $\alpha_D = 0$, the PIQI$^{\rho_{XY}}$ is a function of ρ_{XY} through σ_D^2 so that as ρ_{XY} increases σ_D^2 decreases causing PIQI$^{\rho_{XY}}$ to decrease. Graphically this means that the shaded area under the density curve decreases as ρ_{XY} increases. Panel (d) illustrates that when *D* is skew normal ($i.e.\ \alpha_D \neq 0$), PIQI$^{\rho_{XY}}$ is a function of ρ_{XY} through both σ_D^2 and α_D. Furthermore, in panels (b) and (d), since $\alpha_D > 0$, $E(D) > \mu_D = 2$. The shape of $f_D(d)$ is constrained by $E(D)$, $\mu_D = 2$, and $\alpha_D > 0$. Graphically this means that $f_D(d; \rho_{XY} = -1, \alpha_D > 0)$ represented by the solid line is shifted to the right and there is less shaded area under this density curve than under $f_D(d; \rho_{XY} = 0, \alpha_D > 0)$ represented by the dotted line.

by the dotted line, is what ultimately causes the loss of the strictly decreasing behavior of the PIQI$^{\rho_{XY}}$ as seen in panel (a). Again, the shaded area in panel (d) is illustrating PIQI$^{\rho_{XY}} = P(D < 0)_{\rho_{XY}}$. Due to the shift in $E(D)$ caused by $\alpha_D > 0$, the shaded area under $f_D(d; \alpha_D, \rho_{XY} = -1)$ is less than that of $f_D(d; \alpha_D, \rho_{XY} = 0)$.

Not surprisingly, when $\alpha_D < 0$, the result of $\alpha_{X_J} < \alpha_{Y_J}$ and $k = 1$, there is a similar but opposite effect on the shape of $f_D(d; \alpha_D)$, and therefore, $E(D)$. That is, $E(D) \leq \mu_D$ when $\alpha_D \leq 0$. Propositions 4.1 and 4.2 given next show that $f_D(d; \alpha_D)'s$ shift of to the left for smaller values of ρ_{XY} is important in the development of the PIQI$^{\rho_{XY}}$. In each case it is assumed that *X* and *Y* are distributed as skew normal, $\alpha_{X_J} \leq \alpha_{Y_J}$, and $k = 1$.

❖ **Proposition 4.1**

Given *X* and *Y* are bivariate skew normal, $\alpha_{X_J} \leq \alpha_{Y_J}$, and $k = 1$

$$\alpha_D \text{ is strictly increasing on } -1 < \rho_{XY} < 1$$

Proof: see appendix A.4.3

❖ **Proposition 4.2**

Given *X* and *Y* are bivariate skew normal, $\boldsymbol{\alpha_{X_J} \leq \alpha_{Y_J}}$, and $\boldsymbol{k = 1}$

$$\text{PIQI}^{\rho_{XY}} \text{ is strictly decreasing in } -1 < \rho_{XY} < 1$$

Proof: Follows from proposition 4.1, but a formal proof is given in A.4.4 of the appendix.

Corollary 4.1: Given *X* and *Y* are bivariate skew normal, $\boldsymbol{\alpha_{X_J} \leq \alpha_{Y_J}}$, and $\boldsymbol{k = 1}$

$$\text{PIQI}^{max} = \text{PIQI}^{\rho_{XY}=-1}$$

Proof: Follows from proposition 4.2.

This section has highlighted some additional challenges in the derivation of the PIQI$^{\rho_{XY}}$ under skew normality. Furthermore, important conditions have been set forth under which the PIQI$^{\rho_{XY}}$ behaves similarly for both the normal and the skew normal distribution. These results

indicate that the relationship between the PIQI$^{\rho_{XY}}$ and the PSR will be somewhat more complex when D is distributed as skew normal since the skewness property of D, exhibited through α_D, not only affects the shape of $f_D(d)$, but also $E(D)$. Furthermore, the dependence structure of the bivariate skew normal distribution resulting from $\boldsymbol{\alpha}_J$ and ρ_{XY} described in the previous section causes additional challenges. In preparation for such connections, the next section develops the PSR under the bivariate skew normal model.

4.4. The PSR under Skew Normality

Work on the development of the PSR or density overlap outside the normal distribution has been limited. Inman (see Bradley, 1985) and Bradley and Piantadosi (1992) are credited for work involving non-parametric PSRs, but unfortunately, their work was not published. As with the PIQI, the difficulty in the development of the PSR without the normal constraint originates from the plethora of shapes over which the overlap may be formed. This section provides the structure for the calculation of the PSR within the skew normal model. Once the nature of the PSR has been developed, connections with the PIQI will be explored.

4.4.1. Definitions and Properties of the PSR

Irrespective of the distributions of X and Y, the definition of the PSR given in (2.8) remains unaffected. An illustration of a PSR generated under both X and Y distributed as skew normal is given Figure 4.3 in which the shaded area represents the PSR. As characterized in Section 4.2, the number and location of the crossing points of the density curves of X and Y at which $f_X(u) = f_Y(v)$ $(i.e.\ u = v)$ are important elements of the PSR calculation. Recall that when X and Y are normally distributed and $f_X(u) \neq f_Y(v)$, there are at most two points at which the pdf's of X and Y cross. When there is a single crossing point, when $k = 1$, the crossing point

is denoted by x_C. When $k \neq 1$, two crossing points exist; a lower and an upper denoted by x_L and x_U, respectively. When X and Y are normally distributed, a closed form expression for the derivation of the crossing points x_E, x_L, and x_U was given in (2.9). Not surprisingly, when X and Y are distributed as skew normal, the maximum number of crossing points increases. Furthermore, due to the shape parameter of the skew normal density given in (4.1) a general closed form expression for the crossing points does not exist, although in some cases a formula for the crossing point does exist.

The challenge of calculating crossing points stems from the myriad of shapes possible for $f_X(u)$ and $f_Y(v)$ and their relationship when X and Y are skew normal. These combinations are a product of the marginal shape parameters denoted by α_{X_M} and α_{Y_M} and scale parameters summarized by k. For example, when $\alpha_{X_M} = \alpha_{Y_M}$ and $k = 1$, there is exactly one crossing point. An example of this type is shown in Figure 4.3, where $\alpha_{X_M} = \alpha_{Y_M} = 3$. In this case there is no formula for the crossing point and the PSR must be solved numerically as discussed below. Generally, the crossing points are influenced by whether α_{X_M} and α_{Y_M} are greater than 0, less than 0, equal, whether $\alpha_{X_M} < \alpha_{Y_M}$ or $\alpha_{X_M} > \alpha_{Y_M}$, and whether $k = 1$, $k < 1$, or $k > 1$. The calculation of the PSR depends upon the combination of these parameters. Four examples are given in Figure 4.6. Crossing points x_L, and x_U are used in the same way as for the normal case when two crossing points exist. When there is only one crossing point, it is labeled x_C instead of x_E as it is in the normal case. However, when a combination of the densities of X and Y generate three crossing points, the point x_I is added to the points x_L and x_U to denote an "interior" crossing point. Each crossing point in Figure 4.6 is labeled directly below the dotted line within each panel. Furthermore, common with earlier examples, the shaded area in each panel

identifies the PSR. The details necessary for calculating the PSR in each of the four panels in Figure 4.6 are given in Table 4.1. Although these four examples do not provide a

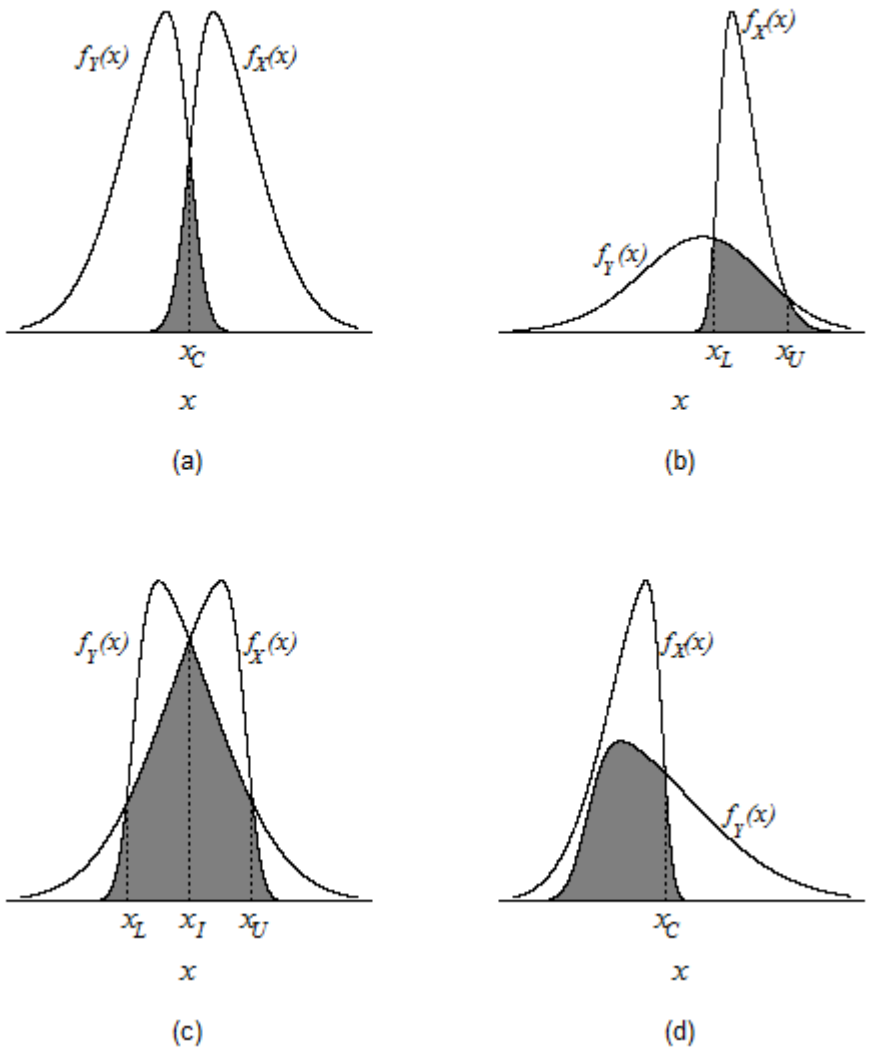


Figure 4.6: Initial PSR examples under the skew normal
This figure illustrates how the number of crossing points and the PSR shown as the shaded area involving two skew normal distributions are function of α_X, α_Y, and k . Although many patterns exist, only four examples are given here. However, the number of crossing points will always equal either one, two, or three. Table 1 complements this figure by provide the details for α_X, α_Y, and k for each panel. The first column of table 1 identifies the apporpriate graph, and column (6) shows the algebraic representation of the PSR.

comprehensive treatment of the number of the potential combinations over which the PSR may exist, they provide insight into how the PSR changes based on changes in α_{X_M}, α_{Y_M}, and k. The letter given in column (1) of Table 4.1 corresponds to the panel given in Figure 4.6. A complete treatment of Table 4.1 and Figure 4.6 is given next to provide a general framework for understanding the PSR and deriving its value.

Table 4.1: **Complement to PSR examples**
This table complements Figure 4.6. Column (1) identifies the corresponding panel in Figure 4.6. Columns (2) – (4) represent the general parameterizations of the skew normal distributions, and columns (5) and (6) show the resulting crossing point(s) and the PSR delineation, respectively.

Figure 4.6 Panel (1)	α_{X_M} **(2)**	α_{Y_M} **(3)**	k **(4)**	**Crossing Point(s) (5)**	**PSR (6)**
a	$\alpha_{X_M} > 0$	$\alpha_{Y_M} < 0$	$k = 1$	x_C	$P(X \leq x_C) + P(Y \geq x_C)$
b	$\alpha_{X_M} > 0$	$\alpha_{Y_M} < 0$	$k > 1$	x_L, x_U	$P(X \leq x_L) + P(x_L \leq Y \leq x_U) + P(X \geq x_U)$
c	$\alpha_{X_M} < 0$	$\alpha_{Y_M} > 0$	$k = 1$	x_L, x_I, x_U	$P(Y \leq x_L) + P(x_L \leq X \leq x_I) + P(x_I \leq Y \leq x_U) + (X \geq x_U)$
d	$\alpha_{X_M} < 0$	$\alpha_{Y_M} > 0$	$k > 1$	x_C	$P(Y \leq x_C) + P(X \geq x_C)$

The nature of α_{X_M}, α_{Y_M}, and k given in columns (2) through (4) do not give specific values, but represent common relationships between the parameters of the distributions of *X* and *Y*. For each these examples and throughout this work, the location parameters are constrained by $\mu_X > \mu_Y$. For simplification, the shape parameters given in columns (2) and (3) are additive inverses of each other $(i.e. \alpha_{X_M} = -\alpha_{Y_M})$ and may take on any value based on this constraint. The relationship between the scale parameters σ_X and σ_Y is captured by k and is restricted to either $k = 1$ or $k > 1$ as identified in column (4). The last two columns of Table 4.1 are for

calculating the PSR. Column (5) gives the number and notation for the crossing points, and column (6) provides a concise statement for the calculation of the PSR as an alternative to the more general formula given in (2.8) or the closed form expression used in the normal case given in (2.10).

Reference to row one of Table 4.1 then gives the description of panel (a) of Figure 4.6. This panel illustrates the general case constrained by $\alpha_{X_M} > 0$ and $\alpha_{Y_M} < 0$, so that the skews are in the opposite direction. Further, since $\alpha_{X_M} = -\alpha_{Y_M}$ and $k = 1$, $Var(X) = Var(Y)$, and $E(X) > E(Y)$. Since there is only one crossing point denoted by x_C as shown in column (5), and the PSR $= P(X \leq x_C) + P(Y \geq x_C)$ as given in column (6). For this case, the crossing point can be obtained in closed form as given in the following result:

Result 4.4: Given X and Y are bivariate skew normal, $\mu_X > \mu_Y$, $\alpha_{X_M} > 0$, $\alpha_{Y_M} < 0$, $\alpha_{X_M} = -\alpha_{Y_M}$ and $k = 1$, there is a single crossing point of $f_X(x)$ and $f_Y(x)$ given by

$$x_C = \frac{\mu_X + \mu_Y}{2}$$

Proof:

$$f_X(x_C) = f_X\left(\frac{\mu_X + \mu_Y}{2}\right) = \frac{2}{\sigma}\phi\left(\frac{\frac{\mu_X + \mu_Y}{2} - \mu_X}{\sigma}\right)\Phi\left(\alpha_X\frac{\frac{\mu_X + \mu_Y}{2} - \mu_X}{\sigma}\right)$$

$$= \frac{2}{\sigma}\phi\left(\frac{\mu_Y - \mu_X}{2\sigma}\right)\Phi\left(-\alpha_Y\frac{\mu_Y - \mu_X}{2\sigma}\right)$$

$$= \frac{2}{\sigma}\phi\left(\frac{\mu_Y - \mu_X}{2\sigma}\right)\Phi\left(\alpha_Y\frac{\mu_X - \mu_Y}{2\sigma}\right)$$

$$f_Y(x_C) = f_Y\left(\frac{\mu_X + \mu_Y}{2}\right) = \frac{2}{\sigma}\phi\left(\frac{\frac{\mu_X + \mu_Y}{2} - \mu_Y}{\sigma}\right)\Phi\left(\alpha_Y\frac{\frac{\mu_X + \mu_Y}{2} - \mu_Y}{\sigma}\right)$$

$$= \frac{2}{\sigma}\phi\left(\frac{\mu_X - \mu_Y}{2\sigma}\right)\Phi\left(\alpha_Y\frac{\mu_X - \mu_Y}{2\sigma}\right),$$

and since $\phi(\cdot)$ is the standard normal pdf $\phi\left(\frac{\mu_Y - \mu_X}{2\sigma}\right) = \phi\left(\frac{\mu_X - \mu_Y}{2\sigma}\right)$. Thus $f_X(x_C) = f_Y(x_C)$. ■

Again, although the crossing point for the case shown in panel (a) has a closed form solution, this is not the case in general.

Having obtained the crossing point, the PSR may then be calculated by application of (2.8) and column (6) as follows:

$$\text{PSR} = \int_{-\infty}^{\infty} min\big(f_X(u), f_Y(v)\big)dx = P(X \le x_C) + P(Y \ge x_C)$$

$$= \int_{-\infty}^{x_C} f_X(u)dx + 1 - \int_{x_C}^{\infty} f_Y(v)dx$$

$$= \int_{-\infty}^{x_C} \frac{2}{\sigma_X}\phi\left(\frac{u-\mu_X}{\sigma_X}\right)\Phi\left(\alpha_X\frac{u-\mu_X}{\sigma_X}\right)du + 1 - \int_{x_C}^{\infty} \frac{2}{\sigma_Y}\phi\left(\frac{v-\mu_Y}{\sigma_Y}\right)\Phi\left(\alpha_Y\frac{v-\mu_Y}{\sigma_Y}\right)dv\,.$$

Since $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal pdf and cdf, respectively, once x_C is obtained, the PSR can be calculated by application of these functions. Alternatively, the *psn* function supplied by the '*sn'* package in R may be utilized for a direct calculation of the PSR. Appendix B (see B.4.2) provides a function for calculating the PSR using the *psn* function, which was used throughout this work.

Reference to Table 4.1 shows that the shift from panel (a) to panel (b) of Figure 4.6 is made by constraining $k > 1$. Given all other constraints from panel (a) are held constant, this constraint generates two crossing points labeled x_L and x_U as shown in panel (b) and cross-referenced in column (5). Similar to the case in which both *X* and *Y* are normally distributed, this change necessitates a change in the calculation of the PSR by including three areas given as $\text{PSR} = P(X \le x_L) + P(x_L \le Y \le x_U) + P(X \ge x_U)$ in column(6). However, due to the

complexity of working with the density of a skew normal distribution given in (4.1), there is not a closed form solution for x_L or x_U in this case. When no closed form solution exists, a numerical algorithm is used, which employs the previously mentioned '*sn*' package in R to solve for the crossing points. The program for finding both the number and value of the crossing points is included in appendix B (see B.4.2). By taking $\varepsilon > 0$ the numerical algorithm can solve for $|f_X(x) - f_Y(x)| < \varepsilon$, and thus the crossing points can be calculated to the desired accuracy. When it is important to recognize when the PSR is being evaluated using numerically solved crossing points, the PSR will be denoted as PSR_N. Otherwise, the standard PSR contraction without the subscript will be employed.

Panel (c) of Figure 4.6 returns to $k = 1$, but reverses the shape parameters so that $\alpha_{X_M} < 0$ and $\alpha_{Y_M} > 0$, which illustrates a case involving three crossing points as identified in column (5). Each point is solved for numerically, and the PSR_N is computed by combining the four shaded areas separated by the dotted lines shown in panel (c), corresponding to column (6) in Table 4.1. The change from panel (c) to panel (d) is an example of the challenge of deriving the PSR when *X* and *Y* are skew normal. By simply constraining $k > 1$ the number of crossing points may go from three to one even though the shape parameters remain unchanged. Again, the crossing point for $f_X(u)$ and $f_Y(v)$ denoted as x_C can be solved for numerically and the PSR_N can be obtained by adding the two components of the shaded area in panel (d) comprised of $P(Y \leq x_C)$ and $P(X \geq x_C)$.

The purpose of the examples given in Figure 4.6 and Table 4.1 is not to provide a comprehensive treatment of all the cases in which the PSR may be observed and calculated when *X* and *Y* are distributed as skew normal. Their purpose is two-fold. First, to express the wide array of possibilities that exist for expressing the overlap and hence the PSR when *X* and *Y* are

not constrained to the normal distribution. Even under the extreme constraints of these four cases considered, the nature of the PSR changes greatly. Second, the purpose is to provide a general strategy for calculating the PSR and an illustration of the values required in order to evaluate the PSR. Although each situation needs to be considered individually with respect to the relationship between and among the location, scale, and shape parameters, the strategy of calculating the crossing points and expressing the PSR in terms of combining mutually exclusive areas allows for both proper expression and calculation of the PSR. The next section discusses an alternate strategy for computing the PSR.

4.4.2. The Kernel PSR

The kernel PSR proposed by Stine and Heyse (2001), denoted herein as PSR_k was introduced in (2.13) as

$$\text{PSR}_k = \int min\left(\hat{f}_{X_k}(x), \hat{f}_{Y_k}(x)\right) dx,$$

where $\hat{f}_{X_k}(x)$ and $\hat{f}_{Y_k}(x)$ are given in (2.12). The primary advantage of the PSR_k is the ability to generate a non-parametric estimate of the PSR when the distribution of either *X* or *Y* or both is not normally distributed, which is ideal for use under the skew normal model used in the current Chapter. Furthermore, in the context the skew normal model, the PSR_k offers an alternative to the PSR_N when *X* and *Y* are skew normal and the crossing points need to be estimated numerically. Recall the example used in Figure 4.3 provides a case where the crossing point had to be solved for numerically using the algorithm provided in appendix B.4.2. In Figure 4.3 $\varepsilon = 0.01$ was used, which resulted in a crossing point estimate of $x_C \cong 4.65$ and a PSR estimate of $\text{PSR}_\text{N} \cong 0.514$. For comparison purposes Figure 4.7 was generated from $n_X = n_Y = 500$ generated values under the same skew normal specifications as Figure 4.3 $\left(i.e.\ \mu_X = 4, \mu_Y =\right.$

$0, \sigma_X = \sigma_Y = 5,$ and $\alpha_{X_M} = \alpha_{Y_M} = 3\big)$. The $\text{PSR}_k \cong 0.513$ displayed as the shaded area in Figure 4.7 is very close to the $\text{PSR}_N \cong 0.514$. The next section uses a small simulation study to formulate the important features of the kernel PSR.

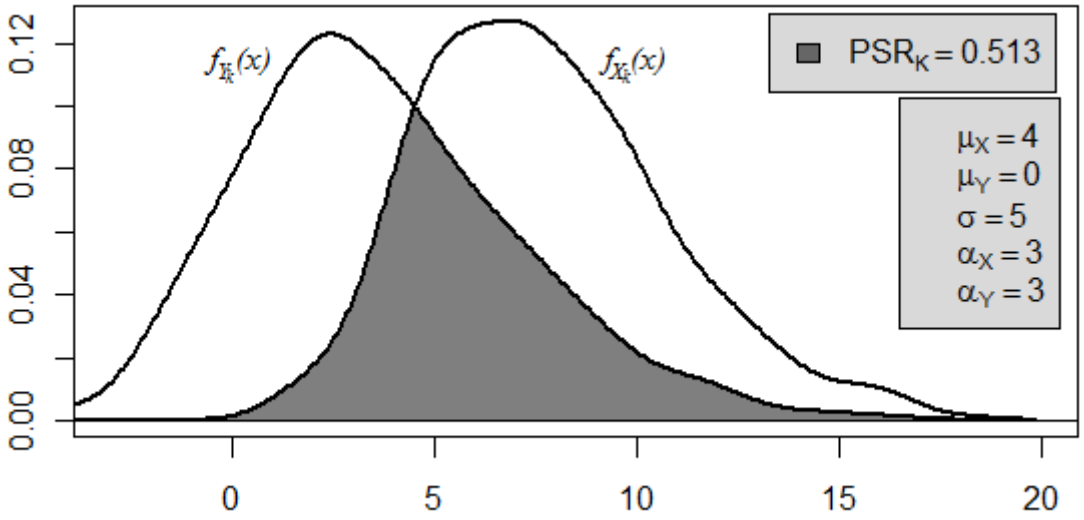


Figure 4.7: Kernel PSR
$f_{X_K}(x)$ and $f_{Y_K}(x)$ are kernel density estimates of $X \sim SN(\mu_X = 4, \sigma_X = 5, \alpha_X = 3)$ and $Y \sim SN(\mu_Y = 0, \sigma_Y = 5, \alpha_Y = 3)$, respectively, from $n = 500$ randomly generated values from each distribution. The shaded area is the kernel estimated PSR equal to 0.513. The same distribuions were used to generate Figure 3 in which PSR = 0.514. The equations for $f_{X_K}(x)$ and $f_{Y_K}(x)$ are given in (2.12), and the bandwidth is a Sheather-Jones plug in estimate.

4.4.3. Simulation Results for the Kernel PSR

In an effort to maximize the use of the PSR_K as an estimate of the PSR, this section discusses a small simulation directed at studying the effects of both the bandwidth choice and sample size effect on the efficiency of the PSR_K. In most cases the kernel component of the kernel density estimators makes very little difference (see Sheather, 2004). Simulations by Stine and Heyse (2001) for the PSR_k suggest this result holds for $\hat{f}_{X_K}(u)$ and $\hat{f}_{Y_K}(v)$ as well. Thus the standard normal kernel $K_x = K_y = \phi(\cdot)$ is used throughout. The bandwidth employed by Stine and Heyse is a scalar of the normal bandwidth method introduced by Silverman (1986), which is

the optimal bandwidth choice if a distribution is normal based on the mean integrated squared error criteria. Since Stine and Heyse investigated the effectiveness of PSR_K exclusively with normal data, the normal bandwidth was a reasonable choice. However, non-normal data is assumed in this section so other bandwidths should be considered. The investigation was conducted via simulation, with the results reported in Table 4.2. The first column of Table 4.2 indicates the two bandwidths compared. The first bandwidth is the Sheather-Jones (SJ) plug-in method with its results given across the first row. The SJ method was chosen for its robust property effective across many distributions (see Sheather, 2004). Secondly, the normal bandwidth method utilized by Stine and Heyse, with its results given across the second row, is shown.

To measure the effectiveness of the PSR_K, it was compared to the PSR_N where the crossing point was calculated using $\varepsilon = .001$. The reference distributions for X and Y are those of Figures 4.3 and 4.7 $\left(i.e.\ \mu_X = 4, \mu_Y = 0, \sigma_X = \sigma_Y = 5, \text{and } \alpha_{X_M} = \alpha_{Y_M} = 3\right)$. 1,000 samples generated 1,000 $\mathrm{PSR}_k's$ from these distributions for each of the sample sizes 50, 100, 250, 500, 1000, and 2000. Then the average generated by $\overline{\mathrm{PSR}}_K = \sum_{i=1}^{1000} \frac{\mathrm{PSR}_{k_i}}{1000}$ was compared to the $\mathrm{PSR}_\mathrm{N} = 0.5143$ computed with $\varepsilon = 0.001$ accuracy. Each cell in the body of Table 4.2 reports the absolute difference of the PSR estimates given as $|\overline{\mathrm{PSR}}_K - 0.5143|$. Thus for a sample size of $n = 50$, the reported value of 0.0579 represents a difference of nearly 6%. With $n = 500$ the difference drops to 0.0229 and drops less than 1% for each of the successive sample size increases at $n = 1000$ and $n = 2000$. Therefore, the PSR_K appears to be very stable for sample sizes of 500 or more. The last row of Table 4.2 shows the effectiveness of using the normal bandwidth estimate used by Stine and Heyse as opposed to the SJ bandwidth reported in row one. Although the Kernel PSR estimates are very close, not surprisingly, the differences for the

SJ bandwidth is consistently less than for the normal bandwidth. For these reasons, the PSR_K based on $n = 500$ and the SJ bandwidth has been used generate the PSR_K used in this work. For convenience, an R program is included in appendix B for generating PSR_K (see B.4.3).

Table 4.2: Simulation check on bandwidth

The summary results of a simulation designed to evaluate the effect of both sample size (n) and choice of bandwidth on the kernel PSR are given in this table. The skew normal densities $f_X(x)$ and $f_Y(x)$, esimated by $f_{X_K}(x)$ and $f_{Y_K}(x)$, respectively, are parameterized as in Figures 4.3 and 4.7, which gave $\text{PSR}_\text{N} = 0.514$. Estimates for the PSR_K were generated from an average of 1000 simulations denoted by $\overline{\text{PSR}}_K$ and the values given in the body of the table are $|\overline{\text{PSR}}_K - 0.5143|$ for each sample size and bandwidth combination.

Bandwidth	**Simulation Size (n)**					
	50	**100**	**250**	**500**	**1000**	**2000**
Sheather Jones	0.0579	0.0462	0.0308	0.0229	0.0173	0.0131
Normal	0.0603	0.0490	0.0337	0.0257	0.0198	0.0154

4.4.4. Skew Normal Marginal Distributions

Result 4.1 states that the marginal distributions of X and Y that come from $\boldsymbol{Z} = (X \quad Y)^T$ given in (4.6) are not independent even when $\rho_{XY} = 0$. This section helps identify part of the reason for this dependence. More importantly, this section develops the marginal shape parameters α_{X_M} and α_{Y_M} vital to the calculation of the PSR. Equations $(4.7a - f)$, given above and used to construct the distribution of D, provide the basis for construction of the distributions of both X and Y by simply taking A^T equal to $(1 \quad 0)^T$ and $(0 \quad 1)^T$, respectively. Application of $(4.7a - f)$ shows that the marginal distributions of both X and Y are skew normal. Furthermore, the marginal location and scale parameters for both X and Y are the same for those of the bivariate distribution. However, the marginal shape parameters are given by

$$\alpha_{X_M} = \frac{\alpha_{X_J} + \alpha_{Y_J}\rho_{XY}}{\sqrt{1 + \alpha_{Y_J}^2(1 - \rho_{XY}^2)}} \tag{4.11a}$$

and

$$\alpha_{Y_M} = \frac{\alpha_{X_J}\rho_{XY} + \alpha_{Y_J}}{\sqrt{1 + \alpha_{X_J}^2(1 - \rho_{XY}^2)}} \tag{4.11b}$$

as shown in the appendix (see A.4.5). Both α_{X_M} and α_{Y_M} illustrate the interdependence of the bivariate skew normal parameters $\alpha_{X_J}, \alpha_{Y_J},$ and ρ_{XY}. Note that even when $\rho_{XY} = 0$ the marginal shape parameters equal

$$\alpha_{X_M} = \frac{\alpha_{X_J}}{\sqrt{1 + \alpha_{Y_J}^2}} \quad \text{and} \quad \alpha_{Y_M} = \frac{\alpha_{Y_J}}{\sqrt{1 + \alpha_{X_J}^2}};$$

identifying the cause of dependency between the marginal distributions of X and Y highlighted in result 4.1. That is both α_{X_M} and α_{Y_M}remain functions of α_{X_J} and α_{Y_J}even when $\rho_{XY} = 0$. Given this dependency, it is instructive to study the effect of the bivariate skew normal parameters on the marginal shape parameters α_{X_M}, α_{Y_M}, and α_D. Figure 4.8 displays α_{X_M}, α_{Y_M}, and α_D as the dashed, dotted, and solid line, respectively as functions of α_{X_J}, α_{Y_J}, ρ_{XY}, and k for some general cases. Panel (a) shows that under the constraint that $k = 1$ and $\alpha_{X_J} < \alpha_{Y_J}$, both α_{X_M} and α_{Y_M} are strictly increasing functions of ρ_{XY}. Furthermore, as ρ_{XY} goes to 1 α_{X_M} and α_{Y_M} converge. Finally, if $\alpha_{X_J} < \alpha_{Y_J}$, then $\alpha_{X_M} \leq \alpha_{Y_M}$ for all ρ_{XY}; a fact that can be seen from application of $(4.11a)$ and $(4.11b)$.

In contrast to panel (a) of Figure 4.8, ρ_{XY} is fixed ($\rho_{XY} = 0.5$) in panel (b) and k is allowed to change over $1 \leq k \leq 5$. The zero slopes of α_{X_M} and α_{Y_M}in panel (b) confirms what is clear from $(4.11a)$ and $(4.11b)$, which is that the marginal shape parameters are not a

function the joint scale parameters. α_D is a function of k as made clear from either panel (b) or inspection of (4.8). The only change in panel (a) to panel (c) is now $\alpha_{X_J} > \alpha_{Y_J}$, making $\alpha_D > 0$. Thus $\alpha_{X_M} \geq \alpha_{Y_M}$ for all ρ_{XY}, but α_{X_M} and α_{Y_M} still converge as ρ_{XY} goes to 1. Panel (d) shows the same effect as panel (b) even though $\alpha_{X_J} > \alpha_{Y_J}$. Ultimately, the important characteristic resulting from the marginal distributions is the PSR, which leads to the next section.

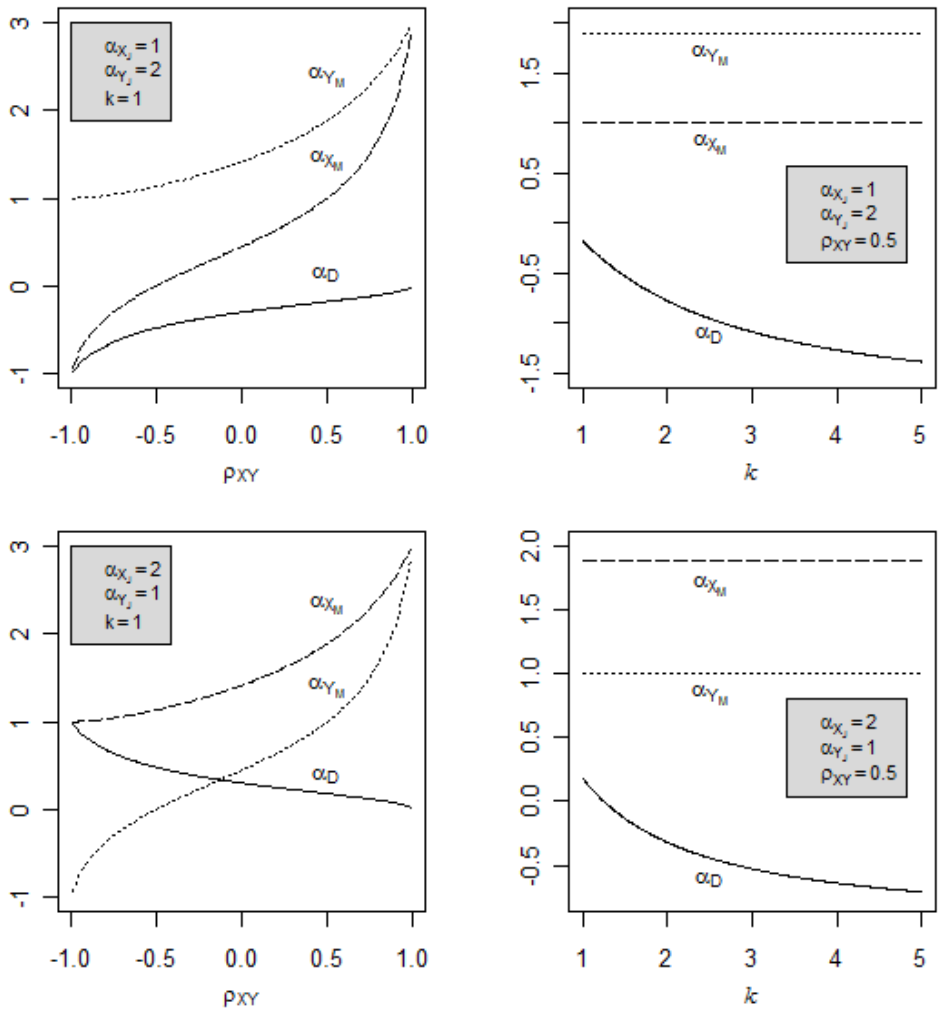


Figure 4.8: Assessment of shape parameters
This figure illustrates the dependency of the shape parameters α_D, α_{X_M}, and α_{Y_M} on ρ_{XY} and k given the distributions of *D*, *X*, and *Y* are skew normal. Each shape parameter is derived from the bivariate skew normal distribution of X and Y using equations (4.8), $(4.11a)$, and $(4.11b)$, respectively. For panels (a) and (b) $\alpha_{X_J} = 1$ and $\alpha_{Y_J} = 2$ and for panels (c) and (d) $\alpha_{X_J} = 2$ and $\alpha_{Y_J} = 1$. Panels (a) and (b) illustrate α_D's dependency on α_{X_J}, α_{Y_J}, k, and ρ_{XY} , while the marginal shape parameters α_{X_M} and α_{X_M} depend only on α_{X_J}, α_{Y_J}, and ρ_{XY}. Furthermore, α_D is a strictly increasing function of ρ_{XY} when $\alpha_D < 0$ and a strictly decreasing function of ρ_{XY} when $\alpha_D > 0$.

4.4.5. The PSR Calculation from Marginal Distribution

The PSR is a function of the marginal distributions of *X* and *Y*. An important result about the bivariate normal distribution of $(X \quad Y)^T$ is that the marginal distributions of *X* and *Y* are not a function of ρ_{XY}. Consequently, the PSR obtained from $f_X(u)$ and $f_Y(v)$ is not a function of ρ_{XY}. This result is illustrated graphically in Figure 4.9 panel (a), where, since $\alpha_{X_J} = \alpha_{Y_J} = 0$ of a skew normal distribution, $(X \quad Y)^T$ is distributed as bivariate normal with $\mu_X = 2,\ \mu_Y = 0,$ and $\sigma_X = \sigma_Y = \sigma = 1$. ½ PSR is given by both the ½PSR_k, displayed as the dotted line, and the ½PSR_N, displayed as the solid line. Each quantity graphically illustrates the PSR's independence of ρ_{XY} since they are constant equal to (2.11) over ρ_{XY}. Note that the PSR_K shows random fluxuations due to the simulation size $(n = 500)$. Conversely, the $\text{PIQI}^{\rho_{XY}}$, displayed as the dashed line, decreases over $-1 \leq \rho_{XY} \leq 1$ since the $\text{PIQI}^{\rho_{XY}}$ is a function of ρ_{XY} through σ_D^2.

When $\alpha_J^T = (\alpha_{X_J} \quad \alpha_{Y_J})^T \neq (0 \quad 0)^T$, the effect of α_{X_M} and/or α_{Y_M} being a function of ρ_{XY} is illustrated in Figure 4.9 panel (b), where $(\alpha_{X_J} \quad \alpha_{Y_J})^T = (1 \quad 3)^T$. Note that as ρ_{XY} increases, ½PSR decreases as demonstrated by both $½PSR_N$ and $½PSR_K$ decrease in a similar fashion as the $\text{PIQI}^{\rho_{XY}}$ shown with the dashed line. As illustrated in Figure 4.5 panel (b), the PIQI is not always a decreasing function of ρ_{XY} when X and Y are skew normal. Such is the case with the PSR. However, the important point is that the PSR is a function of ρ_{XY} when *X* and *Y* are skew normal in a similar way that the PIQI is a function of ρ_{XY}. Panels (c) and (d) of Figure 4.9 provide a graphical illustration of the PSR's dependency on ρ_{XY} from a density 'overlap' perspective, where the bivariate distribution of $(X \quad Y)^T$ is distributed as

$$SN_2 \sim \left(\begin{bmatrix} \mu_X = 2 \\ \mu_Y = 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{bmatrix}, \begin{bmatrix} \alpha_{X_J} = 1 \\ \alpha_{Y_J} = 3 \end{bmatrix} \right)$$

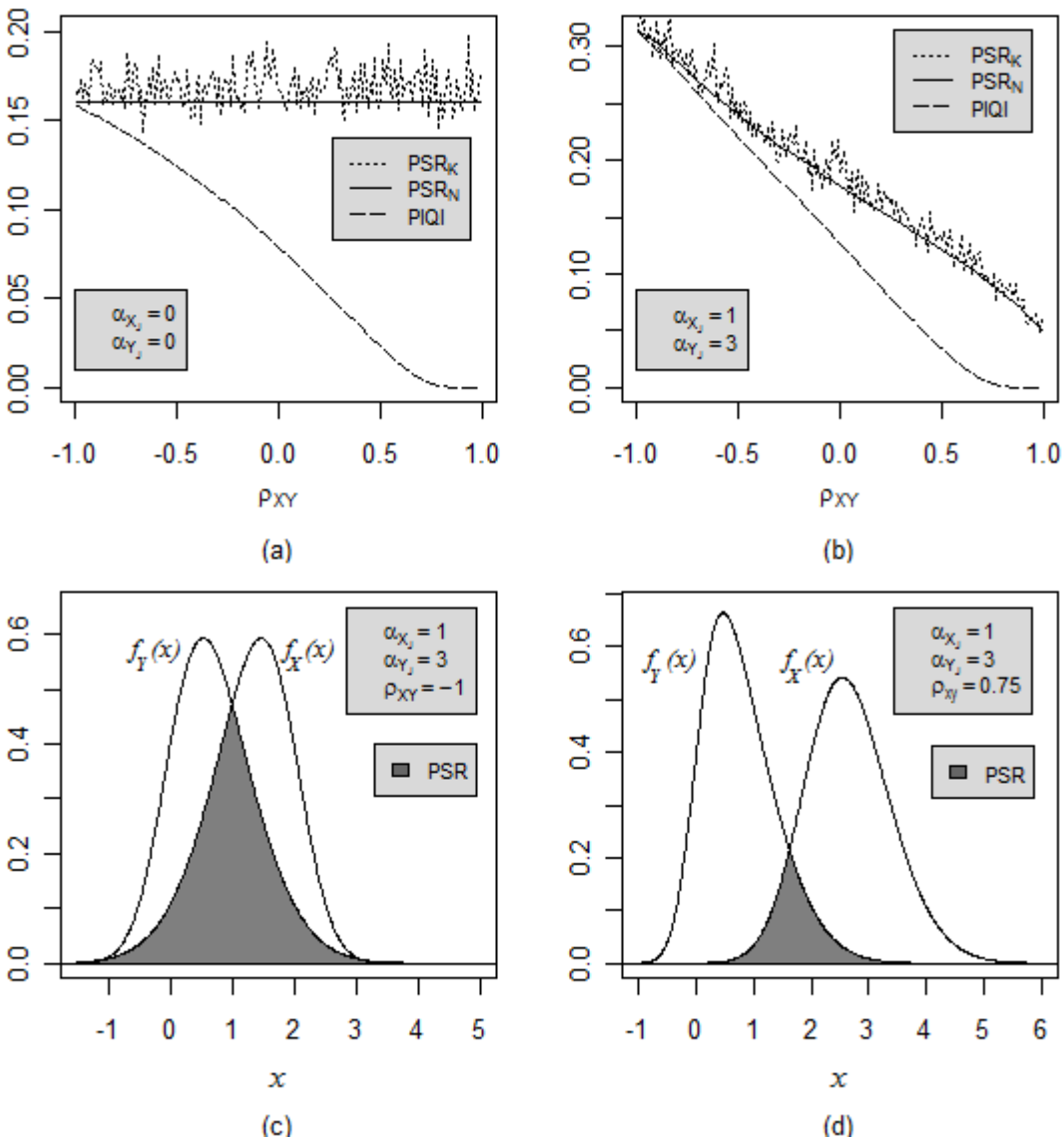


Figure 4.9: An illustration of the PSR's dependency on the correlation
This figure illustrates the PSR's dependency on ρ_{XY} when $(X \quad Y)^T$ is bivariate skew normal. In this illustration the location parameters are set at $\mu_X = 2$ and $\mu_Y = 0$, and the scale parameters are set at $\sigma_X = \sigma_Y = \sigma = 1$ across each panel. In panels (a) and (b) ½ PSR is measured by both the ½PSR_N shown as the solid line and the ½PSR_K using $n = 500$ shown as the dotted line. For reference, the PIQI is illustrated with the dashed line. Panels (a) and (b) show that as $(X \quad Y)^T$ transitions from bivariate normal to bivariate skew normal signified by $(\alpha_{X_J} \quad \alpha_{Y_J})^T$ transitioning from $(0 \quad 0)^T$ to $(1 \quad 3)^T$. Thus the PSR transitions from not being a function of ρ_{XY} in (a) to becoming a strictly decreasing function of ρ_{XY} in (b). The graphical display of panel (b) as depicted in the overlap of $f_X(x)$ and $f_Y(x)$ is illustrated in the transition from panel (c) to panel (d). The shaded area representing both the overlap and ½PSR is significantly reduced as ρ_{XY} changes from -1 in panel (c) to 0.75 in panel (d) in which ½PSR is reduced from 0.315 to 0.090, respectively.

so that $\sigma_X = \sigma_Y = 1$ $(i.e.$ the same distribution as panel (b)). In panel (c) $\rho_{XY} = -1$ and PSR illustrated by the shaded area is approximately equal to 0.629. The marginal shape parameters are equal to $\alpha_{X_M} = 2$ and $\alpha_{Y_M} = -2$. The transition to panel (d) is caused by changing $\rho_{XY} = 0.75$. The increase in ρ_{XY} produces $\alpha_{X_M} \cong 1.46$ and $\alpha_{Y_M} \cong 3.13$, which ultimately reduced the PSR to approximately 0.18 illustrated by the reduction in the shaded area in panel (d). Therefore, similar to superscripting the PIQI with ρ_{XY} to emphasize this dependency, the PSR may need a superscript to emphasize the functional relationship between the PSR and ρ_{XY} when $(X \quad Y)^T$ is bivariate skew normal. Furthermore, for the remainder of this work the PSR will be calculated numerically using algorithm B.4.2 in appendix B and denoted as PSR_N. So dropping the subscript 'N' and adding the superscript ρ_{XY} to the PSR gives $\text{PSR}^{\rho_{XY}}$, which is taken to mean that the PSR is a function of ρ_{XY} and has been calculated by solving for the crossing points numerically. The next section establishes what connections can be made between the PIQI and the PSR under the bivariate skew normal distribution.

4.5. Assessing the PIQI/PSR Relationship under Skew Normality

The relationship between the PSR and the PIQI is taken from the perspective that X and Y come from a bivariate distribution. This was the situation in Chapter 3 when $(X \quad Y)^T$ was distributed as bivariate normal, where, although ρ_{XY} did not impact the PSR, ρ_{XY} played an important role in the PIQI. As established in Section 4.3, given the bivariate skew normal model the PSR is a function of ρ_{XY} as well, so that under this model proposition 2.2, which states that $\text{PIQI}^{max} \geq \frac{1}{2}\text{PSR}^{\rho_{XY}}$, is not necessarily true. Thus there is no guarantee that $\text{PIQI}^{max} \geq \frac{1}{2}\text{PSR}^{\rho_{XY}}$ is true outside the bivariate normal model. However, the main objective of this section is to assess the conditions under which $\text{PIQI}^{max} \geq \frac{1}{2}\text{PSR}^{\rho_{XY}}$ is true, given $(X \quad Y)^T$ is

distributed as bivariate skew normal. The determining factor in establishing whether $\text{PIQI}^{max} \geq$ ½$\text{PSR}^{\rho_{XY}}$ holds is the state of α_D. These states are $\alpha_D > 0$, $\alpha_D = 0$, and $\alpha_D < 0$. The next section presents four cases; each representing specific parameter combinations for a bivariate skew normal distribution. The first case is an example of $\alpha_D > 0$. The second case is an example of $\alpha_D = 0$, and the last two cases are examples of $\alpha_D < 0$. From these cases, some generalizations about the relationship between the PIQI and PSR over the different states of $\alpha_D$can be made.

4.5.1. Cases for Study

Due to the several potential shapes for both the bivariate and marginal skew normal densities, there are many classes under which the relationship between the PIQI and the PSR may be studied, which make it difficult to form connections between the PSR and the PIQI that hold over all classes. The strategy used here is to form sub-classes defined by the state of α_D, the value of k, and the value of ρ_{XY} over the entire class of skew normal distributions so that some generalizations may be formed within the sub-classes. Table 4.3 delineates four cases; each representative of a state of α_D.

The first three columns represent the sub-classes of the skew normal, which are defined over the state of α_D. A particular case number (#) is assigned in column 1, and the PSR computations are given in column three, but only under the restriction that $\rho_{XY} = -1$. The remaining columns provide specific parameterizations that fall under the sub-classes. The exact values of the parameters are not important except that they generate the state of α_D. For example, in case #1 since $\alpha_{X_J} = 2$ and $\alpha_{Y_J} = 1$, from (4.10) we have that $\alpha_D > 0$ over $-1 <$

$\rho_{XY} < 1$ and $k = 1$. In fact, given $\rho_{XY} = -1$ we have that $\alpha_D = 1$, $\alpha_{X_M} = 1$, $\alpha_{Y_M} = -1$, and $x_C = 1$.

Since the values of PIQI and PSR are both functions of k and ρ_{XY}, their relationship is subject to change as ρ_{XY} and k change. Therefore, the PIQI/PSR connections need to be ascertained under different values of each. The important sets of ρ_{XY} are $\rho_{XY} = -1$ and $\rho_{XY} > -1$. The important sets of k are $k = 1$ and $k > 1$. The following sections are dedicated to establishing connections between the PIQI and the PSR under each state of α_D and based on the important sets of ρ_{XY} and k. Furthermore the constraints that $\mu_X \geq \mu_Y$ and $E(X) \geq E(Y)$ are in force throughout.

Table 4.3: Outlining special cases
This table outlines four cases for the relationship between the PIQI and the PSR under the skew normal model. Each case is differentiated by the relationship between the joint shape parameters α_{X_J} and α_{Y_J} as summarized by α_D in column 2 and calculated in (4.10) since $k = 1$. The characterization of the PSR along with the marginal shape parameters and the crossing points are all given under the constraint $\rho_{XY} = -1$. Other parameter specifications include $\mu_X = 2$ and $\mu_Y = 0$.

Sub Classes of the Skew Normal			Parameter Specification							
#	$\alpha_D^{k=1}$	$PSR^{\rho_{XY}=-1}$	α_{X_J}	α_{Y_J}	α_{X_M}	α_{Y_M}	α_D	x_L	$x_{I/C}$	x_U
1	$\alpha_D > 0$	$P(X \leq x_C) + P(Y \geq x_C)$	2	1	1	-1	1	na	1	na
2	$\alpha_D = 0$	$P(X \leq x_C) + P(Y \geq x_C)$	3	3	0	0	0	na	1	na
3	$\alpha_D < 0$	$P(Y \leq x_L) + P(x_L \leq X \leq x_I) + P(x_I \leq Y \leq x_U) + (X \geq x_U)$	1	5	-4	4	-4	-0.4	1	2.4
4	$\alpha_D < 0$	$P(Y \leq x_L) + P(x_L \leq X \leq x_I) + P(x_I \leq Y \leq x_U) + (X \geq x_U)$	-2	2	-4	4	-4	-0.4	1	2.4

4.5.2. Case Study Results over the Correlation

In order to ascertain the relationship between the PIQI and the PSR over ρ_{XY}, in this section the $k = 1$ constraint will be enforced over the sets of $\rho_{XY} = -1$ and $\rho_{XY} > -1$. Also, similar to the location parameters of the normal model, each set of ρ_{XY} state of α_D combination will be evaluated.

4.5.2.1. Case Study Results for $\rho_{XY} = -1$

This section begins with an important proposition. This proposition is motivated by results 3.1, 3.2, and 3.3; foundational for establishing the relationship between the PIQI and the PSR when $(X \quad Y)^T$ was bivariate normal. Recall that results 3.1 and 3.2 establish the fact that when $\rho_{XY} = -1$, there exists a common value for *x* and *y* given by $x_{-1} = \frac{\mu_Y + k\mu_X}{1+k}$. These results were not distribution dependent, so they hold for any bivariate distribution. Result 3.3 showed that $P(D < 0)_{\rho_{XY}=-1} = P(X < x_{-1}) = P(Y > x_{-1})$, but under the constraint that $(X \quad Y)^T$ was bivariate normal. The next proposition does not depend on the bivariate distribution of $(X \quad Y)^T$ nor on k.

❖ **Proposition 4.3**

Given $\mu_X > \mu_Y$ and $E(X) > E(Y)$

$$\text{PIQI}^{\rho_{XY}=-1} = P(D < 0)_{\rho_{XY}=-1} = P(X \leq x_{-1}) = P(Y \geq x_{-1})$$

Proof:

$$P(X \leq x_{-1}, Y \geq x_{-1}) = P(X \leq x_{-1} | Y \geq x_{-1}) \times P(Y \geq x_{-1}) = 1 \times P(Y \geq x_{-1})$$
$$= P(Y \geq x_{-1})$$
$$P(X \leq x_{-1}, Y \geq x_{-1}) = P(Y \geq x_{-1} | X \leq x_{-1}) \times P(X \leq x_{-1}) = 1 \times P(X \leq x_{-1})$$
$$= P(X \leq x_{-1})$$
$$\text{PIQI}^{\rho_{XY}=-1} = P(D < 0)_{\rho_{XY}=-1} = P(X - Y < 0)_{\rho_{XY}=-1} = P(X < Y)_{\rho_{XY}=-1}$$
$$= P(X \leq x_{-1})$$

A proof for result 3.3 was given in Chapter 3. However, as an alternative proof, result 3.3 is also true due to proposition 4.3 and given the bivariate normal distribution $\text{PIQI}^{\rho_{XY}=-1} = \text{PIQI}^{max}$. Recall from Section 4.2 that this fact is not necessarily true for the bivariate skew normal distribution. A boundary for $½\text{PSR}^{\rho_{XY}}$ at $\rho_{XY} = -1$ over each state of α_D is given next.

State: $\boldsymbol{\alpha_D > 0}$ ***(example: case 1)***

Given (4.1), $\rho_{XY} = -1$, and $k = 1$

➔**Boundary**: $½\text{PSR}^{\rho_{XY}=-1} = \text{PIQI}^{\rho_{XY}=-1}$

Proof:

$$\text{PSR}^{\rho_{XY}=-1} = P(X \leq x_C) + P(Y \geq x_C)$$
$$= P(X \leq x_{-1}) + P(Y \geq x_{-1})$$

since $x_C = x_{-1}$, so

$$= P(D \leq 0)_{\rho_{XY}=-1} + P(D \leq 0)_{\rho_{XY}=-1}$$
$$= 2 \times \text{PIQI}^{\rho_{XY}=-1}$$
$$½\text{PSR}^{\rho_{XY}=-1} = \text{PIQI}^{\rho_{XY}=-1} \quad \blacksquare$$

State: $\boldsymbol{\alpha_D = 0}$ ***(example: case 2)***

Given (4.1), $\rho_{XY} = -1$, and $k = 1$

➔**Boundary**: $½\text{PSR}^{\rho_{XY}=-1} = \text{PIQI}^{max}$

Proof:

Since $k = 1$ and $\rho_{XY} = -1$, from (4.10) $\alpha_D = 0$, and from $(4.11a, b)$ α_{X_M} and α_{Y_M} are equal to 0. Therefore, *D*, *X*, and *Y* are normally distributed. Consequently, by proposition 3.2 $½\text{PSR}^{\rho_{XY}=-1} = \text{PIQI}^{\rho_{XY}=-1} = \text{PIQI}^{max}$. ■

State: $\boldsymbol{\alpha_D < 0}$ ***(examples: cases 3 and 4)***

Given (4.1), $\rho_{XY} = -1$, and $k = 1$

➔**Boundary**: $½\text{PSR}^{\rho_{XY}=-1} < PIQI^{max}$

Proof:

$$\begin{aligned}
\text{PSR}^{\rho_{XY}=-1} &= P(Y \le x_L) + P(x_L \le X \le x_C) + P(x_C \le Y \le x_R) + P(X \ge x_R) \\
&= P(Y \le x_L) + P(x_L \le X \le x_{-1}) + P(x_{-1} \le Y \le x_R) + P(X \ge x_R) \\
&< P(X \le x_L) + P(x_L \le X \le x_{-1}) + P(x_{-1} \le Y \le x_R) + P(Y \ge x_R) \\
&= P(X \le x_{-1}) + P(Y \ge x_{-1}) \\
&= 2 \times \text{PIQI}^{\rho_{XY}=-1} \\
&= 2 \times \text{PIQI}^{max}
\end{aligned}$$

$\frac{1}{2}\text{PSR}^{\rho_{XY}=-1} < \text{PIQI}^{max}$. ■

Consequently, for the class of bivariate skew normal distributions defined by $\rho_{XY} = -1$ and $k = 1$, the $\text{PIQI}^{\rho_{XY}=-1} \ge \frac{1}{2}\text{PSR}^{\rho_{XY}=-1}$ boundary holds over all states of α_D. A graphical illustration of each case is displayed in Figure 4.10. The PIQI^{max} represented with the dashed line is equal to or larger than ½PSR represented with the solid line over at the value $\rho_{XY} = -1$. The next section compares the PIQI^{max} and ½ PSR over $\rho_{XY} > -1$, where it will be shown that the boundary does depend on the state of α_D.

4.5.2.2. Case Studies for $\rho_{XY} \ge -1$

State: $\alpha_D > 0$ (example: case 1)

Given (4.1), $\rho_{XY} \ge -1$, and $k = 1$

$$\rightarrow \text{PIQI}^{max} \not\ge \tfrac{1}{2}\text{PSR}^{\rho_{XY}}$$

Thus the PIQI^{max} does not provide a boundary for $\frac{1}{2}\text{PSR}^{\rho_{XY}}$ when $\alpha_D > 0$. The illustration for case 1 in Figure 4.10 shows that the $\text{PIQI}^{max} \not\ge \frac{1}{2}\text{PSR}^{\rho_{XY}}$, since the $\text{PIQI}^{max} = \text{PIQI}^{\rho_{XY}=-0.25} \cong 0.17$ is less than $\frac{1}{2}\text{PSR}^{\rho_{XY}=0.5} \cong 0.224$. The last two columns of Table 4.4, give the values for ½ PSR and the PIQI over seven ascending values of ρ_{XY} to complement Figure 4.10. Consequently, case 1 serves as proof that the PIQI^{max} boundary given in proposition 2.2 does not hold under the bivariate skew normal model. In fact, case 1 is a

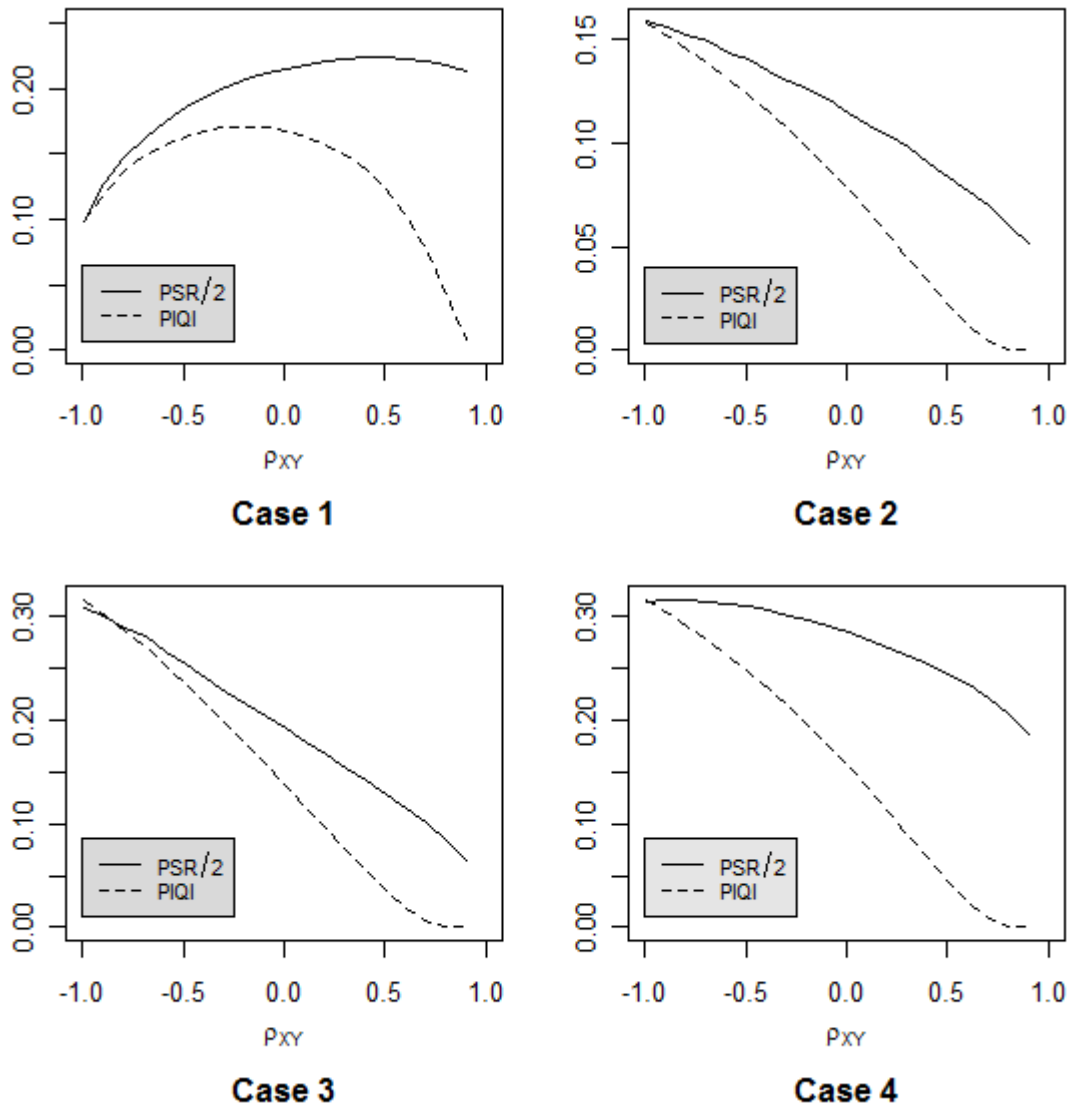


Figure 4.10: A study of the PSR and the PIQI under the special cases
Each of the panels shows ½PSR as the solid line and the PIQI as the dashed line. The cases are those given in Table 4.3. When $\alpha_D = \alpha_{X_M} - \alpha_{Y_M} > 0$, as in case 1, both the ½ PSR^{max} and the PIQI^{max} are achieved at $\rho_{XY} > -1$. However, when $\alpha_D = \alpha_{X_M} - \alpha_{Y_M} \leq 0$, both are achieved at $\rho_{XY} = -1$.

working example in which PIQI$^{max} \geq$ ½PSR$^{\rho_{XY}}$ does not hold for the class of bivariate skew normal distributions defined by $\rho_{XY} > -1, \alpha_D > 0$ and $k = 1$. The reason for this boundary failure under this class of distributions is revealed by a close inspection of $(4.11a, b)$. When

$\alpha_D > 0$ $\left(i.e.\ \alpha_{X_J} > \alpha_{Y_J}\right)$ and $\alpha_{X_J}, \alpha_{Y_J} > 0$, both α_{X_M} and α_{Y_M} are strictly increasing functions of ρ_{XY} (see appendix A.4.6 for proof). Furthermore, as shown in Table 4.4, α_{Y_M} is increasing faster than α_{X_M}. Figure 4.8 panel (c) complements Table 4.4 illustrating graphically that α_{Y_M} is increasing at a faster rate than α_{X_M}. Thus $f_Y(x)$ is shifting to the right at a faster rate than $f_X(x)$ is shifting to the right. The effect here on the marginal distributions of *X* and *Y* are similar to the effect on the distribution of *D* when $\alpha_D > 0$, which causes the PIQI^{max} to occur at $\rho_{XY} > -1$. That is, the maximum PSR occurs at $\rho_{XY} > -1$. Table 4.4 charts the change in the marginal shape parameters in an effort to delineate the faster shift of *Y* over *X* to supplement the illustration in panel (c) of Figure 4.8. Note in particular that although $\alpha_{X_M} \geq \alpha_{Y_M}$ over $\rho_{XY} \geq -1$, both α_{X_M} and α_{Y_M} are converging on 3, while their difference $\alpha_{X_M} - \alpha_{Y_M}$ is going to zero.

Table 4.4: Shows the dependency of the PSR on the correlation
This table complements Figure 4.10 case 1 by illustrating the cause of $PSR^{\rho_{XY}}$ being an increasing function ρ_{XY} over $(-1, 0.5)$. Both α_{X_M} and α_{Y_M} are increasing functions of ρ_{XY} (illustrated in Figure 4.8 panel (c) and proven in Appendix A.4.6), but α_{Y_M} is increasing faster demonstrated by $\alpha_{X_M} - \alpha_{Y_M}$ going to zero. Thus $f_Y(x)$ is shifting to the right faster than $f_X(x)$ causing an increase in both the 'overlap' and ½PSR$^{\rho_{XY}}$. For reference the PIQI$^{\rho_{XY}}$ is affected similarly by $\alpha_{X_M} > \alpha_{Y_M}$ so that $PIQI^{max}$ is achieved at $\rho_{XY} > -1$.

Specification			**Results**					
$\boldsymbol{\rho_{XY}}$	$\boldsymbol{\alpha_{X_J}}$	$\boldsymbol{\alpha_{Y_J}}$	$\boldsymbol{\alpha_{X_M}}$	$\boldsymbol{\alpha_{Y_M}}$	$\boldsymbol{\alpha_{X_M} - \alpha_{Y_M}}$	$\boldsymbol{x_{I/C}}$	**½PSR$^{\rho_{XY}}$**	**PIQI$^{\rho_{XY}}$**
-1	2	1	1	-1	2	1	0.095	0.095
-0.5	2	1	1.13	0	1.13	1.5	0.185	0.163
-0.25	2	1	1.26	0.23	1.03	1.65	0.203	0.170
0	2	1	1.41	0.45	0.96	1.79	0.215	0.168
0.25	2	1	1.62	0.69	0.93	1.91	0.221	0.154
0.5	2	1	1.89	1	0.89	2	0.224	0.125
1	2	1	3	3	0	2.11	0.209	0

State:* $\alpha_D \leq 0$ *(example: cases 2, 3, and 4)

Given (4.1), $\rho_{XY} \geq -1$, $k = 1$, and $\alpha_{X_J} = \alpha_{X_J} = 3$

➔**Case specific Boundary**: $\text{PIQI}^{max} \geq ½\text{PSR}^{\rho_{XY}}$

Note that this result is only applicable to this specific case and not the class of bivariate skew normal distributions over the entire class defined by $\alpha_D \leq 0$. The proof of this boundary comes from the fact that $\text{PSR}^{\rho_{XY}}$ is a strictly decreasing function of ρ_{XY}, which can be verified from direct inspection of Figure 4.10 panels (b), (c), and (d). Thus for any specific case in which $\text{PSR}^{\rho_{XY}}$ can be shown to be a strictly decreasing function of ρ_{XY}, the $\text{PIQI}^{max} \geq ½\text{PSR}^{\rho_{XY}}$ boundary holds. The development of a more general boundary is discussed in Section 4.7. The challenge of establishing $\text{PSR}^{\rho_{XY}}$ as a strictly decreasing function of ρ_{XY} over the entire class of distributions defined by (4.1), $k = 1$, and $\alpha_D \leq 0$ can be seen by inspection of the $\text{PSR}^{\rho_{XY}}$ formula given in Section 4.4, which was given as

$$\text{PSR} = \int_{-\infty}^{\infty} min\big(f_X(u), f_Y(v)\big)dx = P(X \leq x_C) + P(Y \geq x_C)$$

$$= \int_{-\infty}^{x_C} \frac{2}{\sigma_X}\phi\left(\frac{u-\mu_X}{\sigma_X}\right)\Phi\left(\alpha_X\frac{u-\mu_X}{\sigma_X}\right)du + 1 - \int_{x_C}^{\infty} \frac{2}{\sigma_Y}\phi\left(\frac{v-\mu_Y}{\sigma_Y}\right)\Phi\left(\alpha_Y\frac{v-\mu_Y}{\sigma_Y}\right)dv,$$

since the marginal shape parameters for the distributions of X and Y are now denoted σ_{X_M} and σ_{Y_M}, respectively. Inspection of the first term of the PSR reveals that as σ_{X_M} gets larger as ρ_{XY} increases (which is true by A.4.6), the PSR decreases. However, x_C also gets larger over ρ_{XY}, which causes the PSR to get larger.

4.5.3. Case Studies for Unequal Variances

In this section, the class of bivariate skew normal distributions only restriction is $\rho_{XY} = -1$. Thus k may be any value in the interval $(0, \infty)$ and α_D is allowed to take on any of the three states.

State: $\alpha_D \in \mathbb{R}$ (example: cases 1, 2, 3, and 4)

Given (4.1), $\rho_{XY} = -1$, and $k \in (0, \infty)$

➔ **Boundary**: $\text{PIQI}^{\rho_{XY}=-1} \geq \text{PSR}^{\rho_{XY}=-1}$

Proof:

Given $\rho_{XY} = -1$, then from proposition 4.3

$$
\begin{aligned}
2 \times PIQI^{\rho_{XY}=-1} &= 2 \times P(D < 0)_{\rho_{XY}=-1} \\
&= P(X \leq x_{-1}) + P(Y \geq x_{-1}) \\
&= P(X \leq x_L) + P(x_L \leq X < x_{-1}) + P(x_{-1} \leq Y < x_U) + P(Y > x_U) \\
&= P(X \leq x_L) + P(x_L \leq X < x_{-1}) + P(x_L \leq Y < x_U) - P(x_L \leq Y < x_{-1}) + \\
&\qquad P(Y > x_U) + P(X > x_U) - P(X > x_U) \\
&= P(X \leq x_L) + P(x_L \leq Y < x_U) + P(X > x_U) + \\
&\qquad P(x_L \leq X < x_{-1}) - P(x_L \leq Y < x_{-1}) + P(Y > x_U) - P(X > x_U) \\
&= PSR^{\rho_{XY}=-1} + \\
&\qquad P(x_L \leq X < x_{-1}) - P(x_L \leq Y < x_{-1}) + P(Y > x_U) - P(X > x_U)
\end{aligned}
$$

$PIQI^{\rho_{XY}=-1} \geq$ ½$PSR^{\rho_{XY}=-1}$ ■

Figure 4.11 provides a graphical illustration for the relationship between the PIQI and the PSR for each case over $0 \leq k \leq 2$ under the current class of distributions. The relationship displayed in each case shows that the PIQI, depicted as the dashed line, is greater than or equal to ½PSR, shown as the solid line, over all values of $k \in (0,2)$. This result is similar to proposition 2.2. However, the fundamental difference is that in the current class of models neither the $\text{PIQI}^{\rho_{XY}}$ nor the $\text{PSR}^{\rho_{XY}}$ may be maximized at $\rho_{XY} = -1$.

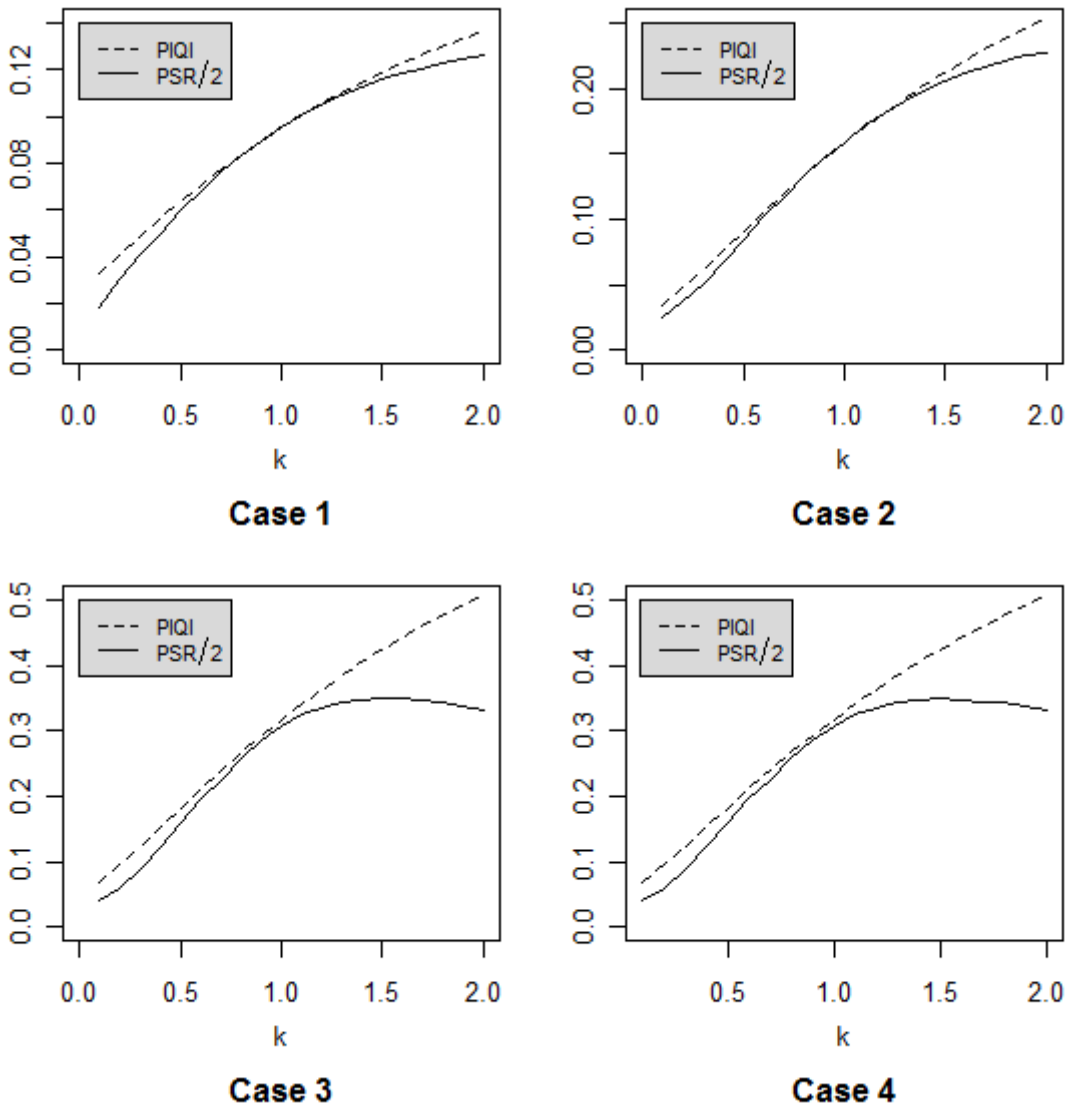


Figure 4.11: The relationship between the PSR and the PIQI over k
Each of the panels shows the PIQI as the dashed line and ½PSR as the solid line constrained by $\rho_{XY} = -1$ for the four cases given in Table 4.3. In all four cases $\text{PIQI}^{\rho_{XY}=-1} \geq \text{PSR}^{\rho_{XY}=-1}$.

4.6. The Observed PSR

The utility of the PSR as a measure of treatment heterogeneity and in particular IQIs is grounded in the idea that X and Y come from a bivariate distribution. Under the potential outcomes structure the bivariate distribution becomes a way to represent an individual's response to two treatments simultaneously. The PSR results from the marginal distributions of X and Y, while the PIQI results from $D = X - Y$; a linear combination of X and Y. When $(X \quad Y)^T$ is normally distributed, as in Chapter 3, the PSR was unaffected by the fact that ρ_{XY} was non-identifiable since the marginal distributions of X and Y are not affected by the correlation between X and Y. Consequently, the PSR was constant across $-1 \leq \rho_{XY} \leq 1$, while the PIQI was a strictly decreasing function of ρ_{XY}. Under the normal model, all parameters of the PSR are identifiable despite the fact that $(X \quad Y)^T$ is a model for potential outcomes. As such, a natural and observable boundary given by $\text{PIQI}^{max} \geq$ ½PSR (proposition 3.2) was formed. Boundaries for ρ_{XY} allow for direct assessment of the relationship between the PIQI and the PSR. For example, given $(X \quad Y)^T$ is normally distributed and $k = 1$, given the lower boundary for ρ_{XY} at -1, $\text{PIQI}^{max} =$ ½PSR .

When $(X \quad Y)^T$ is distributed as skew normal, the relationship between the PSR and the PIQI is not so direct since both are affected by the non-identifiable quantity ρ_{XY}. Table 4.5 is used to illustrate the challenge of assessing the boundaries for the PIQI from the observed PSR by illustrating the number of ways a skew normal distribution may be used to generate Figure 4.3. In fact, two univariate skew normal distributions were used to generate Figure 4.3 based on the parameterizations outlined in row 1 of Table 4.5 along with $\mu_X = 4$ and $\mu_Y = 0$. However, Figure 4.3 could just as easily have been generated from the bivariate parameterizations of any of

the last three rows of Table 4.5 in which each is primarily differentiated based on different values of α_D. For example, if the row 2 of Table 4.5 parameterization was used to generate Figure 4.3, then the ½PSR boundary equal to $½ \times 0.507 \cong 0.254$ does not hold since $\text{PIQI}^{max} = 0.227$. In fact, since $\alpha_D > 0$, $½\text{PSR}^{max} = ½\text{PSR}^{\rho_{XY}=0.10} = ½ \times 0.521 \cong 0.261$, which is of course is still greater than $\text{PIQI}^{max} = \text{PIQI}^{\rho_{XY}=-0.35} = 0.227$.

If X and Y were a result of the bivariate skew normal distribution parameterized by either row 3 or 4, then the boundary $\text{PIQI}^{max} \geq ½\text{PSR}^{\rho_{XY}} = 0.227$ holds, since the $\text{PSR}^{\rho_{XY}}$ is a decreasing function of ρ_{XY} under the $\alpha_D \leq 0$ condition. Which of these three skew normal bivariate distributions generated the overlap of $f_X(x)$ and $f_Y(x)$ and the resulting $PSR^{\rho_{XY}} \cong 0.514$ in Figure 4.3 is not determinable since ρ_{XY} is unidentifiable. Furthermore, the marginal shapes provide little help without constraints on ρ_{XY} since $(4.11a, b)$ makes their dependency on ρ_{XY} clear.

Table 4.5: Non-identifiable nature of the skew normal
Figure 4.3 was generated using two univariate skew normal distributions parameterized by the first row, where $\mu_X = 4$ and $\mu_Y = 0$. However, the skew normal bivariate distributions given in rows 2, 3, and 4 may also be used to generate Figure 4.3 illustrating the challenge of establishing a relationship between the PIQI and the PSR when $(X \quad Y)^T$ is skew normal.

Condition	$\boldsymbol{\rho_{XY}}$	$\boldsymbol{\alpha_{X_J}}$	$\boldsymbol{\alpha_{X_J}}$	$\boldsymbol{\sigma_X}$	$\boldsymbol{\sigma_Y}$	$\boldsymbol{\alpha_D}$	$\boldsymbol{\alpha_{X_J}}$	$\boldsymbol{\alpha_{X_J}}$	**PSR**	$\textbf{PIQI}^{max}$
Univariate	N/A	N/A	N/A	5	5	N/A	3	3	0.514	N/A
$\alpha_D > 0$	0.55	3	2	5.2	4.7	0.20	2.11	1.35	0.507	0.227
$\alpha_D = 0$	0.82	5	5	5	5	0	3	3	0.514	0.345
$\alpha_D < 0$	1.00	1	2	5	5	0	3	3	0.514	0.500

4.7. Future IQI Research using the PSR and the Skew Normal Model

4.7.1. Exploit the Correlation Link

So far the fact that under the skew normal model the PSR is also a function of ρ_{XY} has been considered a setback to furthering the development of the relationship between the PIQI and the PSR. However, the fact that both the PSR and the PIQI are functions of ρ_{XY} could be used to a distinct advantage to developing the relationship. In particular, note that each panel in Figure 4.10 demonstrates that both $\text{PSR}^{\rho_{XY}}$ and the $\text{PIQI}^{\rho_{XY}}$ behave similarly to ρ_{XY} within each of the primary categories of α_D. The similar response to ρ_{XY} comes through the shape parameters; α_D for the $\text{PIQI}^{\rho_{XY}}$ and α_{X_M} and α_{X_M} for the $\text{PSR}^{\rho_{XY}}$, which is where the dependency upon ρ_{XY} originates. The following example demonstrates how the dependency upon ρ_{XY} may be exploited in order to diagnose IQI taken from case 2 of Figure 4.10 where $\alpha_{X_J} = \alpha_{X_J} = 3$. Based on the the condition that $\alpha_D \leq 0$, the boundary $\text{PIQI}^{max} = ½\text{PSR}^{max} \cong 0.309$ may be employed. Suppose *a priori* information exists that sets $\rho_{XY} \geq 0.50$. Consequently, $½\text{PSR}^{0.50} \cong 0.243$ may actually be observed and serve as an upper bound for the $PIQI^{max}$, which is unobservable. Thus the fact that the PSR and the PIQI are linked by their common response to ρ_{XY} makes way to utilize this connection in new and useful ways. However, more structure and study is required to outline the nature of the PSR and the PIQI under the skew normal model.

4.7.2. Develop Relationships for Unequal Scale Parameters

The class of bivariate skew normal distribuitons in Section 4.5.5 are classified in part by the constraint that $k = 1$ implying equal scale parameters for the distributions of *X* and *Y*. The primary reason for this constraint is that the complicated formula for α_D given in (4.9) is

significantly simplified in (4.10) where $k = 1$. Thus the simplification makes it possible to capture the behavior of α_D and the PIQI as functions of ρ_{XY}. As seen in $(4.11a, b)$ the $k = 1$ constraint is not necessary for either of the marginal shape parameters α_{X_M} and α_{X_M}. Thus a greater understanding of how both ρ_{XY} and k simultaneously affect α_D, and as a consequence the PIQI, may yield import relationships and more general results between the PIQI and the PSR.

4.7.3. Proof for the PSR as a Strictly Decreasing function of the Correlation

The boundaries for the class of skew normal distributions defined by $k = 1$ and $\alpha_D \leq 0$ given in Section 4.5.2 were noted as case specific. The challenge presented was to determine that the $\text{PSR}^{\rho_{XY}}$ is strictly decreasing within this class. A closed form expression establishing this fact has not been derived yet.

4.7.4. Incorporate Location Shifts

Recall that under the $k = 1$ constraint, $\alpha_D = \alpha_{X_J} - \alpha_{Y_J} > 0$ shifts the distribution of D to the right as illustrated in Figure 4.5. Consequently, the $\text{PIQI}^{\rho_{XY}}$ is no longer a strictly decreasing function of ρ_{XY} as illustrated in Figure 4.10. Figure 4.10 also illustrates that $\alpha_{X_J} - \alpha_{Y_J} > 0$ has a very similar effect $\text{PSR}^{\rho_{XY}}$. It is important to note the definition of the $\text{PIQI}^{\rho_{XY}} = P(D < 0)_{\rho_{XY}} = \Phi\left(\frac{\mu_X - \mu_Y}{\sigma_D}\right) = \Phi\left(\frac{-\mu_D}{\sigma_D}\right)$ given in (3.1) when $(X \quad Y)^T$ is normally distributed is an expression regarding the difference of population means. However, in the skew normal model $E(D) \neq \mu_X - \mu_Y$, and so the definition of the $\text{PIQI}^{\rho_{XY}}$ may not have the same interpretations as it does when a normal model is specified. A possible solution may be to use a location shift to account for the shifting of the distribution of D caused by α_D.

Chapter 5 - IQI and Bivariate Normal Respsonse

5.1. Introduction to Joint Outcome Variables

Thus far the structure for studying and assessing IQIs was contained to a single outcome variable at a time. This chapter provides a structure for studying IQIs with respect to two outcome variables simultaneously, such as a measure of safety and efficacy. Furthermore, connections are made between the joint PIQI, formed by the study of individual effects over both outcome variables, and the joint PSR, generated by the overlap of the bivariate outcomes. Both outcome variables are assumed to be continuous.

5.1.1. Concurrent Study of Two Outcome Variables

In some cases it is vital that treatment T be more effective than R over both outcome variables. Bristol (2005) studied the simultaneous average effect of T over R on two outcome variables in which one response measured safety and the other measured efficacy. Bristol's methodology was motivated from the need to develop a new treatment T to improve the safety of treatment R, while maintaining at least the same efficacy. In such cases it may be important to know the proportion of the population that may be subject to an IQI over either, or both, outcome variables. Another situation, discussed by Lizotte, Bowling, and Murphy (2010), is motivated by the treatment of schizophrenia, which involves measurements on both symptoms and side effects producing two outcome variables of interest. In this case no treatments are currently available that works best on both outcomes so that an IQI over one outcome may be expected, while IQIs over both outcomes may be considered unacceptable. In situations like these, it may be important to know the proportion of the population that may experience an IQI over both

outcome variables simultaneously, a condition referred to as joint PIQI and denoted as PIQI_J herein. This chapter develops the PIQI_J and its counterpart, the joint PSR denoted by PSR_J.

5.1.2. Sets of Potential Outcomes

The outcome variables are identified as variables 1 and 2 in this chapter. Similar to previous chapters, which studied IQI over a single outcome variable, the response measured on variables 1 and 2 under both treatments T and R may be considered, although not observed. Thus X_1 and X_2 result from application of T and Y_1 and Y_2 result from application of R. Consequently, there are two 'true' effects defined as $D_1 = X_1 - Y_1$ and $D_2 = X_2 - Y_2$ for each individual. In this structure $(X_1,\ Y_1)$ and $(X_2,\ Y_2)$ may be considered two sets of potential outcomes, and although both D_1 and D_2 are definable, they are unobservable. Consistent with previous chapters, the assumption that T is more effective on average than R is conveyed by the constraint that $\mu_{X_1} \geq \mu_{Y_1}$ and $\mu_{X_2} \geq \mu_{Y_2}$ and, without loss of generality, it is assumed that positive values of the response variables are desirable.

5.1.3. Soy Treatment for Reduction in Both Cholesterol and Weight

In the soy-treatment working example the reduction in both cholesterol, say variable 1, and weight, say variable 2, may be considered simultaneously. As demonstrated in Chapter 3, an investigation into IQI may be conducted on cholesterol and weight change individually. However, a simultaneous study of IQI over both variables may provide a very different story. For example, the PIQI^{max} for cholesterol change was found to be approximately 0.34. A reduction in the PIQI_J^{max} over both variables $(i.e.\ \text{PIQI}_J^{max} < 0.34)$ would suggest that individuals in the population show improvement in at least one response, either cholesterol or weight reduction, since the PIQI_J^{max} gets smaller unless both $D_1 < 0$ and $D_2 < 0$.

5.2. The Joint PIQI

5.2.1. The Multivariate Normal Model

The structure established in this chapter for study of IQI over two variables simultaneously begins with placing the two sets of potential outcomes in a multivariate normal model given as

$$\begin{bmatrix} X_{1j} \\ X_{2j} \\ Y_{1j} \\ Y_{2j} \end{bmatrix} \sim iid\ N\left(\boldsymbol{\mu} = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \mu_{Y_1} \\ \mu_{Y_2} \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{X_1}^2 & \rho_X\,\sigma_{X_1}\sigma_{X_2} & \rho_{11}\sigma_{X_1}\sigma_{Y_1} & \rho_{12}\,\sigma_{X_1}\sigma_{Y_2} \\ \rho_X\,\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 & \rho_{21}\,\sigma_{X_2}\sigma_{Y_1} & \rho_{22}\sigma_{X_2}\sigma_{Y_2} \\ \rho_{11}\sigma_{X_1}\sigma_{Y_1} & \rho_{21}\,\sigma_{X_2}\sigma_{Y_1} & \sigma_{Y_1}^2 & \rho_Y\sigma_{Y_1}\sigma_{Y_2} \\ \rho_{12}\,\sigma_{X_1}\sigma_{Y_2} & \rho_{22}\sigma_{X_2}\sigma_{Y_2} & \rho_Y\sigma_{Y_1}\sigma_{Y_2} & \sigma_{Y_2}^2 \end{bmatrix} \right). \quad (5.1)$$

The correlation between outcome variables within treatment is denoted by ρ_X and ρ_Y. The correlation between outcome variables over different treatments is denoted by ρ_{12} and ρ_{21}. The correlation within outcome variables over different treatment is denoted by $\rho_{11}$and ρ_{22}. So $\rho_{11}$and ρ_{22} are the nonestimable correlations discussed in earlier chapters for one outcome variable.

A joint IQI may be defined as an individual that responds better to R than to T over both outcome variables simultaneously. The proportion of these individuals in a population may be defined as the probability that $(X_1 \quad X_2 \quad Y_1 \quad Y_2)^T$ falls in the region of $\mathbb{R}^4$ defined by $(Y_1 < X_1) \bigcap (Y_2 < X_2)$, which is analogous to the shaded region defined by $Y < X$ in consideration of a single outcome variable shown in Figure 1.1 (a). Thus the joint PIQI may be defined as

$$\text{PIQI}_J = P(Y_1 < X_1\ and\ Y_2 < X_2). \quad (5.2)$$

Under (5.1) the PIQI_J is affected by the degree to which both $\mu_{X_1} \geq \mu_{Y_1}$ and $\mu_{X_2} \geq \mu_{Y_2}$ and the characterization of Σ since not all parameters in Σ are estimable. Investigation of PIQI_J herein is simplified by assuming $\rho_X = \rho_Y = 0$ (although these parameters are estimable with observed data) and $\rho_{12} = \rho_{21} = 0$. Under these assumptions Σ may be simplified to

$$\Sigma_{\mathrm{c}} = \begin{bmatrix} \sigma_{X_1}^2 & 0 & \rho_{11}\sigma_{X_1}\sigma_{Y_1} & 0 \\ 0 & \sigma_{X_2}^2 & 0 & \rho_{22}\sigma_{X_2}\sigma_{Y_2} \\ \rho_{11}\sigma_{X_1}\sigma_{Y_1} & 0 & \sigma_{Y_1}^2 & 0 \\ 0 & \rho_{22}\sigma_{X_2}\sigma_{Y_2} & 0 & \sigma_{Y_2}^2 \end{bmatrix}, \tag{5.3}$$

where again ρ_{11} and ρ_{22} quantify within outcome correlations over different treatments, correlations analogous to ρ_{XY} in chapter 3. From the constraints inherent in the distribution of potential outcomes in (5.1), these correlations are not estimable from observed data and an assumption of $\rho_{11} = \rho_{22} = 0$ is not typically justifiable.

5.2.2. The Joint Distribution of D

Under the multivariate normal model in (5.1) and the correlation constraints imposed on Σ_{c}, the two true effects D_1 and D_2 follow a bivariate normal distribution with parameters $\boldsymbol{\mu_D}$ and Σ_{D}, and under assumptions made earlier D_1 and D_2 are independent,. Furthermore, following a similar strategy for the reparameterization of σ_D^2, let $\sigma_{X_1}^2 = \sigma_1^2$, $\sigma_{Y_1}^2 = k_1^2\sigma_1^2$, $\sigma_{X_2}^2 = \sigma_2^2$, and $\sigma_{Y_2}^2 = k_2^2\sigma_2^2$ giving

$$\boldsymbol{D} = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \sim N\left(\boldsymbol{\mu_D} = \begin{bmatrix} \mu_{X_1} - \mu_{Y_1} \\ \mu_{X_2} - \mu_{Y_2} \end{bmatrix}, \Sigma_{\mathrm{D}} = \begin{bmatrix} \sigma_1^2(1 + k_1^2 - 2k_1\rho_{11}) & 0 \\ 0 & \sigma_2^2(1 + k_2^2 - 2k_2\rho_{22}) \end{bmatrix}\right). \tag{5.4}$$

Thus the PIQI_J given in (5.2) may now be expressed as

$$\mathrm{PIQI}_J = P(D_1 < 0, D_2 < 0) = \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{\sigma_1\sqrt{1 + k_1^2 - 2k_1\rho_{11}}}, \frac{\mu_{Y_2} - \mu_{X_2}}{\sigma_2\sqrt{1 + k_2^2 - 2k_2\rho_{22}}}\right), \tag{5.5}$$

and bounds for PIQI_J may be given as

$$\Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{\sigma_1\sqrt{1 + k_1^2 - 2k_1}}, \frac{\mu_{Y_2} - \mu_{X_2}}{\sigma_2\sqrt{1 + k_2^2 - 2k_2}}\right) \leq \mathrm{PIQI}_J \leq \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{\sigma_1\sqrt{1 + k_1^2 + 2k_1}}, \frac{\mu_{Y_2} - \mu_{X_2}}{\sigma_2\sqrt{1 + k_2^2 + 2k_2}}\right). \tag{5.6}$$

Although $\boldsymbol{D}$ is unobservable, the bivariate distribution of $\boldsymbol{D}$ is illustrated with a contour plot in Figure 5.1 panel (a) under a presumed characterization of Σ_{D} in which $\sigma_1^2 = \sigma_2^2 = 0.5$, $k_1 = k_2 = 1$, and $\rho_{11} = \rho_{22} = 0$. The distribution, centered at $\boldsymbol{\mu_D}$ represented with an '×' and the shaded

area, is a depiction of the area over which (5.5) may be calculated. Panel (b) displays $f(d_1, d_2)$ over the shaded region displayed in panel (a). The area under $f(d_1, d_2)$ provides a graphical display of the quantity given in (5.5).

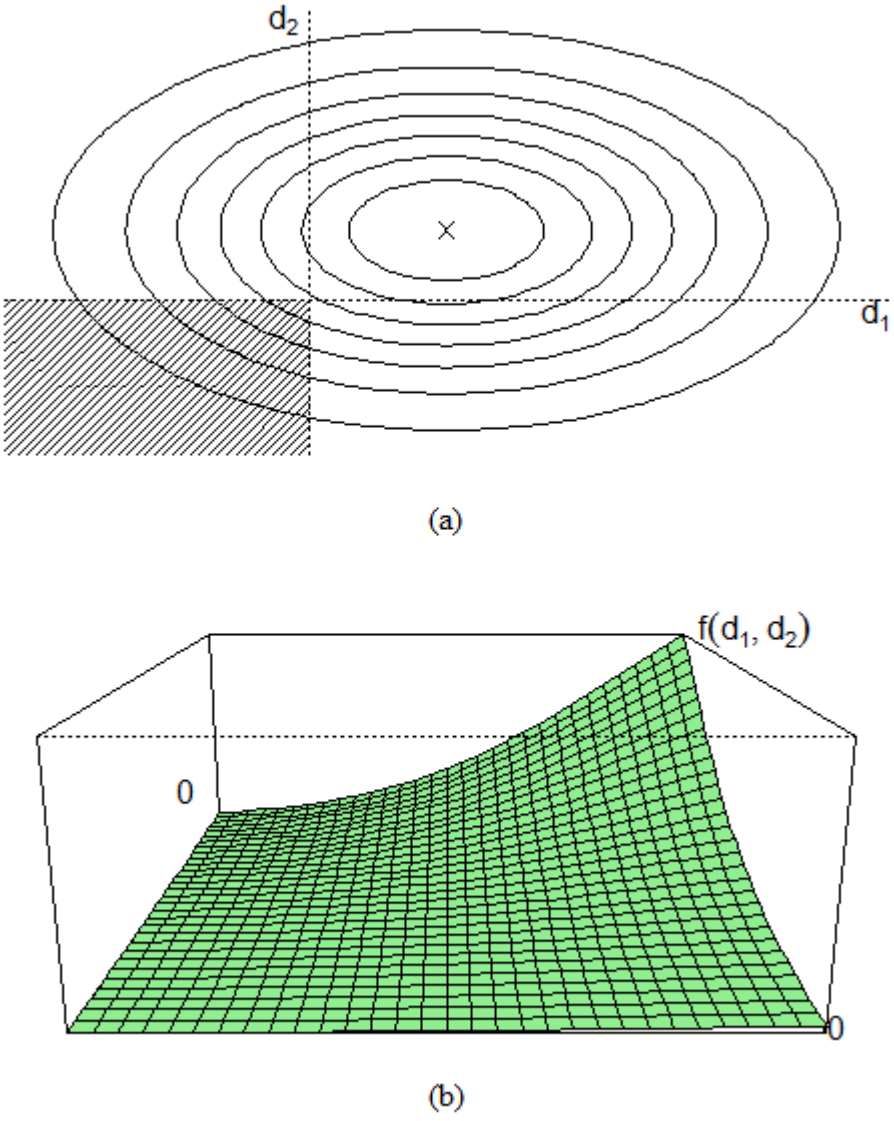


Figure 5.1: Illustration of the bivariate density of *D*
In panel (a) the bivariate pdf of D is illustrated with a contour plot. Panel (b) is the shaded portion of $f_{\boldsymbol{D}}(d_1, d_2)$ from panel (a).

5.3. The Joint PSR

5.3.1. Definition and Calculation of the Joint PSR

A joint PSR may be defined as the proportion of similar responses or 'overlap' of two bivariate distributions. Although a technical report given by Bradley and Piantadosi (1982) references development of a joint overlap, their work has been inaccessible. Given two bivariate densities $f_{X_1X_2}(u,v)$ and $f_{Y_1Y_2}(u,v)$, the joint PSR may be given as

$$\text{PSR}_J = \int\int min\left(f_{X_1X_2}(u,v), f_{Y_1Y_2}(u,v)\right) d_u d_v \,. \tag{5.7}$$

Under bivariate normality, which will be assumed in this chapter, the joint distributions of $(X_1 \quad X_2)^T$ and $(Y_1 \quad Y_2)^T$ may be derived under appropriate linear combinations of (5.1), where the off diagonal elements of Σ are equal to zero. Figure 5.2 illustrates the PSR_J under the constraints that both $(X_1 \quad X_2)^T$ and $(Y_1 \quad Y_2)^T$ are bivariate normal and $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Since $\mu_{X_1} \geq \mu_{Y_1}$ and $\mu_{X_2} \geq \mu_{Y_2}$, $(\mu_{Y_1} \quad \mu_{Y_2})^T$ is denoted with a '+' and $(\mu_{Y_1} \quad \mu_{Y_2})^T$ is denoted with a '×'. The axes are depicted with dotted lines and the PSR_J is illustrated by the overlap of the density contours.

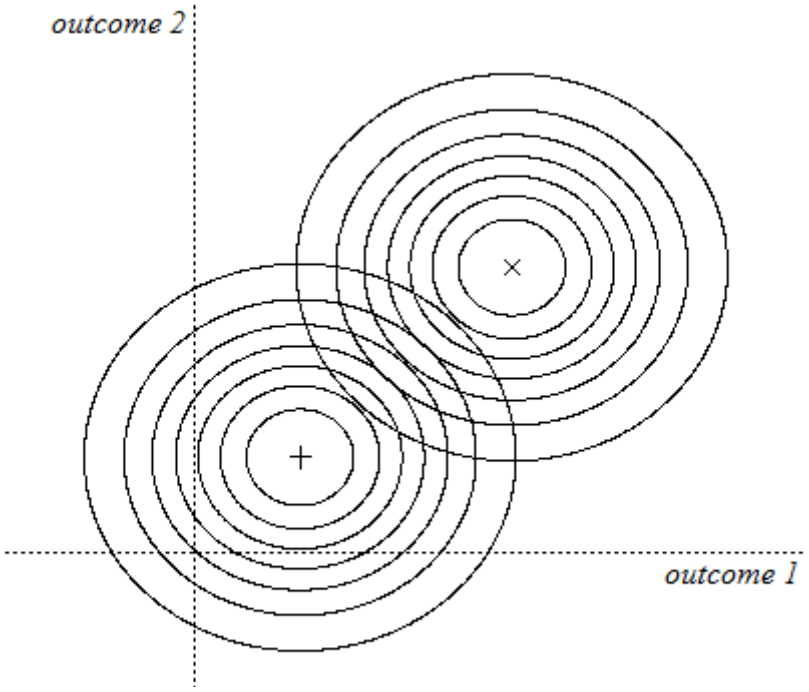


Figure 5.2: Joint PSR
The contour plots of the distributions of $(X_1 \quad X_2)^T$ centered at $'\times'$ and $(Y_1 \quad Y_2)^T$ centered at $'+'$. The overlap of the contours is the joint PSR. Plotting region represents points (x_1, x_2) and (y_1, y_2).

5.3.2. Calculation of joint PSR under the Normal Model

Based on (5.7) the calculation of PSR_J is made possible by identifying the points at which the pdf's of $(X_1 \quad X_2)^T$ and $(Y_1 \quad Y_2)^T$ are equal. Similar to the point of equality x_E in the univariate setting for the case when $k = 1$ (see Figure 2.3), the PSR_J depends on a line of equality over the points (x_E, y_E), which is denoted by $y_E = \beta_{0_E} + \beta_E x_E$. The solid line in Figure 5.3 panel (a) displays an example of such a line, which splits the overlap area in half, and where $f_{X_1X_2}(x_E, y_E) = f_{Y_1Y_2}(x_E, y_E)$, and otherwise $f_{X_1X_2}(x, y) \neq f_{Y_1Y_2}(x, y)$. The calculation of β_{0_E} and β_E are given in the next proposition.

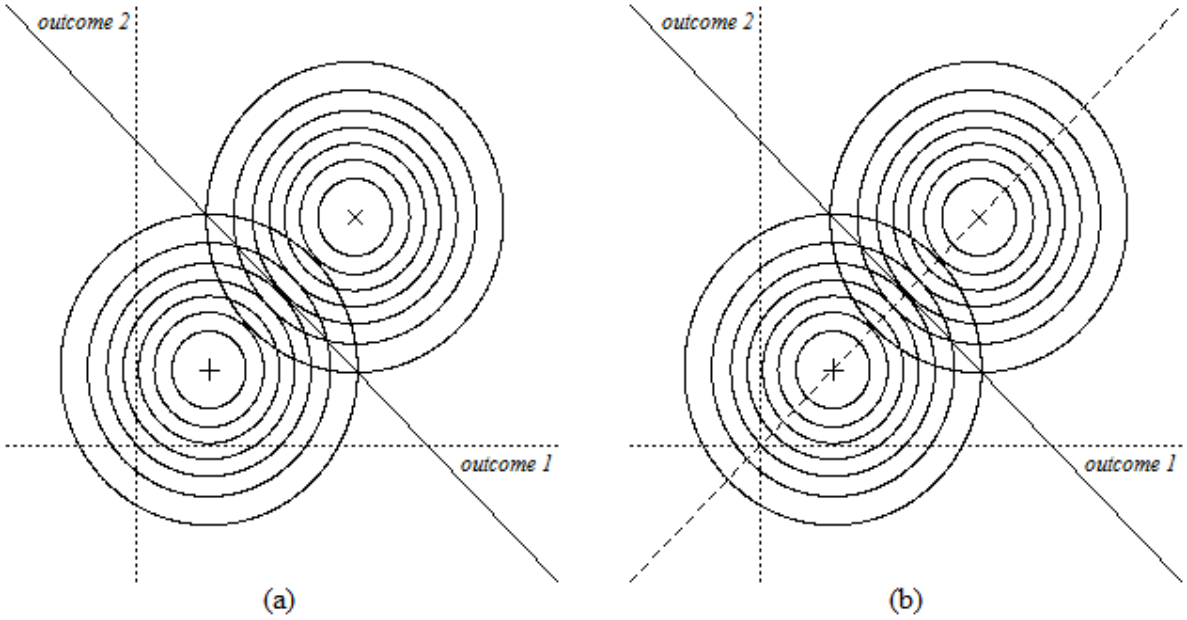


Figure 5.3: The lines of equality where the densities are equal
The solid line is called the line of equality found by (5.8). The dashed line is perpendicular to the line of equality and found by (5.9).

❖ **Proposition 5.1**

Given (5.1), $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and $k_1 = k_2 = 1$, the line of equality is defined by $y_E = \beta_{0_E} + \beta_E x_E$, where

$$\beta_{0_E} = \frac{\mu_{X_1}^2 + \mu_{X_2}^2 - \mu_{Y_1}^2 - \mu_{Y_2}^2}{2(\mu_{X_2} - \mu_{Y_2})} \quad \text{and} \quad \beta_E = \frac{\mu_{Y_1} - \mu_{X_1}}{\mu_{X_2} - \mu_{Y_2}} \tag{5.8}$$

Proof:

From (5.1) let $\boldsymbol{\mu'_X} = \begin{pmatrix}\mu_{X_1} & \mu_{X_2}\end{pmatrix}^T$, $\boldsymbol{\mu'_Y} = \begin{pmatrix}\mu_{Y_1} & \mu_{Y_2}\end{pmatrix}^T$, $\sum_X = \begin{pmatrix}\sigma_{X_1} & 0 \\ 0 & \sigma_{X_2}\end{pmatrix}$, and $\sum_Y = \begin{pmatrix}\sigma_{Y_1} & 0 \\ 0 & \sigma_{Y_2}\end{pmatrix}$,

then if $k_1 = k_2 = 1$ and $\boldsymbol{x}' = (x\ y)$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{Y}}(\boldsymbol{x})$$

$$\frac{1}{2\pi|\sum_X|^{1/2}} exp\left\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_X})' \sum\nolimits_X (\boldsymbol{x}-\boldsymbol{\mu_X})\right\} = \frac{1}{2\pi|\sum_Y|^{1/2}} exp\left\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_Y})' \sum\nolimits_Y (\boldsymbol{x}-\boldsymbol{\mu_Y})\right\}$$

$$(\boldsymbol{x}-\boldsymbol{\mu_X})'(\boldsymbol{x}-\boldsymbol{\mu_X}) = (\boldsymbol{x}-\boldsymbol{\mu_Y})'(\boldsymbol{x}-\boldsymbol{\mu_Y})$$

$$\begin{pmatrix}x-\mu_{X_1} \\ y-\mu_{X_2}\end{pmatrix}' \begin{pmatrix}x-\mu_{X_1} \\ y-\mu_{X_2}\end{pmatrix} = \begin{pmatrix}x-\mu_{Y_1} \\ y-\mu_{Y_2}\end{pmatrix}' \begin{pmatrix}x-\mu_{Y_1} \\ y-\mu_{Y_2}\end{pmatrix}$$

$$2y\mu_{X_2} - 2y\mu_{Y_2} = 2x\mu_{Y_1} - 2X\mu_{X_1} + \mu_{X_1}^2 + \mu_{X_2}^2 - \mu_{Y_1}^2 - \mu_{Y_2}^2$$

$$y = \frac{\mu_{X_1}^2 + \mu_{X_2}^2 - \mu_{Y_1}^2 - \mu_{Y_2}^2}{2\left(\mu_{X_2} - \mu_{Y_2}\right)} + \frac{\mu_{Y_1} - \mu_{X_1}}{\mu_{X_2} - \mu_{Y_2}} x. \quad \blacksquare$$

Calculation of the PSR_J under (5.1) may be facilitated by the calculation of a line perpendicular to the line of equality that goes through the points $\left(\mu_{X_1}, \mu_{X_2}\right)$ and $\left(\mu_{Y_1}, \mu_{Y_2}\right)$, which is termed the line of perpendicularity and denoted by $y_P = \beta_{0_P} + \beta_P x_P$. The intercept and slope of this line can be calculated by

$$\beta_{0_P} = \frac{\mu_{X_2}\mu_{Y_1} - \mu_{X_1}\mu_{Y_2}}{\mu_{Y_1} - \mu_{X_1}} \quad \text{and} \quad \beta_P = \frac{\mu_{Y_2} - \mu_{X_2}}{\mu_{Y_1} - \mu_{X_1}}. \tag{5.9}$$

The dashed line added to panel (a) of Figure 5.3 shown panel (b) illustrates the line of perpendicularity. The intersection of the line of equality and the line of perpendicularity divides the overlap into four equal areas. When $\beta_E = 0$, the PSR_J can be expressed in closed form as

$$\text{PSR}_J = P\left(X_1 \le \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) + P\left(X_1 > \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right)$$

$$+P\left(Y_1 > \frac{\mu_{X_1} + \mu_{Y_1}}{2}, Y_2 \ge \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) + P\left(Y_1 \le \frac{\mu_{X_1} + \mu_{Y_1}}{2}, Y_2 \ge \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right). \tag{5.10}$$

The fact that these four terms are equal is shown in the appendix (see A.5.1). Representation of the PSR_J in the form of (5.10) is important, since a simple transformation can change the means of the distributions so that $\mu_{X_1} = \mu_{Y_1}$ without changing the PSR_J (or 'overlap'). Thus any PSR_J generated under the constraints of (5.1), (5.3), $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and $k_1 = k_2 = 1$ is composed of four equal areas, and can be transformed to meet the conditions of (5.10).

5.4. Relationship between the joint PIQI and the joint PSR

❖ **Proposition 5.2**

Given $(\mathbf{5.1})$, $(\mathbf{5.3})$, $\boldsymbol{\sigma_1^2 = \sigma_2^2 = \sigma^2}$, and $\boldsymbol{k_1 = k_2 = 1}$

$$\text{PIQI}_J^{max} \le ¼\text{PSR}_J$$

Proof:

Part 1: Show that $\text{PICI}_J^{max} = ¼\text{PSR}_\text{J}$ at $\beta_E = 0$

Let $\beta_E = 0$, then $\mu_{X_1} = \mu_{Y_1}$ and from (5.8)

$$\text{PIQI}_J^{max} = \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right) = \Phi\left(0, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right)$$

and from (5.10)

$$¼\text{PSR}_J = P\left(X_1 \le \mu_{X_1}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right)$$

$$= P\left(Z_1 \le \frac{\mu_{X_1} - \mu_{X_1}}{\sigma}, Z_2 \le \frac{\frac{\mu_{X_2} + \mu_{Y_2}}{2} - \mu_{X_2}}{\sigma}\right)$$

$$= P\left(Z_1 \le 0, Z_2 \le \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right) = \Phi\left(0, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right) \blacksquare$$

Part 2: Show that $\text{PIQI}_J^{max} = ¼\text{PSR}_\text{J}$ at $\beta_P = 0$

Let $\beta_P = 0$, then $\mu_{X_2} = \mu_{Y_2}$ and from (5.8)

$$\text{PIQI}_J^{max} = \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right) = \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, 0\right)$$

and from (5.10)

$$¼\text{PSR}_J = P\left(X_1 \le \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \mu_{X_2}\right)$$

$$= P\left(Z_1 \le \frac{\frac{\mu_{X_1} + \mu_{Y_1}}{2} - \mu_{X_1}}{\sigma}, Z_2 \le \frac{\mu_{X_2} - \mu_{X_2}}{\sigma}\right)$$

$$= P\left(Z_1 \le \frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, Z_2 \le 0\right) = \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, 0\right) \blacksquare$$

Part 3: Show that $\text{PIQI}_J^{max} < ¼\text{PSR}_\text{J}$ at $\beta_E < 0$

Let $\beta_E < 0$, then $\mu_{X_1} > \mu_{Y_1}$ and $\mu_{X_2} > \mu_{Y_2}$ and from (5.9)

$$\text{PIQI}_J^{max} = \Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right)$$

and from (5.10) both

$$\Phi\left(\frac{\mu_{Y_1} - \mu_{X_1}}{2\sigma}, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right) < \Phi\left(0, \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right)$$

and

$$\Phi\left(\frac{\mu_{Y_1}-\mu_{X_1}}{2\sigma},\frac{\mu_{Y_2}-\mu_{X_2}}{2\sigma}\right) < \Phi\left(\frac{\mu_{Y_1}-\mu_{X_1}}{2\sigma},0\right) \blacksquare$$

An example of proposition 5.3 is illustrated in Figure 5.4 under the constraints of proposition 5.3 and $\sigma^2 = 1$. In panel (a) $(\mu_{X_1} \quad \mu_{X_2})^T = (3 \quad 3)^T$ and $(\mu_{Y_1} \quad \mu_{Y_2})^T = (1 \quad 1)^T$. Thus the distance between the origin $(0 \quad 0)^T$ and $(\mu_{Y_1} \quad \mu_{Y_2})^T$ is equal to $\sqrt{2}$ and the distance between $(\mu_{Y_1} \quad \mu_{Y_2})^T$ and $(\mu_{X_1} \quad \mu_{X_2})^T$ is equal to $\sqrt{8}$. In panel (a) the $\text{PICI}_J^{max} = 0.025$ and $¼\text{PSR}_\text{J} = 0.039$. In panel (b) the means of the bivariate distributions change to $(\mu_{X_1} \quad \mu_{X_2})^T = (\sqrt{2}+\sqrt{8} \quad \sqrt{2})^T$ and $(\mu_{Y_1} \quad \mu_{Y_2})^T = (\sqrt{2} \quad \sqrt{2})^T$ so that the distances between $(\mu_{Y_1} \quad \mu_{Y_2})^T$ and $(\mu_{X_1} \quad \mu_{X_2})^T$ remains equal to $\sqrt{8}$. Thus in panel (b) $\text{PICI}_J^{max} = ¼\text{PSR}_\text{J} = 0.039$.

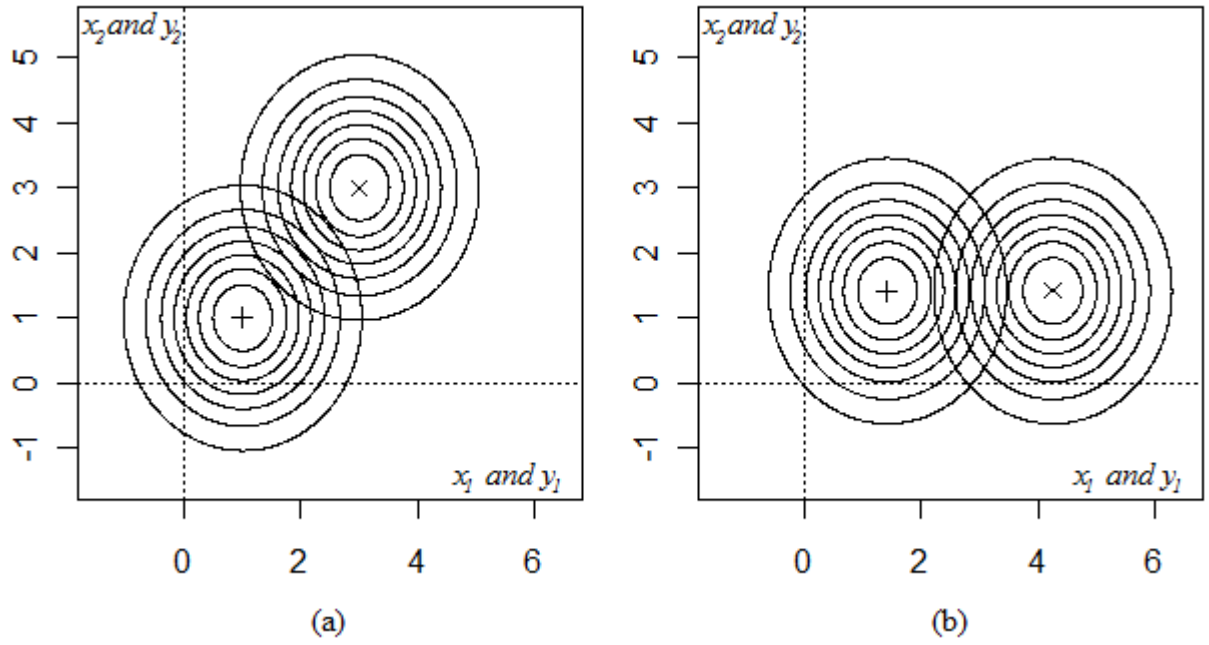


Figure 5.4: An illustration of proposition 5.3
The contour plots of the distributions of $(X_1 \quad X_2)^T$ centered at $'\times'$ and $(Y_1 \quad Y_2)^T$ centered at $'+'$. The overlap of the contours is the joint PSR. Although the PSR remains unchanged across both panels, the PIQI^{max} is smaller in panel (a). In panel (a) $\text{PIQI}^{max} = 0.025$, while $¼\text{PSR}_\text{J} = 0.039$. However, in panel (b) $\text{PIQI}^{max} = ¼\text{PSR}_\text{J} = 0.039$.

5.5. Future work

The relationship between the PIQI_J^{max} and the PSR_J given in proposition 5.2 is constrained by both $\sigma_1^2 = \sigma_2^2$ and $k_1 = k_2 = 1$. Such constraints may be overly restrictive for broad use of the boundary given in proposition 5.2. Consequently, allowing for more flexibility by eliminating one or both of these constraints is recommended for further study. Allowing σ_1^2 and σ_2^2 to be different will impact both quantities. First, the diagonal elements of Σ_D may no longer be equal, affecting the PIQI_J. Second, the shapes of $f(x_1, x_2)$ and $f(y_1, y_2)$ may be very different, thus the line of equality wherein $f(x_1, x_2) = f(y_1, y_2)$ will be affected, impacting the PSR_J. Allowing for different variances within the bivariate distributions of $(X_1 \quad X_2)^T$ and $(Y_1 \quad Y_2)^T$ should have similar consequences.

Constraints on the correlations given in (5.1) may also inhibit the development of a more general connection between PIQI_J and the PSR_J. Future research is recommended which allows both ρ_{12} and ρ_{12} to be bound by -1 and 1 and both ρ_X and ρ_Y to be estimated from the data. The variance-covariance matrix Σ in (5.1) provides the structure for studying the PICI_J and the PSR_J and any unidentified relationships that may exist.

The approach taken in this chapter is that a joint IQI is defined as an individual that experiences an IQI on both outcome variables concurrently. The PIQI_J quantity is then motivated by the shaded area represented in Figure 5.1 panel (a). However, recall the interest in IQIs motivated by Bristol (2005), wherein an IQI across either one or both outcome variables is of significance. Figure 5.5 illustrates the area of interest given as the shaded region over the distribution of $\boldsymbol{D}$. That is, the safety of a variable may be improved for an individual, but any IQI for the efficacy, which is expected to not change, is of primary interest.

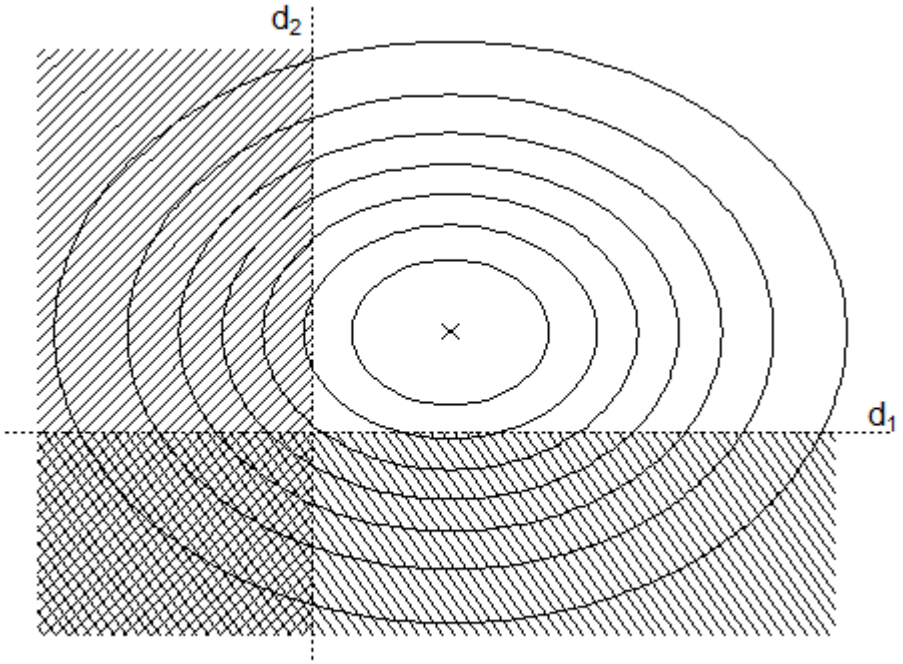


Figure 5.5: An alternative joint PIQI Definition
The entire shaded area represents the area over the density of D where an IQI over either outcome variable may exhibit an IQI. The double shaded area represents the space of D over which both outcome variables exhibit an IQI.

Chapter 6 - Summary and Future Direction

Sharma (2010, p. 1) states that "the problem of averages" is that "our entire medical philosophy of 'evidence-based' medicine seems built on the assumption that averages can reflect the true benefit (or risk) of a drug, when in real life (or medical practice) there is no such thing as the truly average patient." Treatment heterogeneity, a particularly challenging phenomenon to analyze at the individual level, is at the root of Sharma's issue with 'averages.' Evidence that treatment heterogeneity has been of interest to scientists and medical practitioners for well over a century has been presented herein (e.g. Darwin, 1871). In the modern era, formal recognition of the problems associated with treatment heterogeneity among patients in clinical trials has been recognized as early as 1977 when individual bioequivalent regulation was introduced (Hwang et al., 1978). Before that, policy for evaluating bioequivalence occurred exclusively at the level of average effects of T with respect to R. Interestingly, the statement by Chow and Liu (2000, p. 2719) that "Important clinical and/or scientific issues in design and analysis of individual bioequivalence still remain unsolved" reflects a continued lack of treatment heterogeneity analysis today. Much of the work presented in this dissertation is intended to provide both a structure and methods to improve and enhance the analysis of treatment heterogeneity in a randomized controlled trial. The demand for individualized treatment regimes suggests that the study of treatment heterogeneity at the individual level is a worthwhile effort. This work can help provide a basis for new discoveries.

6.1. Summary of Dissertation

Chapter 3 introduced a new structure for studying individual treatment heterogeneity. The fact that S-T interaction can be decomposed into two components was shown. This decomposition led to a new quantity; an IQI. This quantity puts the focus on individuals, as suggested by Sharma, by looking for those individuals that respond opposite to the response suggested by a study of average effects. A positive probability of an IQI (PIQI) means that there are individuals in the population that respond better to R than T even though the average effects suggest otherwise. In Chapter 3 maximum and minimum bounds were placed on the PIQI using existing techniques. The development of the PIQI permitted new connections to be made between S-T interaction and two existing methods for studying treatment heterogeneity; subset interaction and the PSR, or density overlap. These connections led to a series of simple diagnostics' used to evaluate the PIQI in a population or a subpopulation. Longford (1999) suggested that all clinical trial results should include some assessment of individual effects. The diagnostics suggested in this work provide direct information about the more severe component of S-T interaction with very little investment in time and no additional data or structure imposed on the trial. For example, the ideas presented herein may help establish individualized bioequivalence standards such as requiring the maximum PIQI be below a certain threshold.

The connections between the PIQI and the PSR in Chapter 3 led to the development of Chapters 4 and 5. Since Chapter 3 is constrained to the normal model, Chapter 4 explores the connections between the PIQI and the PSR using a skew normal model. This chapter led to a completely new structure and development for both the PIQI and the PSR. Chapter 5 also presents connections between the PIQI and the PSR, but for bivariate responses. As a result, the probability of an IQI over two sets of responses to T with respect to R may be quantified.

6.2. Future Direction

6.2.1. The PIQI used to Detect Response Improvements

The PIQI focuses on the most severe form of S-T interaction. An alternative strategy may be to define a similar quantity as the IQI, such as IQI^{+}, to represent individuals who respond in a positive direction to treatment such as $\text{PIQI}^{+} = P(D > 0)$. Thus it may be possible to detect subpopulations that respond positively to a treatment that has been shown to have no effect on average over the entire population (see Zhao, Dmitrienko, and Tamura, 2010).

6.2.2. The PIQI minimum

It has been reported that when treatment variances are unequal ($k \neq 1$), individual responses to T with respect to R would not be constant (Stine and Heyse, 2001). This idea is reflected in the PIQI^{min} since, when $k \neq 1, \text{PIQI}^{min} > 0$. This idea is also reflected in the overlap of two densities. Suppose *X* and *Y* follow a normal distributions. Then the overlap of the distributions will not be complete (equal 100%) unless *X* and *Y* have equal variances, even if $\mu_X = \mu_Y$. Likewise, if $k \neq 1$, PSR cannot equal 1. The PIQI^{min} is an important quantity, since, when $\text{PIQI}^{min} > 0$, there are individuals in the population that will exhibit an IQI for treatments T and R. Despite the importance of PIQI^{min} quantity, it has received little attention here. Part of the reason is that there was no connection found between the PIQI^{min} and the PSR. However, there may be alternatives to studying the PIQI^{min} outside the PSR that can lend insight to this important quantity. For example, differences among the quantiles of the distributions of *X* and *Y* might also suggest that $\text{PIQI}^{min} > 0$. Thus relationships between quantiles and the PIQI^{min} may be found that will provide more understanding of IQI within a population. Recall that Gail and Simon (1985) formalized tests for a QI. Tests for differences in distributions by quantile comparisons include Kemp, Yang, Perng, and Nelson (1993) and Elmore, Hettmansperger, and

Xuan (2006). Such methods may provide information about the PIQI and the PIQI^{min} in particular.

6.2.3. Reduction in the Variance of D using the Correlation

Other improvements revolve around methods to tighten the bounds on the PIQI. There may be several ways to do this, most of which will center on reducing σ_D^2. Section 3.6 uses covariates to reduce σ_D^2. This can be done when there is more than one observation on the same treatment within groups. An improvement to this strategy may come from Perrett and Higgins (2004). Perrett and Higgins estimate, what they call, the between unit component of variation when there is no replication within treatment group. Thus, following a similar strategy, it may be possible to remove the constraints on the within groups replication and still reduce σ_D^2

6.2.4. Calculation of the PIQI Conditioning on the Observed Values

An alternative idea is motivated by the soy treatment example. As described in Section 3.2, although data from the soy treatment has been observed, the PIQI as defined herein only uses the treatment means to specify the distribution for $(X \quad Y)^T$. That is, $(X \quad Y)^T$ is drawn from the population of interest without prior knowledge of an individual's response to either treatment. However, given the observed data, the value of the missing data may be modeled conditioned on the observed value of the response. From this perspective, new bounds on the PIQI could be formed that may be smaller, conditioned on the observed outcome. An example of how this might be done is given using the soy treatment example.

Figure 6.1 panel (a) is a duplication of Figure 1.1. Recall that the open triangles along the x axis are the observed values for the first 8 patients on treatment T and the open circles along the y axis are the observed values for patients 9 through 16 on treatment R. These values

are given in Table 1.1 and given again here in Table 6.1 (a). One strategy for investigating evidence of an IQI with the soy and reference treatments is to impute the missing data values $x_9, x_{10}, \cdots, x_{16}$ and $y_1, y_2, \cdots, y_8$ under a bivariate imputation model for direct comparison with the observed values. Since $x_1, x_2, \cdots, x_8, y_9, y_{10}, \cdots, y_{16}$ are observed, the model would need to generate pairs of values. Under the assumption that $(X, Y)^T$ is distributed as bivariate normal, models for the imputed values of X and Y may be given by

$$X_{j,impute} \sim N\left((\mu_X - \beta_Y \mu_Y) + \beta_X y_j,\ \sigma_X^2(1-\rho_{XY}^2)\right) \qquad (6.1a)$$

and

$$Y_{j,impute} \sim N\left((\mu_Y - \beta_X \mu_X) + \beta_Y x_j,\ \sigma_Y^2(1-\rho_{XY}^2)\right), \qquad (6.1b)$$

respectively, where $j = 1, 2, \cdots, 16$. Recall that despite the missing data, all the parameters of $(6.1a, b)$ can be estimated from the data with the exception of the correlation ρ_{XY},. Thus in addition to the distributional assumption of normality, to employ (6.1) ρ_{XY} must be assumed as well. However, leaving the observed values in the contour plots given in Figure 6.1 panels (b), (c), and (d) as opposed to Figure 3.2 panels (b), (c), and (d), respectively gives additional insight about possible IQIs in the sample, and therefore, the population.

As an example, consider Figure 6.1 panel (b). If the correlation is postulated at $\rho_{XY} = -0.85$, then the two largest values of y being $y_{10} = 22$ and $y_{13} = 31$ are almost certain to produce an x_{10} and an x_{13} pair using $(6.1a)$ such that (x_{10}, y_{10}) and (x_{13}, y_{13}) are in the IQI

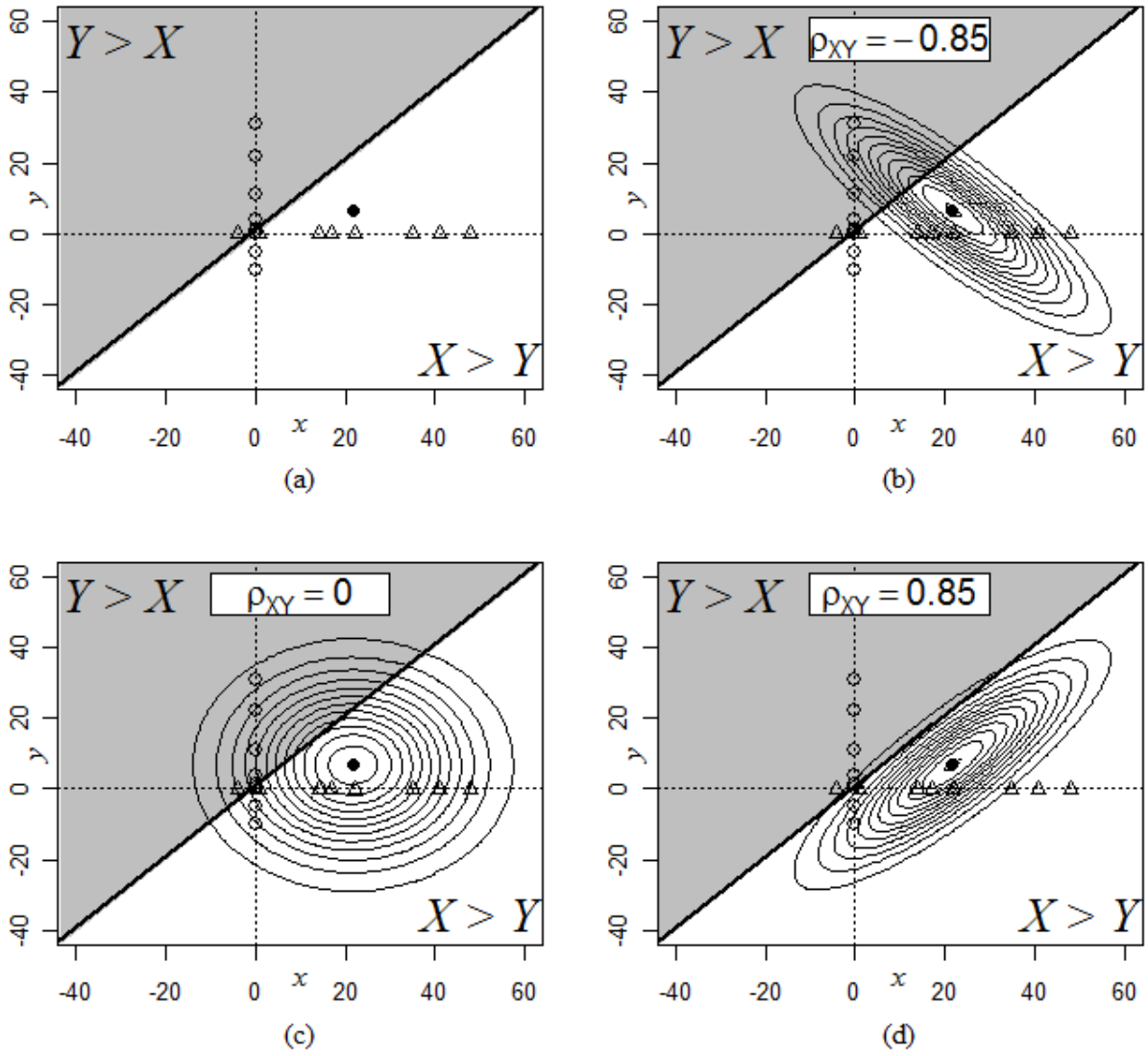


Figure 6.1: The PIQI generated from the conditional distributions of *X* and *Y*
This figure illustrates the development of the PIQI from the conditional univariate distributions conditioned on the observed responses. The shape of the distribution of $(X \quad Y)^T$, and therefore, the conditional distributions of X and Y are dependent upon the unidentifiable value of ρ_{XY}.

Table 6.1: Assessing treatment heterogeneity by imputation
Panel (a) illustrates the difficulty of assessing individual treatment heterogeneity with observable data. Panels (b) – (d) illustrate one method by imputing missing values. Any method requires assumptions about the individual correlation structure ρ_{XY}. Each panel references a corresponding panel from Figure 6.1.

(a) Univariate Normal

Observed and Missing Responses			Individual Response Comparisons	
Patient	Soy(X)	Ref. (Y)	$X-Y$	$I(Y>X)$
1	17	NA	?	?
2	41	NA	?	?
3	35	NA	?	?
4	-4	NA	?	?
5	22	NA	?	?
6	48	NA	?	?
7	14	NA	?	?
8	1	NA	?	?
9	NA	1	?	?
10	NA	22	?	?
11	NA	-5	?	?
12	NA	11	?	?
13	NA	31	?	?
14	NA	0	?	?
15	NA	4	?	?
16	NA	-10	?	?
Average	**21.75**	**6.75**	?	?

(b) $\rho_{XY} = -0.85$

Observed and Imputed Responses			Individual Response Comparisons	
Patient	Soy(X)	Ref. (Y)	$X-Y$	$I(Y>X)$
1	17	**7**	10	0
2	41	**-25**	66	0
3	35	**-6**	41	0
4	-4	**30**	-34	1
5	22	**6**	16	0
6	48	**-25**	73	0
7	14	**5**	9	0
8	1	**43**	-42	1
9	**16**	1	15	0
10	**9**	22	--13	1
11	**41**	-5	46	0
12	**28**	11	17	0
13	**-3**	31	-34	1
14	**30**	0	30	0
15	**18**	4	14	0
16	**38**	-10	48	0
Average	**21.94**	**5.56**	**16.37**	**0.25**

(c) $\rho_{XY} = 0$

Observed and Imputed Responses			Individual Response Comparisons	
Patient	Soy(X)	Ref. (Y)	$X-Y$	$I(Y>X)$
1	17	**1**	16	0
2	41	**16**	25	0
3	35	**8**	27	0
4	-4	**-7**	3	0
5	22	**26**	-4	1
6	48	**-7**	55	0
7	14	**10**	4	0
8	1	**4**	-3	1
9	**40**	1	39	0
10	**47**	22	25	0
11	**0**	-5	5	0
12	**12**	11	1	0
13	**24**	31	-7	1
14	**10**	0	10	0
15	**60**	4	56	0
16	**21**	-10	31	0
Average	**24.25**	**6.56**	**17.69**	**.1875**

(d) $\rho_{XY} = 0.85$

Observed and Imputed Responses			Individual Response Comparisons	
Patient	Soy(X)	Ref. (Y)	$X-Y$	$I(Y>X)$
1	17	**0**	17	0
2	41	**12**	29	0
3	35	**15**	20	0
4	-4	**-25**	21	0
5	22	**-10**	32	0
6	48	**33**	15	0
7	14	**14**	0	0
8	1	**-17**	18	0
9	**33**	1	32	0
10	**38**	22	16	0
11	**1**	-5	6	0
12	**46**	11	35	0
13	**61**	31	30	0
14	**10**	0	10	0
15	**20**	4	16	0
16	**3**	-10	13	0
Average	24.13	4.75	19.38	0

region (shaded region). The reason being both $y_{10}$and y_{13} are considerably larger than $\mu_Y = 6.75$ and even larger than $\mu_X = 21.75$ Reference to Table 6.1 (b) shows that use of $(6.1a, b)$ produced an IQI for both individuals, where the individuals that exhibited and IQI are identified by the shaded rows. In fact, the imputed data in Table 6.1 (b) shows that, using $\rho_{XY} = -0.85$ there were four IQIs for that case.

This information may be used in many ways. One of which may be to see what value of ρ_{XY} may be required before the pairs such as (x_{10}, y_{10}) and (x_{13}, y_{13}) have a high probability of not falling within the IQI region. Figure 6.1 panel (c) illustrates that even when $\rho_{XY} = 0$, the imputed (x_{10}, y_{10}) and (x_{13}, y_{13}) pairs have a higher probability of falling within the shaded IQI region than outside the shaded region due to the large within treatment variability. On the other hand one may ask what value of ρ_{XY} should be necessary before the chance of an IQI is eliminated. Figure 6.1 (d) illustrates the example with $\rho_{XY} = 0.85$. Surprisingly, even with X and Y so highly correlated, the contour plot in panel (d) shows that the probability mass over the shaded area is positive. In fact, in 1000 imputed data sets approximately half had at least one IQI. The methods of utilizing conditional distributions based on the observed data have not been fully explored. Likewise, the imputation models of $(6.1\ a, b)$ have not been fully developed. For example, there is additional variance considerations associated with imputed data sets. Also, the models may be improved by use of other observed covariates. The increased information should yield important results in making statements about σ_D^2, which may lead to an improved IQI analysis based on observed data in a study.

References

Allison, D.B., Gadbury, G.L., Schwartz, L.G., Murugesan, R., Kraker, J., Heshka, S., Fontaine, K.R., and Heymsfield, S.B. (2003). "A novel soy-based meal replacement formula for weight loss among obese individuals: a randomized controlled clinical trial." *European Journal of Clinical Nutrition*. **57**, 514 – 522.

Azzalini, A. (1985). "A class of distributions which include the normal ones." *Scandinavian Journal of Statistics.* **12**(2), 171 – 178.

Azzalini, A., Capitanio, A. (1999). "Statistical applications of the multivariate skew normal distribution." *Journal of the Royal Statistical Society. Series B (Statistical Methodology).* **61**(3), 579 – 602.

Azzalini, A., Valle, (1996). "The multivariate skew-normal distribution." *Biometrika*, **33**(4), 715 –726.

Balaam, L. N. (1968), "A Two-period Design with *t*2 Experimental Units," *Biometrics*, 24, 61– 73.

Bradley, E. L. and Piantadosi, S. (1992). The overlapping coefficient as a measure of agreement between distributions. *Tech. Report, Department of Biostatistics and Biomathematics, University of Alabama at Birmingham.*

Bristol, D. R. (2005). "Superior safety in noninferiority trials," *Biometrical Journal*. **47**(1), 75 – 81.

Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics*. **36**, 69 – 79.

Byar, D. P. and Corle, D. K., (1977). Selecting optimal treatment in clinical trials using covariate information. *Journal of Chronic Diseases*. **30**, 445 – 459.

Chow, S-C and Liu, J-P. (2000). Preface. *Statistics in Medicine*. **19**, 2719.

Cohen, J. (1969) Statistical Power Analysis for the Behavioral Sciences. (pp. 9 –12). Academic Press, Inc. New York.

Cox, D. R., *The Planning of Experiments*. Wiley: New York, 1958, 19.

Cox, D. R., (1992). Causality: Some statistical aspects. *Journal of the Royal Statistical Society. Series A.* **155**, 291 – 301.

Darwin, C. (1871), *On the Origin of Species* (5th ed.), New York: D. Appleton and Company, p. 26.

Dawid, A. P. (2000). "Causal inference without counterfactuals." *Journal of the American Statistical Association*. **95**, 407 – 424.

Elmore, R.T., Hettmansperger, T.P., Xuan, F. (2006). "A fully nonparametric test for one-way layouts." *Australia New Zealand Journal of Statistics*. **48**(4), 477 – 490.

Endrenyi, L. and Tothfolusi, L. (1999). Subject-by-formulation interaction in determinations of individual bioequivalence: bias and prevalence. *Pharmaceutical Research*. 16, 186 – 190.

Fang, C., Tseng, M., Daly, M.B. (2005). "Correlates of soy food consumption in women at increased risk for breast cancer," *Journal of American Diet Association*. 105, 1552 – 1558.

FTC Statement. Retrieved on Sept. 10, 2010 from http://www.ftc.gov/bcp/edu/pubs/business/adv/bus60.pdf.

Gadbury, G. L. and Iyer, H. K. (2000). Unit-Treatment interaction and its practical consequences. *Biometrics*. **56**, 882 – 885.

Gadbury, G. (2001). Randomization inference and bias of standard errors. Journal of the American Statistical Association. **55**, 1 – 4.

Gadbury, G. L., Iyer, H. K., and Allison, D. B. (2001). Evaluating subject-treatment interaction when comparing two treatments. *Journal of Biopharmaceutical Statistics*. **11**, 313 – 333.

Gadbury, G.L. (2004). Subject-Treatment Interaction. In *Encyclopedia of Biopharmaceutical Statistics, Second Edition, Revised and Expanded*. Edited by Shein-Chung Chow. Marcel Dekker, Inc., New York. On-line published 03/09/2004. 1-7.

Gail, M. and Simon, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. **41**, 361 – 372.

Gastwirth, J. L. (1975). Statistical measures of earnings differentials. *The American Statistician*. **29**, 32 – 35.

Hauck, W. W., Hyslop, T., Mei-Ling, C., Patnaik, R., Williams, R. L. (2000). Subject-by-formulation interaction in bioequivalence: Conceptual and statistical issues. *Pharmaceutical Research*. **17**, 375 – 380.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*. **81**, 945 – 960.

Hwang, S., Huber, P. B., Hesney, M., Kwan, K. C. (1978). Bioequivalence and Interchangeability. *Journal of Pharmaceutical Sciences*. **67**, Introduction.

Inman, H. F. and Bradley, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics: Theory and Methods*. **18**, 3851 – 3874.

Jones, M. C., Marron, J. S., Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*. 91, 401 – 407.

Kemp, K.E., Yang, S.S., Perng, S.K., and Nelson, P.I. (1993). "An asymptotically distribution free test for assessing the separation between two distributions." *Nonparametric Statistics*. **2**, pp. 235 – 248.

Kent, D. M. and Hayward, R. A., (2009). Limitations of applying summary results of clinical trials to individual patients: The need for risk stratification. *Journal of the American Medical Association*. **298**, 1209 – 1212.

Lader, M. Retrieved March 1, 2011 from http://www.news-medical.net/health/Sedative-Abuse.aspx

Lewis, K. D. and Burton-Freeman, B. M., (2010). The role of innovation and technology in meeting individual nutritional needs. *The Journal of Nutrition*. **140**, 426S – 436S.

Li, J. and Chan, I. S. F. (2006). Detecting qualitative interactions in clinical trials: An extension of range test. Journal of Biopharmaceutical Statistics, **16**, 831 – 841.

Lizotte, D., Bowling, M., Murphy, S.A. (2010). "Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis." *Proceedings of the 27th International Conference on Machine Learning (ICML).* 695 – 702.

Lucretius. Retrieved Sept. 9, 2010 from http://classics.mit.edu/Carus/nature_things.4.iv.html

Longford, N. T. (1999). Selection bias and treatment heterogeneity in clinical trials. *Statistics in Medicine. **18**, 1467 – 1474.*

Marshall, A. (1997). Laying the foundations for personalized medicines. *Nature Biotechnology*. **15**, 954 – 957.

Marchant, G.E. (2001), "Genetics and Toxic Torts," *Seton Hall Law Review,* 31, 949.

Marchant, G.E. (2010). Retrieved on Sept. 11, 2010 from http://www.law.asu.edu/files/Programs/Sci-Tech/Commentaries/Marchant%20Formatted.rev.doc

Milliken, G. A. and Johnson, D. E. (1984). *Analysis of Messy Data*. Lifetime Learning Publications. Belmont, CA. p. 113.

Neyman, J., Dabrowska, D. M., Speed, T. P. (1990). On the application of probability theory to agricultural experiments. Essay on Principles. Section 9. *Statistical Science*. 5, 465 – 472.

O'Hagan, A., Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*. 63(1), 201 – 203.

Perrett, J, Higgins, J. (2004). "Using prior information on the intraclass correlation coefficient to analyze data from unreplicated and under-replicated experiments." Dissertation, Kansas State University.

Peto, R. (1982). Statistical aspects of cancer trials. Chapman and Hall. 867 – 871.

Poulson, R. S., Gadbury, G. L., Allison, D. B. (2011). "Treatment heterogeneity and individual crossover interaction." *The American Statistician* (under revision).

Rom D. M. and Hwang, E. (1996). Testing for individual and population equivalence based on the proportion of similar responses. *Statistics in Medicine*. **15**, 1489 – 1505.

Roses, A. D. (2000). Pharmacokinetics and the practice of medicine. *Nature*. **405**, 857 – 865.

Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*. **70**(1), 41 – 55.

Rubin, D. B. (1974). Estimating causal effects for treatments in randomized and nonrandomized studies. *Journal of Educational Psychology.* **66**, 688 – 701.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*. **6**, 34 – 58.

Senn, S. (1997). Letter to the editor. *Statistics in Medicine*. **16**, 1301 – 1306.

Senn, S, (2001). Individual therapy: New dawn or false dawn? *Drug Information Journal*. **35**, 1479 – 1494.

Senn, S. (2006a), "Cross-over Trials in *Statistics in Medicine*: The First '25' Years," *Statistics in Medicine*, 25, 3430–3442.

Senn, S. (2006b). Letter to the editor. *Statistics in Medicine*. **25**, 3944 – 3948.

Setchell, K.D., Zimmer-Nechemias, L., Cai, J, and Heubi, J.E. (1998) "Isoflavone content of infant formulas and the metabolic fate of these early phytoestrogens in early life," *American Journal of Clinical Nutrition*, Supplement, **68**, 1453S-1461S.

Sharma, A.M. Retrieved July 10, 2010 from http://www.drsharma.ca/why-in-obesity-treatment-averages-are-not-good-enough.html

Sheather, S. J. (2004). Density estimation. *Statistical Science*. 19, 588 – 597.

Simon, R. (1982). Patient subsets and variation in therapeutic efficacy. *British Journal of Clinical Pharmacology*. **14**, 465 – 481.

Stine, R. A. and Heyse, J. F. (2001). Non-parametric estimates of overlap. *Statistics in Medicine*. **20**, 215 – 236.

Silvapulle, M. J. (2001). Tests for qualitative interaction: exact critical values and robust tests. *Biometrics*. **57**, 1157 – 1165.

Smith, C.M. (2005), "Origin and Uses of Primum Non Nocere--Above All, Do No Harm!," Journal *of Clinical Pharmacology*, 45, 371–377.

Tilton, J.W. (1937). "The measurement of overlapping." *Journal of Educational Psychology*, 28, 656-662.

Yang, M., and Stufken, J. (2008), "Optimal and Efficient Crossover Designs for Comparing Test Treatments to a Control Treatment Under Various Models," *Journal of Statistical Planning and Inference*, 138, 278–285.

Zhao, Y.D., Dmitrienko, A., and Tamura, R. (2010). "Design and analysis considerations in clinical trials with a sensitive subpopulation." *American Statistical Association Statistics in Biopharmaceutical Research.* 2(1), 72 – 83.

Appendix A - Proofs and Derivations

Chapter 3 Results

Unless otherwise stated, the results for chapter 3 are derived assuming $(X \quad Y)^T$ is distributed bivariate normal with $\mu_X > \mu_Y$, $\sigma_X^2 = \sigma^2$, $\sigma_Y^2 = k^2\sigma^2$, and $k \geq 1$.

A.3.1: Equation (3.4)

$$\frac{(\mu_Y - k^2\mu_X) \pm \sqrt{k^2(\mu_X^2 + \mu_Y^2 - 2\mu_X\mu_Y) - k^2 2\sigma^2 \ln(k)(1-k^2)}}{(1-k^2)}$$

Proof:

$$f_X(x) = f_Y(x)$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma}(x-\mu_X)^2} = \frac{1}{\sqrt{2\pi k^2\sigma^2}} e^{-\frac{1}{2k\sigma}(x-\mu_Y)^2}$$

$$x^2(1-k^2) - x(2\mu_Y - k^2 2\mu_X) + \mu_Y^2 - k^2\mu_X^2 + 2\sigma^2 k^2 ln(k) = 0$$

Solving for x gives the desired result. ■

A.3.2: Second equation of Result 2

$$x_L < x_{-1} < \mu_X$$

Proof:

Part 1: Show $x_{-1} < \mu_X$.

Suppose $x_{-1} \geq \mu_X$, then

$$x_{-1} = \frac{\mu_Y + k\mu_X}{1+k} \geq \mu_X$$

$$\mu_Y + k\mu_X \geq \mu_X + k\mu_X$$

$\mu_Y \geq \mu_X$, a contradiction.

Part 2: Show $x_{-1} > x_L$.

From (3.4)

$$x_L = \frac{(\mu_Y - k^2\mu_X) + \sqrt{k^2(\mu_X^2 + \mu_Y^2 - 2\mu_X\mu_Y) - k^2 2\sigma^2 \ln(k)(1-k^2)}}{(1-k^2)}$$

$$= \frac{(\mu_Y - k^2\mu_X)}{(1-k^2)} + \frac{\sqrt{k^2(\mu_X^2 + \mu_Y^2 - 2\mu_X\mu_Y) - k^2 2\sigma^2 \ln(k)(1-k^2)}}{(1-k^2)}.$$

(a) Show

$$x_{-1} > \frac{(\mu_Y - k^2\mu_X)}{(1-k^2)}$$

Suppose

$$x_{-1} = \frac{\mu_Y + k\mu_X}{(1+k)} < \frac{\mu_Y - k^2\mu_X}{(1-k^2)}$$

$$(1-k^2)(\mu_Y + k\mu_X) < (1+k)(\mu_Y - k^2\mu_X)$$

$$\mu_Y + k\mu_X - k^2\mu_Y - k^3\mu_X < \mu_Y - k^2\mu_X + k\mu_Y - k^3\mu_X$$

$$\mu_X - k\,\mu_Y < \mu_Y - k\,\mu_X, \text{a contradition.}$$

(b) Since

$$\frac{\sqrt{k^2(\mu_X^2 + \mu_Y^2 - 2\mu_X\mu_Y) - k^2 2\sigma^2 \ln(k)(1-k^2)}}{(1-k^2)} < 0$$

when $k^2 > 1$, then

$$x_{-1} > \frac{(\mu_Y - k^2\mu_X)}{(1-k^2)} > \frac{(\mu_Y - k^2\mu_X)}{(1-k^2)} + \frac{\sqrt{k^2(\mu_X^2 + \mu_Y^2 - 2\mu_X\mu_Y) - k^2 2\sigma^2 \ln(k)(1-k^2)}}{(1-k^2)} = x_L. \blacksquare$$

A.3.3: Equation (3.19)

$$S_{D|Z}^2 = S_{X|Z}^2 + S_{Y|Z}^2 - 2R_{XY|Z}S_{X|Z}S_{Y|Z}$$

Derivation: Given $(3.18a, b)$

$$S_{D|Z}^2 = \frac{SSE_Z}{g(n-1)}$$

$$= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(e_{Xij} - e_{Yij})^2}{g(n-1)}$$

$$= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}^2 + \sum_{i=1}^{g}\sum_{j=1}^{n} e_{Yij}^2 - 2\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}\, e_{Yij}}{g(n-1)}$$

$$= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2 + \sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2 - 2\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})(Y_{ij} - \bar{Y}_{i.})}{g(n-1)}$$

$$= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2}{g(n-1)} + \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2}{g(n-1)}$$

$$-2\,\frac{\frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})(Y_{ij} - \bar{Y}_{i.})}{\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2 \sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2}} \times \sqrt{\frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2 \sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2}{1}}}{g(n-1)}$$

$$= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2}{g(n-1)} + \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2}{g(n-1)}$$

$$-2R_{XY|Z}\,\frac{\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2 \sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2}}{\sqrt{g(n-1)}\sqrt{g(n-1)}}$$

$$= S_{X|Z}^2 + S_{Y|Z}^2 - 2R_{XY|Z}S_{X|Z}S_{Y|Z}. \quad \blacksquare$$

A.3.4: Proposition 3.8

Given SSE is generated from the model $(3.18a)$ and $(\varepsilon_{Xij}, \varepsilon_{Yij})$ are distributed as in $(3.18b)$

$$\frac{ng-1}{g(n-1)}S_D^2 = S_{D|Z}^2 + \frac{n\sum_{i=1}^{g}\left(\bar{D}_{z_i}-\bar{D}\right)^2}{g(n-1)}$$

Proof:

$$S_D^2 = \frac{SSE}{g(n-1)} = \frac{SSE_Z+(SSE-SSE_Z)}{g(n-1)}$$

$$= S_{D|Z}^2 + \frac{\sum_{j=1}^{ng} e_{Xj}^2+\sum_{j=1}^{ng} e_{Yij}^2-2\sum_{j=1}^{ng} e_{Xj}e_{Yj}-\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}^2-\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Yij}^2+2\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}e_{Yij}}{g(n-1)}$$

$$= S_{D|Z}^2 + \frac{\sum_{j=1}^{ng} e_{Xj}^2-\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}^2+\sum_{j=1}^{ng} e_{Yj}^2-\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Yij}^2-2\sum_{j=1}^{ng} e_{Xj}^2e_{Yj}^2+2\sum_{i=1}^{g}\sum_{j=1}^{n} e_{Xij}^2e_{Yij}^2}{g(n-1)}$$

$$= S_{D|Z}^2 + \frac{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}-\bar{X}_{..})^2-\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}-\bar{X}_{i.})^2+\sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_{..})^2-\sum_{i=1}^{g}\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_{i.})^2}{g(n-1)}$$

$$- \frac{2n\sum_{i=1}^{g}(\bar{X}_{i.}-\bar{X}_{..})(\bar{Y}_{i.}-\bar{Y}_{..})}{g(n-1)}$$

$$= S_{D|Z}^2 + \frac{n\{\sum_{i=1}^{g}(\bar{X}_{i.}-\bar{X}_{..})^2+)\sum_{i=1}^{g}(\bar{Y}_{i.}-\bar{Y}_{..})^2-2\sum_{i=1}^{g}(\bar{X}_{i.}-\bar{X}_{..})(\bar{Y}_{i.}-\bar{Y}_{..})\}}{g(n-1)}$$

$$= S_{D|Z}^2 + \frac{n\{\sum_{i=1}^{g}[(\bar{X}_{i.}-\bar{X}_{..})-(\bar{Y}_{i.}-\bar{Y}_{..})]^2\}}{g(n-1)}$$

$$= S_{D|Z}^2 + \frac{n\sum_{i=1}^{g}\left(\bar{D}_{z_i}-\bar{D}\right)^2}{g(n-1)}. \quad \blacksquare$$

A.3.5: Equation (3.21)

The model from a balanced two-way ANOVA (given usual assumptions) with Treatment at two levels $h = 1,2$ (T and R) and treatment Z (a grouping variable) at $i = 1,2,\dots,g$ levels and $j = 1,\dots,\frac{n}{2}$, where both g and $\frac{n}{2}$ are greater than 1 may be given as $R_{hij} = \mu + T_h + Z_i + TZ_{hi} + \varepsilon_{hij}$, where R_{hij} is the response to individual j given treatment h from group i. The sum of squares portion (SSTZ) of an F- test for SI is a scalar of

$$\frac{\frac{n}{2}\sum_{i=1}^{g}[(\bar{x}_{i.}-\bar{y}_{i.})-(\bar{x}_{..}-\bar{y}_{..})]^2}{g\left(\frac{n}{2}-1\right)}$$

an estimate of

$$\frac{n\sum_{i=1}^{g}\left(\bar{D}_{z_i}-\bar{D}\right)^2}{g(n-1)}.$$

Proof:

Given SSTZ is the sums of squares due to the interaction term TZ_{hi}, then

$$\begin{aligned}
\text{SSTZ} &= \sum_{h=1}^{2}\sum_{i=1}^{g}\sum_{j=1}^{n/2}(\bar{r}_{hi.}-\bar{r}_{h..}-\bar{r}_{.i.}+\bar{r}_{...})^2 = \frac{n}{2}\sum_{h=1}^{2}\sum_{i=1}^{g}(\bar{r}_{hi.}-\bar{r}_{h..}-\bar{r}_{.i.}+\bar{r}_{...})^2 \\
&= \frac{n}{2}\sum_{i=1}^{g}\left(\bar{x}_{i.}-\bar{x}_{..}-\frac{\bar{x}_{i.}-\bar{y}_{i.}}{2}+\frac{\bar{x}_{..}+\bar{y}_{..}}{2}\right)^2+\frac{n}{2}\sum_{i=1}^{g}\left(\bar{y}_{i.}-\bar{y}_{..}-\frac{\bar{x}_{i.}-\bar{y}_{i.}}{2}+\frac{\bar{x}_{..}+\bar{y}_{..}}{2}\right)^2 \\
&= \frac{n}{8}\sum_{i=1}^{g}[(\bar{x}_{i.}-\bar{y}_{i.}-\bar{x}_{..}+\bar{y}_{..})^2+(\bar{y}_{i.}-\bar{x}_{i.}-\bar{y}_{..}+\bar{x}_{..})^2] \\
&= \frac{n}{4}\sum_{i=1}^{g}[(\bar{x}_{i.}-\bar{y}_{i.})-(\bar{x}_{..}-\bar{y}_{..})]^2 \quad \blacksquare
\end{aligned}$$

Chapter 4 Results

Unless otherwise stated, the results for chapter 4 are derived assuming X and Y are skew normally distributed with $\mu_X > \mu_Y$, $E(X) \geq E(Y)$, $\sigma_X^2 = \sigma^2$, $\sigma_Y^2 = k^2\sigma^2$, and $k > 1$.

A.4.1: Equation (4.8)

$$\alpha_D = \frac{\frac{\alpha_{X_J}(1-k\rho_{XY}) - \alpha_{Y_J}(k-\rho_{XY})}{\sqrt{1+k^2-2k\rho_{XY}}}}{\sqrt{1+\alpha_{X_J}^2\left[1-\frac{(1-k\rho_{XY})^2}{1+k^2-2k\rho_{XY}}\right]+2\alpha_{X_J}\alpha_{Y_J}\left[\rho_{XY}-\frac{(1-k\rho_{XY})(\rho_{XY}-k)}{1+k^2-2k\rho_{XY}}\right]+\alpha_{Y_J}^2\left[1-\frac{(\rho_{XY}-k)^2}{1+k^2-2k\rho_{XY}}\right]}}$$

Proof (partial):

$$\alpha_D = \frac{\sqrt{\Sigma_D}\Sigma_D^{-1}B_D^T\boldsymbol{\alpha}}{\sqrt{1+\boldsymbol{\alpha}^T(P-B_D^T\Sigma_D^{-1}B_D^T)\boldsymbol{\alpha}}}$$

Note that

$$\Sigma_D = \sigma_D^2 = \sigma^2(1+k^2-2k\rho_{XY})$$
$$B_D = \left(\sqrt{\Sigma_\sigma}\right)^{-1}\Sigma A_D$$
$$\Sigma_\sigma = \begin{pmatrix}\sigma & 0\\ 0 & \sigma k\end{pmatrix}$$
$$\Sigma = \begin{pmatrix}\sigma^2 & \sigma^2 k\rho_{XY}\\ \sigma^2 k\rho_{XY} & \sigma^2 k^2\end{pmatrix}$$

so that

$$B_D = \begin{pmatrix}\frac{1}{\sigma} & 0\\ 0 & \frac{1}{\sigma k}\end{pmatrix}\begin{pmatrix}\sigma^2 & \sigma^2 k\rho\\ \sigma^2 k\rho & \sigma^2 k^2\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix} = \begin{pmatrix}\sigma(1-k\rho)\\ \sigma(\rho-k)\end{pmatrix}$$

From here matrix algebra yields (4.8). ■

A.4.2: Equation 4.10

$$\alpha_D(k=1) = \frac{\left(\alpha_{X_J}-\alpha_{Y_J}\right)\left(\frac{1-\rho_{XY}}{2}\right)^{\frac{1}{2}}}{\sqrt{1+\left(\frac{1+\rho_{XY}}{2}\right)\left(\alpha_{X_J}+\alpha_{Y_J}\right)^2}}$$

Proof:

$$\alpha_D = \frac{\frac{\alpha_{X_J}(1-\rho_{XY}) - \alpha_{Y_J}(1-\rho_{XY})}{\sqrt{1+k^2-2k\rho_{XY}}}}{\sqrt{1+\alpha_{X_J}^2\left[1-\frac{(1-\rho_{XY})^2}{2-2\rho_{XY}}\right]+2\alpha_{X_J}\alpha_{Y_J}\left[\rho_{XY}-\frac{(1-\rho_{XY})(\rho_{XY}-1)}{2-2\rho_{XY}}\right]+\alpha_{Y_J}^2\left[1-\frac{(\rho_{XY}-1)^2}{2-2\rho_{XY}}\right]}}$$

$$\alpha_D(k=1)$$

$$= \frac{\frac{\alpha_{X_J}(1-\rho_{XY}) - \alpha_{Y_J}(1-\rho_{XY})}{\sqrt{2(1-\rho_{XY})}}}{\sqrt{1+\alpha_{X_J}^2\left[1-\frac{(1-\rho_{XY})^2}{2(1-\rho_{XY})}\right]+2\alpha_{X_J}\alpha_{Y_J}\left[\rho_{XY}-\frac{(1-\rho_{XY})(\rho_{XY}-1)}{2(1-\rho_{XY})}\right]+\alpha_{Y_J}^2\left[1-\frac{(\rho_{XY}-1)^2}{2(1-\rho_{XY})}\right]}}$$

$$= \frac{\frac{\left(\alpha_{X_J}-\alpha_{Y_J}\right)(1-\rho_{XY})}{\sqrt{1-\rho_{XY}}\sqrt{2}}}{\sqrt{1+\alpha_{X_J}^2\left[1-\frac{(1-\rho_{XY})}{2}\right]+2\alpha_{X_J}\alpha_{Y_J}\left[\rho_{XY}-\frac{(1-\rho_{XY})(\rho_{XY}-1)}{2(1-\rho_{XY})}\right]+\alpha_{Y_J}^2\left[1-\frac{(\rho_{XY}-1)^2}{2(1-\rho_{XY})}\right]}}$$

$$= \frac{\left(\alpha_{X_J}-\alpha_{Y_J}\right)\left(\frac{1-\rho_{XY}}{2}\right)^{\frac{1}{2}}}{\sqrt{1+\alpha_{X_J}^2\left[\frac{1+\rho_{XY}}{2}\right]+2\alpha_{X_J}\alpha_{Y_J}\left[\frac{1+\rho_{XY}}{2}\right]+\alpha_{Y_J}^2\left[\frac{1+\rho_{XY}}{2}\right]}}$$

$$= \frac{\left(\alpha_{X_J}-\alpha_{Y_J}\right)\left(\frac{1-\rho_{XY}}{2}\right)^{\frac{1}{2}}}{\sqrt{1+\left(\frac{1+\rho_{XY}}{2}\right)\left(\alpha_{X_J}+\alpha_{Y_J}\right)^2}} \cdot \blacksquare$$

A.4.3: Proposition 4.1

Given *X* and *Y* are distributed as skew normal, $\alpha_{X_J} \leq \alpha_{Y_J}$ and $k = 1$

$$\alpha_D(\rho_{XY}, k = 1) \text{ is strictly increasing on } -1 < \rho_{XY} < 1$$

Proof:

Let $-1 < \rho_{XY} + \varepsilon < 1$, where $\varepsilon > 0$, then

$$\alpha_D^{\rho_{XY}} = \frac{\left(\alpha_{X_J} - \alpha_{Y_J}\right)\left(\frac{1-\rho_{XY}}{2}\right)^{\frac{1}{2}}}{\sqrt{1 + \left(\frac{1+\rho_{XY}}{2}\right)\left(\alpha_{X_J} + \alpha_{Y_J}\right)^2}} < \alpha_D^{\rho_{XY}+\varepsilon} = \frac{\left(\alpha_{X_J} - \alpha_{Y_J}\right)\left(\frac{1-(\rho_{XY}+\epsilon)}{2}\right)^{\frac{1}{2}}}{\sqrt{1 + \left(\frac{1+(\rho_{XY}+\varepsilon)}{2}\right)\left(\alpha_{X_J} + \alpha_{Y_J}\right)^2}}.$$

over all $-1 < \rho_{XY} < 1$. This is true since in the numerator

$$\left(\frac{1-\rho_{XY}}{2}\right)^{\frac{1}{2}} > \left(\frac{1-(\rho_{XY}+\epsilon)}{2}\right)^{\frac{1}{2}}$$

over all $-1 < \rho_{XY} < 1$ because $0 > -\varepsilon$, and in the denominator

$$\sqrt{1 + \left(\frac{1+\rho_{XY}}{2}\right)\left(\alpha_{X_J} + \alpha_{Y_J}\right)^2} < \sqrt{1 + \left(\frac{1+(\rho_{XY}+\varepsilon)}{2}\right)\left(\alpha_{X_J} + \alpha_{Y_J}\right)^2}$$

over all $-1 < \rho_{XY} < 1$ because $0 < \varepsilon$. ■

A.4.4: Proposition 4.2

Given *X* and *Y* are distributed as skew normal, $\alpha_{X_J} \leq \alpha_{Y_J}$ and $k = 1$

$$\text{PIQI}^{\rho_{XY}}(k = 1) \text{ is strictly decreasing in } -1 < \rho_{XY} < 1$$

Proof:

Let $\varepsilon > 0$ and $-1 < \rho_{XY} + \varepsilon < 1$

$$\text{PIQI}^{\rho_{XY}}(k = 1) = \int_{-\infty}^{0} f_D(d; \mu_D, \sigma_D, \alpha_D, \rho_{XY}) dd$$

$$= \int_{-\infty}^{0} \frac{2}{\sigma_D^{\rho_{XY}}} \phi\left(\frac{d-\mu_D}{\sigma_D^{\rho_{XY}}}\right) \Phi\left(\alpha_D^{\rho_{XY}} \frac{d-\mu_D}{\sigma_D^{\rho_{XY}}}\right) dd$$

$$> \int_{-\infty}^{0} \frac{2}{\sigma_D^{\rho_{XY}+\varepsilon}} \phi\left(\frac{d-\mu_D}{\sigma_D^{\rho_{XY}+\varepsilon}}\right) \Phi\left(\alpha_D^{\rho_{XY}+\varepsilon} \frac{d-\mu_D}{\sigma_D^{\rho_{XY}+\varepsilon}}\right) dd$$

$$= \int_{-\infty}^{0} f_D(d; \mu_D, \sigma_D, \alpha_D, \rho_{XY}+\varepsilon) dd$$

$$= \text{PIQI}^{\rho_{XY}+\varepsilon}(k=1)$$

The inequality holds since,

$\sigma_D^{\rho_{XY}} > \sigma_D^{\rho_{XY}+\varepsilon}$ (by definition in Section 3)

$\alpha_D^{\rho_{XY}} < \alpha_D^{\rho_{XY}+\varepsilon}$ (by proposition 4.3),

each component of $f_D(d; \mu_D, \sigma_D, \alpha_D, \rho_{XY})$ (the integrand) is larger for $\rho_{XY}+\varepsilon$ over the range $(-\infty, 0)$. That is,

$$\frac{2}{\sigma_D^{\rho_{XY}}} < \frac{2}{\sigma_D^{\rho_{XY}+\varepsilon}}$$

$\phi\left(\frac{d-\mu_D^{\rho_{XY}}}{\sigma_D^{\rho_{XY}}}\right) < \phi\left(\frac{d-\mu_D^{\rho_{XY}+\varepsilon}}{\sigma_D^{\rho_{XY}+\varepsilon}}\right)$ over range $(-\infty, 0)$

$$\Phi\left(\alpha_D^{\rho_{XY}} \frac{d-\mu_D^{\rho_{XY}}}{\sigma_D^{\rho_{XY}}}\right) < \Phi\left(\alpha_D^{\rho_{XY}+\varepsilon} \frac{d-\mu_D^{\rho_{XY}+\varepsilon}}{\sigma_D^{\rho_{XY}+\varepsilon}}\right)$$

so that less area exists over the range $(-\infty, 0)$.

Consequently, $\text{PIQI}^{\rho_{XY}}(k=1) > \text{PIQI}^{\rho_{XY}+\varepsilon}(k=1)$ for $-1 < \rho_{XY} < 1$, $-1 < \rho_{XY}+\varepsilon < 1$, and $d \in (-\infty, 0)$. ■

A.4.5: Equation $(4.11a)$

$$\alpha_{X_M} = \frac{\alpha_{X_J} + \alpha_{Y_J}\rho_{XY}}{\sqrt{1+\alpha_{Y_J}^2(1-\rho_{XY}^2)}}$$

Proof (partial):

$$\alpha_{X_M} = \frac{\sqrt{\Sigma_X}\Sigma_X^{-1}B_X^T\boldsymbol{\alpha}}{\sqrt{1+\boldsymbol{\alpha}^T(P - B_X^T\Sigma_X^{-1}B_X^T)\boldsymbol{\alpha}}}$$

Note that

$$A = \begin{pmatrix}1\\0\end{pmatrix}$$

$$\Sigma_X = \sigma_X^2 = \sigma^2$$

$$B_X = \left(\sqrt{\Sigma_\sigma}\right)^{-1}\Sigma A_X = \frac{1}{\sigma}\begin{pmatrix}\sigma^2 & \sigma^2 k\rho_{XY}\\ \sigma^2 k\rho_{XY} & \sigma^2 k^2\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\sigma\\ \sigma k\rho\end{pmatrix}$$

so that the numerator is

$$(\sigma \quad \sigma k\rho)\begin{pmatrix}\alpha_{X_J}\\ \alpha_{Y_J}\end{pmatrix} = \alpha_{X_J} + \alpha_{Y_J}\rho_{XY}$$

and the denominator is

$$\left\{1 + (\alpha_{X_J} \quad \alpha_{Y_J})\left[\begin{pmatrix}1 & \rho_{XY}\\ \rho_{XY} & 1\end{pmatrix} - \begin{pmatrix}\sigma\\ \sigma\rho_{XY}\end{pmatrix}\frac{1}{\sigma^2}(\sigma \quad \sigma\rho_{XY})\right]\begin{pmatrix}\alpha_{X_J}\\ \alpha_{Y_J}\end{pmatrix}\right\}^{\frac{1}{2}} = \sqrt{1+\alpha_{Y_J}^2(1-\rho_{XY}^2)} \quad \blacksquare$$

The result for α_{Y_M} $(4.11b)$ can be found similarly.

A.4.6:

Given $k = 1$ and $\alpha_D > 0$,

α_{X_M} and α_{Y_M} are strictly increasing functions of ρ_{XY}.

Proof : Provided for α_{X_M} in the case for when $\alpha_{X_J}, \alpha_{Y_J} > 0$.

Let $\varepsilon > 0$ and $-1 < \rho_{XY} + \varepsilon < 1$. When $\alpha_D > 0$, $\alpha_{X_J} > \alpha_{Y_J}$. So from the numerator of $(4.11a)$ we have that

$$\alpha_{X_J} + \alpha_{Y_J}\rho_{XY} < \alpha_{X_J} + \alpha_{Y_J}(\rho_{XY} + \varepsilon)$$

and from the denominator that

$$\sqrt{1+\alpha_{Y_J}^2(1-\rho_{XY}^2)} > \sqrt{1+\alpha_{Y_J}^2(1-[\rho_{XY}^2+\varepsilon])}.$$

Therefore,

$$\frac{\alpha_{X_J} + \alpha_{Y_J}\rho_{XY}}{\sqrt{1+\alpha_{Y_J}^2(1-\rho_{XY}^2)}} < \frac{\alpha_{X_J} + \alpha_{Y_J}(\rho_{XY}+\varepsilon)}{\sqrt{1+\alpha_{Y_J}^2(1-[\rho_{XY}^2+\varepsilon])}}. \quad \blacksquare$$

Chapter 5 Results

A.5.1: Equation (5.10)

Given the constraints of (5.1), (5.3), $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and $k_1 = k_2 = 1$,

$$\mathrm{PSR}_J = P\left(X_1 \le \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) + P\left(X_1 > \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right)$$

$$+P\left(Y_1 > \frac{\mu_{X_1} + \mu_{Y_1}}{2}, Y_2 \ge \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) + P\left(Y_1 \le \frac{\mu_{X_1} + \mu_{Y_1}}{2}, Y_2 \ge \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right).$$

Proof:

Term 1:

$$P\left(X_1 \le \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) = P\left(X_1 \le \mu_{X_1}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right)$$

$$= P\left(Z_1 \le 0, Z_2 \le \frac{\frac{\mu_{X_2} + \mu_{Y_2}}{2} - \mu_{X_2}}{\sigma}\right) = P\left(Z_1 \le 0, Z_2 \le \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right)$$

Term 2:

$$P\left(X_1 > \frac{\mu_{X_1} + \mu_{Y_1}}{2}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) = P\left(X_1 > \mu_{X_1}, X_2 \le \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right)$$

$$= P\left(Z_1 > 0, Z_2 \le \frac{\frac{\mu_{X_2} + \mu_{Y_2}}{2} - \mu_{X_2}}{\sigma}\right) = P\left(Z_1 > 0, Z_2 \le \frac{\mu_{Y_2} - \mu_{X_2}}{2\sigma}\right)$$

Term 3:

$$P\left(Y_1 > \frac{\mu_{X_1} + \mu_{Y_1}}{2}, Y_2 \ge \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right) = P\left(Y_1 > \mu_{Y_1}, Y_2 \ge \frac{\mu_{X_2} + \mu_{Y_2}}{2}\right)$$

$$= P\left(Z_1 > 0, Z_2 \le \frac{\frac{\mu_{X_2} + \mu_{Y_2}}{2} - \mu_{Y_2}}{\sigma}\right) = P\left(Z_1 > 0, Z_2 \le \frac{\mu_{X_2} - \mu_{Y_2}}{2\sigma}\right)$$

Term 4:

$$P\left(Y_1 \leq \frac{\mu_{X_1}+\mu_{Y_1}}{2}, Y_2 \geq \frac{\mu_{X_2}+\mu_{Y_2}}{2}\right) = P\left(Y_1 \leq \mu_{Y_1}, Y_2 \geq \frac{\mu_{X_2}+\mu_{Y_2}}{2}\right)$$

$$= P\left(Z_1 \leq 0, Z_2 \leq \frac{\frac{\mu_{X_2}+\mu_{Y_2}}{2}-\mu_{Y_2}}{\sigma}\right) = P\left(Z_1 \leq 0, Z_2 \leq \frac{\mu_{X_2}-\mu_{Y_2}}{2\sigma}\right)$$

and since

$$P\left(Z_1 > 0, Z_2 \leq \frac{\mu_{Y_2}-\mu_{X_2}}{2\sigma}\right) = P\left(Z_1 \leq 0, Z_2 \leq \frac{\mu_{X_2}-\mu_{Y_2}}{2\sigma}\right),$$

the terms are equal. ■

Appendix B - R Programs

B.4.1: Equation (4.9) (Calculation of PIQI under the skew normal model)

```
 #Function for PIQI under skew normality
#Output: PIQI
#Name of function: f.piqi.sn
#Equation (4.9) to accomodate skew normality
#sn package is for skew normal

#Packages:
library(MASS)
library(sn)

###JOINT DISTRIBUTION SPECIFICATIONS
#Location
#mu=c(mu.x,mu.y)

#Scale
#sigma=diag(c(sigma.x,sigma.y))

#Shape
#alpha.j=matrix(c(alpha.x.j,alpha.y.j),nrow=2)

#Correlation
#rho

f.piqi.sn=function(mu,sigma,alpha.j,rho){
P=matrix(c(1,rho,rho,1),nrow=2)
Sigma=sigma%*%P%*%sigma

A.D=matrix(c(1,-1),nrow=2)  #Tha A matrix from pages 584-585 Azzalinni
        mu.D=t(A.D)%*%mu
        Sigma.D=t(A.D)%*%Sigma%*%A.D
        sigma.D=sqrt(Sigma.D)
        B.D=solve(sigma)%*%Sigma%*%A.D
        alpha.D.num=sigma.D*solve(Sigma.D)%*%t(B.D)%*%alpha.j
        alpha.D.den=sqrt(1+t(alpha.j)%*%(P-B.D%*%solve(Sigma.D)%*%
        t(B.D))%*%alpha.j)
        alpha.D=alpha.D.num/alpha.D.den

PIQI=psn(0,location=mu.D,scale=sigma.D,shape=alpha.D)
}
```

B.4.2: PSR and Crossing Points

```
#Function for PSR under skew normality
	#Output: PSR and the crossing points
	#Name of function: f.psr.sn
	#Equation (2.8) to accomodate skew normality
	#sn package is skew normal
	#E(X)<E(Y) for f.psr.sn to operate

	library(MASS)
	library(sn)

	###Function parameters
	#Location parameters
	#mu.x
	#mu.y

	#Scale parameters
	#sigma.x
	#sigma.y

	#Shape parameters
	#alpha.x.m
	#alpha.y.m

	#delta: specification of numerical accuracy for cross points
	#Range of x: x.lower, x.upper

	f.psr.sn<-function(mu.x,mu.y,sigma.x,sigma.y,alpha.x.m,alpha.y.m,delta,x.lower,x.upper){
	x.range=seq(x.lower,x.upper,delta)
	fX=dsn(x.range,location=mu.x,scale=sigma.x,shape=alpha.x.m)#desnity for X
	fY=dsn(x.range,location=mu.y,scale=sigma.y,shape=alpha.y.m)#density for Y

	#########FIND CROSSING POINTS xL , xI, and xU

	count.x=ifelse(fY>fX,1,0)
	count.y=ifelse(fY<fX,1,0)
	count.diff=count.x-count.y
	for(i in 1:length(count.diff)){
		ifelse(count.diff[i]==0,count.diff[i]<-1,count.diff[i]<-count.diff[i])
		}

	#LOOP FOR FINDING NUMBER OF CROSSPOINTS
	cross_points.vector=rep(0,length(count.diff))
	for(i in 1:length(count.diff-1)){
		ifelse(count.diff[i]==count.diff[i+1],cross_points.vector[i]<-0,
		cross_points.vector[i]<-1)
		}
	cross_points.number=sum(cross_points.vector)
```

```
#SOLUTION FOR ONE CROSSING POINT
if (cross_points.number==1){
        xL=x.range[sum(count.x)]
        xI=1000
        xU=1000

        prob.X.num=psn(xL,location=mu.x,scale=sigma.x,shape=alpha.x.m)
        prob.Y.num=1-psn(xL,location=mu.y,scale=sigma.y,shape=alpha.y.m)

        PSR=prob.X.num+prob.Y.num
        }

#SOLUTION FOR TW0 CROSSING POINTS
if ((cross_points.number==2)&(count.diff[1]<0)){ #K<1
        #split into 3 X segments
        count.seg.1=0
        seq.1=1
        while(count.diff[seq.1]<0)
                {count.seg.1<-count.seg.1+1
                seq.1<-seq.1+1
        }
        xL=x.range[count.seg.1]

        count.seg.2=0
        seq.2=1
        while(count.diff[count.seg.1+seq.2]>0)
                {count.seg.2<-count.seg.2+1
                seq.2<-seq.2+1
        }
        xU=x.range[count.seg.2+count.seg.1]
        xI=1000

        prob.Y.1.num=psn(xL,location=mu.y,scale=sigma.y,shape=alpha.y.m)
        prob.X.2.num=psn(xU,location=mu.x,scale=sigma.x,shape=alpha.x.m)-
                        psn(xL,location=mu.x,scale=sigma.x,shape=alpha.x.m)
        prob.Y.3.num=1-psn(xU,location=mu.y,scale=sigma.y,shape=alpha.y.m)

        PSR=prob.Y.1.num+prob.X.2.num+prob.Y.3.num
        }

if ((cross_points.number==2)&(count.diff[1]>0)){ #K>1
        #split into 3 X segments
        count.seg.1=0
        seq.1=1
        while(count.diff[seq.1]>0)
                {count.seg.1<-count.seg.1+1
                seq.1<-seq.1+1
        }
        xL=x.range[count.seg.1]
        count.seg.2=0
```

```
        seq.2=1
        while(count.diff[count.seg.1+seq.2]<0)
                {count.seg.2<-count.seg.2+1
                seq.2<-seq.2+1
        }
        xU=x.range[count.seg.2+count.seg.1]
        xI=1000
        prob.X.1.num=psn(xL,location=mu.x,scale=sigma.x,shape=alpha.x.m)
        prob.Y.2.num=psn(xU,location=mu.y,scale=sigma.y,shape=alpha.y.m)-
                        psn(xL,location=mu.y,scale=sigma.y,shape=alpha.y.m)
        prob.X.3.num=1-psn(xU,location=mu.x,scale=sigma.x,shape=alpha.x.m)

        PSR=prob.X.1.num+prob.Y.2.num+prob.X.3.num
        }

#SOLUTION FOR THREE CROSSING POINTS
if (cross_points.number==3){ #ALPHA.X.M=inverse of ALPHA.Y.M
        #split into 4 X segments
        count.seg.1=0
        seq.1=1
        while(count.diff[seq.1]<0)
                {count.seg.1<-count.seg.1+1
                seq.1<-seq.1+1
        }
        xL=x.range[count.seg.1]
        count.seg.2=0
        seq.2=1
        while(count.diff[count.seg.1+seq.2]>0)
                {count.seg.2<-count.seg.2+1
                seq.2<-seq.2+1
        }
        xI=x.range[count.seg.2+count.seg.1]
        count.seg.3=0
        seq.3=1
        while(count.diff[count.seg.1+count.seg.2+seq.3]<0)
                {count.seg.3<-count.seg.3+1
                seq.3<-seq.3+1
        }
        xU=x.range[count.seg.3+count.seg.2+count.seg.1]
                prob.Y.1.num=psn(xL,location=mu.y,scale=sigma.y,shape=alpha.y.m)
        prob.X.2.num=psn(xI,location=mu.x,scale=sigma.x,shape=alpha.x.m)-
                        psn(xL,location=mu.x,scale=sigma.x,shape=alpha.x.m)
        prob.Y.3.num=psn(xU,location=mu.y,scale=sigma.y,shape=alpha.y.m)-
                        psn(xI,location=mu.y,scale=sigma.y,shape=alpha.y.m)
        prob.X.4.num=1-psn(xU,location=mu.x,scale=sigma.x,shape=alpha.x.m)
        PSR=prob.Y.1.num+prob.X.2.num+prob.Y.3.num+prob.X.4.num
        }
f.psr.ns_out=return(PSR,xL,xI,xU)
}
```

B.4.3: Kernel PSR calculation

```
#Kernel PSR Calculation
#Equation (2.13)
library(sn)
library(KernSmooth)

##Parameter Specifications
#Location parameters
#mu.x
#mu.y
#Scale parameters
#sigma.x
#sigma.y
#Shape parameters
#alpha.x.m
#alpha.y.m

#SIMULATION TO PRODUCE MARGINALS

n=500
X<-rsn(n,mu.X,sigma.X,alpha.x.m)
Y<-rsn(n,mu.Y,sigma.Y,alpha.y.m)

hx=bw.SJ(X,method="dpi") #Sheather Jones plug-in bandwidth
hy=bw.SJ(Y,method="dpi") #Sheather Jones plug-in bandwidth
#hx<-1.06*sqrt(var(X))*n^(-1/5) #Proposed by Stine and Heyse
#hy<-1.06*sqrt(var(Y))*n^(-1/5) #Proposed by Stine and Heyse
kde.X<-bkde(X,bandwidth=hx,range.x=c(min(X,Y),max(X,Y)))
kde.X.x<-kde.X$x
kde.X.y<-kde.X$y
kde.Y<-bkde(Y,bandwidth=hy,range.x=c(min(X,Y),max(X,Y)))
kde.Y.x<-kde.Y$x
kde.Y.y<-kde.Y$y
plot(kde.X,type="l",lwd=2,axes=FALSE)
lines(kde.Y,lwd=2,lty=2)

l<-length(kde.X.x)
d<-kde.X.x[2]-kde.X.x[1]

d.psr<-numeric(l)
for(i in 1:l){d.psr[i]<-min(kde.X.y[i],kde.Y.y[i])*d
        }
PSR_k=sum(d.psr)
```