NEW METHODS IN COMMON FACTOR MODELING AND EXPERIMETRICS

by

Jason Alexander Parker

APPROVED BY SUPERVISORY COMMITTEE:

Donggyu Sul, Chair

Rodney J. Andrews

Gary Bolton

James C. Murdoch

This dissertation is dedicated to my parents

who, despite my twenty-six years of mooching,

have never once failed to support me.

NEW METHODS IN COMMON FACTOR MODELING AND EXPERIMETRICS

by

JASON ALEXANDER PARKER, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ECONOMICS

THE UNIVERSITY OF TEXAS AT DALLAS

August 2014

UMI Number: 3634473

UMI

Dissertation Publishing

UMI 3634473

ProQuest

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

www.manaraa.com

ACKNOWLEDGMENTS

NEW METHODS IN COMMON FACTOR MODELING AND EXPERIMETRICS

Publication No. _____

Jason Alexander Parker, PhD
The University of Texas at Dallas, 2014

Supervising Professor: Donggyu Sul

The first chapter, Identification of Unknown Factors: Leaders and Followers, concerns finding an economic interpretation of an underlying factor model. Previously, factors have been considered an unimportant nuisance parameter in the data. With factor identification, factors can be economically interpreted, creating more clear descriptions of the underlying patterns in the data. The authors also propose a leadership model where one individual in the panel is the common factor, the source of cross-sectional dependence (CSD) in the data.

In the second chapter, Heavy Traffic: Determination of Homicide Rates across the 50 United States, the author uses these new methods to analyze state-level crime rates in the U.S. from 1971 to 2011. Few papers have previously explored CSD between states in the crime data. While property crimes may be influenced by the common U.S. business cycle, it is somewhat puzzling that murder rates exhibit CSD between states. Using the leadership method detailed in the first chapter, the source of CSD in the data is found to be equivalent to the murder rates in California and Texas. The reason for this effect is likely to be related to drug-trafficking from Mexico through these key states and into the U.S. as a whole.

In the third chapter, Improved Two-Sample Comparisons for Laboratory Data, the author proposes a new approach to comparing the distributions of randomly drawn, independent samples. The Wilcoxon-Mann-Whitney (WMW) rank sum test is criticized for its complicated hypothesis and for having low power while laboratory data is quite costly. In contrast to this nonparametric approach, the author proposes a parametric test using the beta distribution to gain a significant power advantage. Further, the author introduces the idea of comparing higher-moments, e.g., skewness, of the respective distributions to find differences which first-moment comparisons will miss. Altogether, this dissertation provides new methods for understanding economic and laboratory data in a more meaningful way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

IDENTIFICATION OF UNKNOWN COMMON FACTORS:
LEADERS AND FOLLOWERS

Authors – Jason Alexander Parker, Donggyu Sul

The Department of Economics, GR31

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

PREFACE

The following paper is an equal contribution from Jason Parker and Donggyu Sul. Donggyu Sul is responsible for the original ideas of identification through factor number estimation and dominant leadership modeling. These ideas and the others were developed by both authors throughout the research process. Donggyu Sul wrote the introduction to the paper. Both authors were involved in the rest of the writing. The proofs, simulations, and examples were composed by Jason Parker.

2

ABSTRACT

First, this paper develops a new criterion for testing if a particular time series variable is a common factor in the conventional approximate factor model. Second, by modeling observed factors as a set of potential factors to be tested, this paper reveals how to easily pin down the factor without performing a large number of tests. This allows the researcher to check whether or not each individual in the panel is the underlying common factor. When multiple individuals appear to be factors, identification becomes confounded. A clustering mechanism is designed to identify the underlying factors in such a situation. Asymptotically, the developed procedure correctly identifies the factor when N and T jointly approach infinity under the minimal assumptions of Bai and Ng (2002). The procedure is shown to be quite effective in the finite sample by means of Monte Carlo simulation. The procedure is then applied to three empirical examples, demonstrating that the newly-developed method identifies the unknown common factors accurately.

3

1.1 Introduction

In the last two decades, there has been rapid development in analyzing cross-sectional dependence by using the approximate common factor structure. Among many others, Ahn and Horenstein (2013), Amengual and Watson (2007), Bai and Ng (2002), Hallin and Liska (2007), Harding (2013), Kapetanios (2010), Onatski (2009), and Perez and Ahn (2007) suggest consistent estimation procedures for the number of common factors, while Bai (2003; 2004), Bates, Plagborg-Møller, Stock, and Watson (2013), Choi (2012), Forni, Hallin, Lippi, and Reichlin (2000, 2005), and Stock and Watson (2002a; 2002b) propose consistent estimators for the common factors.

However, the most thorny challenge in this literature is the identification of these unknown common factors. Without identification, a common factor model of an economic phenomenon is fundamentally incomplete. Presently, empirical researchers have two general identification strategies. First, some researchers are forced to settle for simply describing the factors using their shape, correlation to observed series, and factor loadings (e.g., Ludwigson and Ng 2007; Reis and Watson 2010). The problem with this approach is that the factor is only described, not pinned down. Sometimes researchers name the factor, but that name is completely arbitrary. The other approach is directly to compare a $(m \times 1)$ vector of potential true factors P_t with the $(r \times 1)$ vector of unknown statistical factors F_t. Of course, the true statistical factor F_t is not observable, so Bai and Ng (2006) propose several tests to check whether or not a linear combination among the principal component estimators of F_t becomes identical to one or a linear combination of the potential factors P_t. Their idea was novel but unfortunately the finite sample performance is rather disappointing. As we will show later but the Bai and Ng (2006)'s tests suffer from serious size distortion even with large N and T (the dimensions of the panel). Another problem of the Bai and Ng (2006)'s tests is that even when P_t were identified as one of the common factors, there was no suggested clustering algorithm to identify which component of P_t is matching with which component

of F_t. The solution to this problem does not seem to exist until the estimated factors are identical to the true statistical factors. Along with this line, Bai and Ng (2013) provide three sets of restrictions to make the principal component (PC hereafter) estimators, $\hat{F}_t$, be the true statistical factors, F_t.

The purpose of this paper is to provide a novel and intuitive approach to identify whether or not an observed time series is asymptotically equal to an unobserved true factor. The newly suggested identification strategy does not require any identification restrictions for the PC estimators. The underlying logic is based on the notion of an 'asymptotically weak factor'. When a panel data has only asymptotically weak factors, the true number of common factors to the panel data becomes zero as both N and T go to infinity. For example, the PC estimates of the idiosyncratic components are asymptotically weak factors. Obviously, conventional factor number estimation such as Bai and Ng (2002) or Hallin and Liska (2007) will estimate a factor number of zero with panel data which only has asymptotically weak factors. We are utilizing this principle to identify whether or not a vector of potentially true factors, P_t, are indeed one of the true statistical factors, F_t. Let P_{jt} be the jth element of P_t. Then, it is easy to show that the regression residuals from the regression of one of the potential factors and any $(r-1)$ vector of the estimated common factors have only asymptotically weak factors, so that the conventional factor number estimators can be used to examine whether or not a potential factor is the true common factor. Of course, if P_{jt} is not a true factor, then the regression residuals must have at least one strong factor. This simple but novel idea does not require any identification restrictions either on the PC estimators or a rotation matrix H.

While the newly-developed method can be used to test exogenous time series, this paper models the factor as potentially being one particular individual which appears in the panel. When one individual is exactly equal to the factor, we call this individual a 'dominant leader.' If the factor is not a leader in the finite sample, but becomes a leader as N and T go to infinity, the individual is called an 'approximate dominant leader.'

This leadership model has a powerful interpretation: one or more individuals become the source(s) of the cross-sectional dependence in the factor model. Consider the following hypothetical example of leadership: In industrial organization, one or a few dominant firms can set a price for their product, and the rest of the market more or less adopts that price. This type of causal relationship can be observed in many areas including social, agricultural, and behavioral sciences. In natural science, earthquakes and the spread of viruses are potential examples of this pattern. In such situations, a few individuals or locations become leaders or sources of epidemic events. Therefore, an important task is to identify the leaders from a set of individuals.

Since the factor identification strategy above must be performed separately for each individual, there could be some failure probability when N, the number of individuals in the panel, is large. To control this probability, we provide a method based on ranking R^2 values from regressions of the estimated PC factors on each individual time series separately. Individuals with high R^2 are considered 'leader candidates' to be potentially identified as leaders.

If multiple leaders are found in the data, some individuals could be approximate dominant leaders for the same factor. Accordingly, we provide a clustering procedure to pin down which individual is related to which factor. This clustering is based on observing correlations and checking the residual factor number from including multiple individuals as regressors.

It is worth mentioning that two papers have already used our identification strategy. Gaibulloev, Sandler and Sul (2012) find that Lebanon is the main determinant of transnational terrorism. Greenaway-McGrevy, Mark, Sul and Wu (2012) (henceforth GMSW) utilize our method to find three key currencies as the main determinants for local exchange rates.

The remainder of the paper is organized as follows. Section 1.2 provides information about the setting as well as the definition of weak factors. Section 1.3 discusses leadership modeling and testing. Detailed asymptotic analyses are also provided. Section 1.4 demonstrates the finite sample performance of our test and also compares our results with Bai and

Ng (2006). Section 1.5 provides three empirical examples to show the effectiveness of our test. Section 1.6 concludes. Mathematical proofs are provided in Appendix A. Gauss code for the procedures as well as extra Monte Carlo simulations are available at the author's website.

1.2 Preliminary

Before we proceed, we define the variables that are used in the paper. y_{it} is the panel data of interest where the cross-sectional dependence can be expressed in a static common factor representation. G_t is the $r \times 1$ vector of true common factors, F_t is the $r \times 1$ vector of true static common factors. See (1.2) below for the difference between G_t and F_t. $\hat{F}_t$ is the $r \times 1$ vector of the principal component estimator, H is the $r \times r$ rotating matrix. $\#(y_{it})$ is the true number of common factors of y_{it} and $\hat{\#}(y_{it})$ is the estimated number of common factors of y_{it}. y_{it}^o is the idiosyncratic component to y_{it}.

To provide an intuitive explanation of how factor number estimation can be used to identify the true factors, we consider the following static factor structure with two factors $(r = 2)$ as an example.

$$y_{it} = \alpha_{1i}G_{1t} + \alpha_{2i}G_{2t} + y_{it}^o, \tag{1.1}$$

where α_{ji} is the true factor loading coefficient for the ith individual and to the jth factor. Here we exclude any non-zero constant terms for notational simplicity. The inclusion of a constant term does not change the result at all. Further assume that the two true factors are correlated with each other. Let $G_t \sim d(0, \Omega)$, then there exists a unique Cholesky decomposition such that

$$\begin{bmatrix} G_{1t} \\ G_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} F_{1t} \\ F_{2t} \end{bmatrix} = AF_t, \tag{1.2}$$

where F_{1t} is independent from F_{2t} but both F_{1t} and F_{2t} have the unitary variance. That is, $F_t \sim d\left(0, I_2\right)$. By using (1.2), (1.1) can be rewritten as

$$y_{it} = \lambda_{1i}F_{1t} + \lambda_{2i}F_{2t} + y_{it}^o, \tag{1.3}$$

where $\lambda_{1i} = \alpha_{1i}a_{11} + \alpha_{2i}a_{21}$, and $\lambda_{2i} = \alpha_{2i}a_{22}$. Following GMSW, we call G_t the empirical true factors while F_t are called the statistical true factors. The number of factors in y_{it}^o is naturally zero, independent of how the common factors are defined (empirically or statistically). Interestingly, the estimate of y_{it}^o – the panel of regression residuals of $\hat{y}_{it}^o$ from running y_{it} on either G_t or F_t – does not include any significant common factor either as long as the least squares estimator for the factor loading coefficients is consistent. That is,

$$\hat{y}_{it}^o = y_{it} - \hat{\alpha}_{1i}G_{1t} - \hat{\alpha}_{2i}G_{2t} = y_{it} - \hat{\lambda}_{1i}F_{1t} - \hat{\lambda}_{2i}F_{2t} \tag{1.4}$$

$$= y_{it}^o + (\alpha_{1i} - \hat{\alpha}_{1i})G_{1t} + (\alpha_{2i} - \hat{\alpha}_{2i})G_{2t} = y_{it}^o + O_p\left(T^{-1/2}\right). \tag{1.5}$$

The first equality holds exactly due to (1.2). Even though $\hat{y}_{it}^o$ has two common factors in the finite sample, asymptotically $\hat{y}_{it}^o$ does not have any common factors since the common components vanish asymptotically. We call such factors 'asymptotically weak factors.' Let x_{it}^o be the random variables which satisfy Bai and Ng (2002)'s Assumption C for the idiosyncratic components. Define $x_{it} = \psi_i'Z_t + x_{it}^o$ where ψ_i and Z_t are factor loadings and common factors to x_{it}, respectively. Then formally, the asymptotically weak factor can be defined as

Definition: (Asymptotically Weak Factors) *x_{it} has asymptotically weak factors if and only if $\psi_i Z_t = O_p\left(C_{NT}^{-1}\right)$ where $C_{NT} = \min\left[\sqrt{N}, \sqrt{T}\right]$.*

Note that Chudik and Pesaran (2013) use the terminology of 'weak factor' to define the cross-sectionally weak dependence where the common factor is $O_p(1)$ but the factor loadings are $O_p\left(N^{-1/2}\right)$. Hence, the notion of asymptotically weak factors used in this paper is weaker

than the concept of 'weak factor.' Next, the following lemma can be directly established. Recall that in the beginning of this section we defined $\#\left(x_{it}\right)$ as the true factor number to x_{it} and $\hat{\#}\left(x_{it}\right)$ as the estimator for the factor number of x_{it}.

Lemma 1 (Asymptotic Factor Number for Weak Factors) *As $N, T \to \infty$ jointly,*

$$\lim_{N,T\to\infty} \Pr\left[\hat{\#}\left(x_{it}\right) = 0\right] = 1. \tag{1.6}$$

See Appendix A for the proof. Intuitively, if x_{it} has only asymptotically weak factors, then asymptotically the cross-sectional dependence among x_{it} becomes equivalent to that among x_{it}^o which leads to (1.6). According to Lemma 1, it becomes obvious that $\Pr\left[\hat{\#}\left(\hat{y}_{it}^o\right) = 0\right] \to 1$ as $N, T \to \infty$. Hence if G_t are true factors to y_{it}, then the regression residuals, $\hat{y}_{it}^o$, should not have any strong factors. However the opposite is not true in general. Consider a variable W_t which is not correlated with y_{it} at all. Then as long as W_t is included as a regressor with G_t together, the new regression residuals will only have asymptotically weak factors. That is, consider the following regression:

$$y_{it} = \alpha_{1i}G_{1t} + \alpha_{2i}G_{2t} + \alpha_{3i}W_t + y_{it}^o,$$

and define the new residuals as

$$\hat{y}_{it}^o = y_{it} - \hat{\alpha}_{1i}G_{1t} + \hat{\alpha}_{2i}G_{2t} + \hat{\alpha}_{3i}W_t = y_{it}^o + o_p\left(1\right).$$

Since $\hat{\alpha}_{3i} \to^p 0$ as $T \to \infty$, W_t becomes an asymptotically weak factor to y_{it}^o. But it does not mean that W_t is a true factor to y_{it}.

Hence naturally, our interest becomes identifying potential variables one at a time. If the statistic factors were known, such identification could be achieved. Note that F_t is not observable but can be estimated by $H'^{-1}\hat{F}_t$ where H is an invertible 2×2 rotation matrix (in this example). Rewrite (1.3) as

$$y_{it} = \hat{\lambda}_i'\hat{F}_t + \hat{y}_{it}^o, \tag{1.7}$$

where the new residual, $\hat{y}_{it}^o$, is defined as

$$\hat{y}_{it}^o = y_{it}^o - \left(\hat{\lambda}_i' - \lambda_i' H'^{-1}\right)\hat{F}_t - \lambda_i' H'^{-1}\left(\hat{F}_t - H'F_t\right). \tag{1.8}$$

It is well known that under suitable conditions given in Bai (2003), $\hat{\lambda}_i' - \lambda_i' H'^{-1} = O_p\left(T^{-1/2}\right)$ and $\left(\hat{F}_t - H'F_t\right) = O_p\left(N^{-1/2}\right)$. In other words, the new regression residuals of $\hat{y}_{it}^o$ in (1.8) also have only asymptotically weak factors.

The true common factors can be rewritten as

$$G_t = AF_t = AH'^{-1}\hat{F}_t + AH'^{-1}\left(H'F_t - \hat{F}_t\right) = B\hat{F}_t + B\varepsilon_t, \tag{1.9}$$

where $B = AH'^{-1}$ and $\varepsilon_{it} = H'F_t - \hat{F}_t = O_p\left(N^{-1/2}\right)$. To be specific, following the example considered above, we rewrite (1.9) as

$$\begin{bmatrix} G_{1t} \\ G_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} \hat{F}_{1t} \\ \hat{F}_{2t} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{1.10}$$

where $\varepsilon_{1t} = O_p\left(N^{-1}\right)$ and $\varepsilon_{2t} = O_p\left(N^{-1}\right)$. Next, consider the following regression

$$y_{it} = \beta_{1i}G_{1t} + \beta_{2i}\hat{F}_{2t} + u_{it}. \tag{1.11}$$

The regression coefficient and error can be derived by combining (1.10) with (1.7): $\beta_{1i} = \hat{\lambda}_{1i}/b_{11}$, $\beta_{2i} = \hat{\lambda}_{2i} - \hat{\lambda}_{1i}b_{12}/b_{11}$, and $u_{it} = \hat{y}_{it}^o - \left(\hat{\lambda}_{1i}/b_{11}\right)\varepsilon_{1t}$ where $\hat{y}_{it}^o$ is defined in (1.8). Note that the LS estimator $\hat{\beta}_{1i}$ becomes biased due to the temporal correlation between G_{1t} and ε_{1t} but is consistent since $T^{-1}\sum_{t=1}^T \varepsilon_{1t}^2 = O_p\left(N^{-1}\right)$. The regression residuals can be written as

$$\hat{u}_{it} = u_{it} + \left(\beta_{1i} - \hat{\beta}_{1i}\right)G_{1t} + \left(\beta_{2i} - \hat{\beta}_{2i}\right)\hat{F}_{2t}.$$

Since $\hat{u}_{it}$ has only asymptotically weak factors, it is obvious that $\lim_{N,T\to\infty}\Pr\left[\hat{\#}\left(\hat{u}_{it}\right) = 0\right] = 1$. For G_{2t}, similar to (1.11), one can switch $\hat{F}_{2t}$ with $\hat{F}_{1t}$ and G_{1t} with G_{2t}. Then the regression residuals should not have any strong factors. It is important to note that in the place of $\hat{F}_{2t}$

in (1.11), one may consider using $\hat{F}_{1t}$ instead. As long as $b_{12} \neq 0$, G_{1t} can be identified by including $\hat{F}_{1t}$ rather than $\hat{F}_{2t}$ in (1.11).

Next, consider a time series variable L_t such that $L_t = \delta G_{1t} + v_t$, for $v_t = O_p(1)$. Due to the random error of v_t, L_t is not a true factor. Similar to (1.11), we can consider the following regression.

$$y_{it} = \beta_{1i}^* L_t + \beta_{2i} \hat{F}_{2t} + u_{it}^*,$$

where $\beta_{1i}^* = \beta_{1i}/\delta$ and $u_{it}^* = u_{it} - v_t$. Then it is easy to show $\hat{\beta}_{1i}^* \not\to^p \beta_{1i}^*$ due to the correlation between L_t and u_{it}^*. Hence it is straightforward to show that $\lim_{N,T \to \infty} \Pr\left[\hat{\#}\{\hat{u}_{it}^*\} = 0\right] = 0$.

In sum, as long as we are interested in identifying whether or not $G_{jt} \in \{F_t\}$, we do not need any identification restriction on the rotation matrix, H. For example, Bai and Ng (2013) consider three identification restrictions for the rotation matrix, H, to be the identity matrix, I, so that $\hat{F}_t = F_t + o_p(1)$. The method used in our paper does not require such identification restrictions. We formally present the identification procedure in the next section.

1.3 Definitions and Identification Procedure

Before we start to provide identification procedures and strategies, we provide conceptual definitions of the empirical true factors: Dominant and approximate dominant leaders.

Definitions

Let $P_t = [P_{1t}, ..., P_{mt}]'$ be the $m \times 1$ vector of potential true factors which researchers want to examine. Note that m is not necessarily equal to r. We will discuss the reason shortly. If P_t are the true factors, then the inclusion of P_t into the panel data $y_t = [y_{1t}, ..., y_{Nt}]$ always leads to more accurate estimation of the common factors (See Boivin and Ng, 2006). Also, it is possible that a few leaders become the true common factors of the panel data. An example

of this endogenous estimation appears in Gaibulloev, Sandler and Sul (2012) which finds that transnational terrorism in Lebanon is the main determinant of transnational terrorism for the rest of the world. Hence without loss of generality, we can include P_t as a part of the panel data $\{y_{it}\}$ and re-order them as $\{y_{1t}, ..., y_{mt}, y_{m+1,t}, ..., y_{N+m,t}\}$ so that the first m individuals are the potential true common factors to $\{y_{it}\}$.

Definition (Dominant Leaders): *The jth unit becomes an exact dominant leader if and only if $y_{jt} = G_{jt}$.*

In general, the maximum number of dominant leaders should be the same as the number of true common factors. However sometimes, the number of leaders could be larger than the number of the factors especially when there are many approximate dominant leaders which can be defined as

Definition (Approximate Dominant Leaders): *The jth unit becomes an approximate dominant leader for the jth empirical true factor if and only if $G_{jt} = y_{jt} + \zeta_{jt}$ for $j = 1, ..., r$ where $\zeta_{jt} = \epsilon_{jt}/\sqrt{T}$ and $\epsilon_{jt} \sim d\left(0, \sigma_j^2\right)$ with $\max \sigma_j^2 = \bar{\sigma}^2 < \infty$.*

When $\sigma_j^2 = 0$, the jth unit becomes a dominant leader. The non-zero variance of σ_j^2 implies that the jth unit may lose its leadership temporarily. That is,

$$
y_{jt} = \begin{cases} G_{jt} & \text{if } t \notin \Omega_o \\ G_{jt} + \epsilon_{jt} & \text{if } t \in \Omega_o \end{cases} \quad \text{for } \epsilon_{jt} \sim d\left(0, \sigma_j^2\right)
$$

Let $t = 1, ..., T$ and Ω_o be a fixed set of time periods. Call the number of elements of Ω_o, p, which is fixed as $N, T \to \infty$. So y_{jt} is not the leader for p time periods. Then the variance of the deviation between the true factor and the dominant leader, $y_{jt} - G_{jt}$, becomes

$$
\sigma_{j,T}^2 = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T}(y_{jt} - G_{jt})^2\right] = \frac{p\sigma_j^2}{T} \text{ for a small constant } p > 1.
$$

We will later show empirical evidence for the existence of approximate dominant leaders. When there are approximate dominant leaders, then the number of these leaders can be larger than the number of true common factors. Also note that, asymptotically, the approximate dominant leaders cannot be distinguished from the dominant leader since as $T \to \infty$, approximate dominant leaders become the dominant leader.

Identification Procedures

The identification procedures differ depending on whether or not potential leaders are given to or selected by researchers. We first consider the simplest case where potential leaders are given or known. Here we assume that the number of the true factors is known. This assumption is fairly reasonable since Bai and Ng (2002)'s criteria perform fairly well when the panel data are rather homogeneous.[1] Note that we are identifying whether or not a time series is either a dominant or approximate dominant leader for $\hat{F}_{st}$ for $s = 1, ..., r$ since the true statistical factors are unknown.

To make a clear presentation, we use the case of $r = 2$ but $m = 3$ throughout this section. That is, G_t, F_t, and $\hat{F}_t$ become 2×1 vectors but the potential variable, P_t, is a 3×1 vector.

Even when H is an identity matrix, the first PC estimator $\hat{F}_{1t}$ can be F_{2t} depending on the values of factor loadings in (1.3), which also depend on the values of α in (1.1) and the variance matrix of G_t, Ω. However regardless of the ordering, the point of interest becomes whether or not $\hat{F}_{1t}$ can be identified by P_{jt} for $j = 1, .., m$. Let

$$G_{it} - P_{jt} = \gamma_{ij}\epsilon_{jt}/\sqrt{T} + \delta_{ij}\xi_{jt}, \tag{1.12}$$

---

[1]When we refer to panel homogeneity in this paper, we are specifically referring to the panel being constructed of one central variable, such as state-level unemployment rates over time. In terms of the factor structure, homogeneity appears when the order of integration is the same across cross-sectional units and when the idiosyncratic variances are not seriously heterogeneous.

where $\epsilon_{jt} \sim d\left(0, \sigma_{\epsilon,j}^2\right)$, $\xi_{jt} \sim d\left(0, \sigma_{\xi,j}^2\right)$ for a finite constant of $\sigma_{\epsilon,j}^2$ and $\sigma_{\xi,j}^2$. By definition, if $\delta_{ij} = 0$ but $\gamma_{ij} \neq 0$, then P_{jt} becomes the approximate dominant leader for G_{it}. All P_{jt} can be approximate dominant leaders for each particular $j = 1, ..., r$.

To identify whether or not P_{jt} is either a dominant or approximate dominant leader for $\hat{F}_{st}$ for $s = 1, ..., r$, we suggest examining whether or not the regression residuals from the following regressions have any strong common factors.

$$y_{it} = \beta_{2,ji} P_{jt} + \lambda_{2,ji} \hat{F}_{2t} + y_{2j,it}^o, \tag{1.13}$$

$$y_{it} = \beta_{1,ji} P_{jt} + \lambda_{1,ji} \hat{F}_{1t} + y_{1j,it}^o, \tag{1.14}$$

Suppose that $P_{1t} = G_{1t}$ exactly. However, even in this case, $P_{1t} \neq \hat{F}_{1t}$ but P_{1t} becomes a function of $\hat{F}_{1t}$ and $\hat{F}_{2t}$. Depending on the value of the off-diagonal element in B in (1.10), or alternatively the depending on the values of λ_{1i} and λ_{2i}, the estimated number of the common factors in either or both of $\hat{y}_{2j,it}^o$ and $\hat{y}_{1j,it}^o$ becomes zero. For $r \geq 2$, (1.13) and (1.14) can be written as

$$y_{it} = \beta_{s,ji} P_{jt} + \lambda_{i,-s}' \hat{F}_{-s,t} + y_{sj,it}^o \text{ for } s = 1, ..., r, \tag{1.15}$$

where $\hat{F}_{-s,t} = [F_{1t}, ..., F_{s-1,t}, F_{s+1,t}, ..., F_{rt}]$ and $\lambda_{i,-s} = [\lambda_{i1}, ..., \lambda_{is-1}, \lambda_{is+1}, ..., \lambda_{rt}]$. When $r = 1$, $\hat{F}_{-s,t}$ and $\lambda_{i,-s}'$ are not present in (1.15). Then more formally, we have

Theorem 1.1 (Identification of Estimated Factors: Known Potential Leaders)
Under the assumptions in Bai and Ng (2002),

(i) If $\delta_{ij} = 0$, then the probability that the regression residuals in (1.15) have zero factor becomes

$$\lim_{N,T \to \infty} \Pr\left[\hat{\#}\left(\hat{y}_{1j,it}^o\right) = 0 \text{ or } \hat{\#}\left(\hat{y}_{2j,it}^o\right) = 0 \text{ or,..., } \hat{\#}\left(\hat{y}_{rj,it}^o\right) = 0\right] = 1. \tag{1.16}$$

(ii) If $\delta_{ij} \neq 0$, then

$$\lim_{N,T \to \infty} \Pr\left[\hat{\#}\left(\hat{y}_{1j,it}^o\right) = 0 \text{ or } \hat{\#}\left(\hat{y}_{2j,it}^o\right) = 0 \text{ or,..., } \hat{\#}\left(\hat{y}_{rj,it}^o\right) = 0\right] = 0 \tag{1.17}$$

From Lemma 1, the proof of Theorem 1.1 becomes obvious. If all off-diagonal elements of the B matrix become zero – that is, when G_{1t} and G_{2t} are independent of each other and $H = I_2$, then either $\hat{y}^o_{1j,it}$ or $\hat{y}^o_{2j,it}$ has asymptotically weak factors. In this case, $\hat{\#}\left(\hat{y}^o_{2j,it}\right) \to 0$ but $\hat{\#}\left(\hat{y}^o_{1j,it}\right) \to 1$ if $P_{1t} = G_{1t}$. However this is an extreme case. In general, G_{1t} can be correlated with G_{2t} and also the rotating matrix H may not be the identity matrix, so that the off-diagonal elements of $B = AH^{-1}$ matrix are not equal to zero. Hence usually both $\hat{y}^o_{1j,it}$ and $\hat{y}^o_{2j,it}$ have only asymptotically weak factors if $\delta_{ij} = 0$. Hence $\hat{\#}\left(\hat{y}^o_{2j,it}\right) \to 1$ but $\hat{\#}\left(\hat{y}^o_{1j,it}\right) \to 1$ if $\delta_{1j} = 0$.

Many times when leaders are unknown and N is large, applying our criterion to each individual in the panel could lead to over-estimation of the number of approximate dominant leaders, since the 'size' of the procedure is non-zero. One solution to this problem is to run the following regression:

$$\hat{F}_{st} = c_{ss}P_{jt} + c_{s.-s}\hat{F}_{-s,t} + \varepsilon^*_{st} \text{ for each } P_j \text{ for each } s$$

and obtain the R^2-statistics. For each factor, $\hat{F}_{st}$, the individuals, P_j, with high R^2 value have high estimated partial correlation to the factor. Choosing to test only these individuals avoids over-estimation of the number of approximate dominant leaders. It is easy to show that this procedure is consistent as N and T go to infinity.

By running (1.13) and (1.14) for $r = 2$, or (1.15) for $r \geq 1$, approximate dominant leaders can be identified for any G_{st} but the dominant leaders for a particular G_{st} are not known. To achieve this, we suggest the following sieve method to cluster approximate dominant leaders to each G_{st}.

Clustering Method Suppose that there are k approximate dominant leaders identified by (1.16). If some P_{jt} are the approximate dominant leaders for G_{st}, then the variance ratio

between any asymptotically equal dominant leaders must not be different from one with a large T.

$$\begin{aligned}\lim_{T\to\infty} Corr\left(P_{jt}, P_{it}\right) = 1 & \quad \text{if } P_{jt} \in G_{st} \ \& \ P_{it} \in G_{st} \\ \lim_{T\to\infty} Corr\left(P_{jt}, P_{it}\right) = c < 1 & \quad \text{if } P_{jt} \in G_{st} \ \& \ P_{it} \notin G_{st}\end{aligned}, \tag{1.18}$$

For a finite T, P_{jt} can be clustered into r subgroups at once by checking the correlation orders. Alternatively, one can choose P_{jt}s for which correlation coefficients are relatively high. We will show later in the empirical example section that clustering P_{jt}s into a single subgroup is rather easy as long as the correlation between G_{1t} and G_{st} for $s \neq 1$ is relatively small.

However, the clustering mechanism based on the correlation order is not airtight in the finite sample. To verify whether or not the clustering order is well chosen, the following method is recommended. Obtain the LS residuals from the following regression by including a set of clustered leaders.

$$y_{it} = \sum_{j \in s} \alpha_{ji}^* P_{jt} + y_{s,it}^* \tag{1.19}$$

Then asymptotically the estimated number of the common factors should be equal to $r - 1$ if all P_{jt}s are approximate dominant leaders for a particular G_{st}.

$$\lim_{N,T\to\infty} \Pr\left[\hat{\#}\left(\hat{y}_{s,it}^*\right) = r - 1\right] = \begin{cases} 1 \text{ if all } P_{jt} \in G_{st} \\ 0 \text{ if any } P_{jt} \notin G_{st} \end{cases}. \tag{1.20}$$

Once all P_{jt}s are found for a particular G_{st}, the whole process can be repeated until there are no remaining dominant leaders.

Comparison to Extant Testing Methods

Bai and Ng (2006) considers a similar problem. Their test is originally designed to examine whether or not observed vectors of variables, P_t, are true factors, F_t. They do not explicitly discuss whether P_t can be members of $\{y_{it}\}$ and consider only "outside" macro and financial

variables in their empirical application. However it is straightforward to extend their method to identify dominant or sub-dominant leaders in the contour factor model.

Their test is based on (1.9). That is, if P_{jt} is one of the dominant leaders, then the following should hold.

$$P_{jt} = A_j F_t + \pi_{jt}, \text{ for } \pi_{jt} = 0 \text{ all } t.$$

Their infeasible version of the test is examining whether or not π_{jt} are zero statistically for all t. However this logic fails when P_{jt} is an approximate dominant leader. In this case, $\pi_{jt} = \epsilon_{jt}/\sqrt{T}$. Evidently this test requires an extremely large T to have $\pi_{jt} = 0$ for all t when P_{jt} is an approximate dominant leader.

Also the true statistical factors are not observable, the Bai and Ng (2006)'s test is based on the PC estimator of $\hat{F}_t$. Then from (1.10), if P_t were exact dominant leaders, then

$$\pi_t = B\varepsilon_t = O_p\left(N^{-1/2}\right).$$

In this case, π_t must be zero for all t as $N \to \infty$. Hence their test is based on the following statistic

$$\tau_t\left(j\right) = \frac{\hat{P}_{jt} - P_{jt}}{\sqrt{V\left(\hat{P}_{jt}\right)}}.$$

In their Monte Carlo simulation, they found that the performance of the max τ_t test works well. The max τ_t test is defined as

$$M\left(j\right) = \max_{1 \leq t \leq T} \left|\hat{\tau}_t\left(j\right)\right|,$$

where $\hat{\tau}_t\left(j\right)$ is obtained with the estimate of $V\left(\hat{P}_{jt}\right)$. Their test is designed to be conservative so that the test is supposed to reject the null of $P_t = \hat{F}_t$ as long as π_{jt} has a positive variance as N or $T \to \infty$. However if π_t is bounded by $O_p\left(T^{-1/2}\right)$, it is easy to show that the max τ_t test fails. As we discussed above, if P_t are approximate dominant leaders, then $\pi_t = O_p\left(C_{NT}^{-1}\right)$. Hence if $N > T$, then the max τ_t test fails.

Nonetheless, we will show later but the max τ_t test rejects the null hypothesis too often even when P_t is an exact dominant leader. In other words, the size distortion of the test becomes extremely serious.

Utilizing Other Factor Number Estimation Methods

In this paper, the Bai and Ng (2002) criterion is used to estimate the number of factors in the residuals. Recently, other procedures have been suggested for estimating the number of factors in panel data sets. For instance, Onatski (2009) and Hallin and Liska (2007) have suggested alternative methods. These methods were found to perform similarly to the Bai and Ng (2002) criteria. For more detailed results and Gauss code, see the corresponding author's website.

Let the ith eigenvalue of the $(NT)^{-1} \hat{y}^{o\prime} \hat{y}^{o}$ matrix be denoted as ϱ_i. The null hypothesis for the Onatski (2009) test is $H_0 : \# (\hat{y}_{it}^o) = 0$ versus the alternative, $H_1 : 0 < \# (\hat{y}_{it}^o) \leq r_{\max}$. The test statistic used is

$$Q_{Onat} = \max_{0 < k \leq k_{\max}} \left[\ln \left(\varrho_i - \varrho_{i+1} \right) - \ln \left(\varrho_{i+1} - \varrho_{i+2} \right) \right]$$

The null hypothesis is rejected if Q_{Onat} is greater than some critical value. The Bai and Ng (2002) procedure can be written as

$$Q_{BN} = \max_{0 < k \leq k_{\max}} \left[\ln \left(\sum_{i=1}^{N} \varrho_i \right) - \ln \left(\sum_{i=k+1}^{N} \varrho_i \right) - k \times p \left(N, T \right) \right].$$

The null hypothesis is rejected if Q_{BN} is less than than zero. The primary benefit of the Bai and Ng (2002) criterion is that it does not require a distribution to find the critical value. The Onatski (2009) procedure has the benefit that the operator can control the size of the test.

The Hallin and Liska (2006) criterion uses subsamples of the panel, $0 < N_1 < N_2 < \cdots < N_L = N$ and $0 < T_1 < T_2 < \cdots < T_M = T$. Denote $\varrho_i^{(l,m)}$ as the eigenvalues of $(N_l T_m)^{-1} \hat{y}^{o\prime} \hat{y}^{o}$

computed only using the values of $\hat{y}_{it}^o$ where $i \leq N_l$ and $t \leq T_m$. Let

$$\hat{k}_{BN}(c, l, m) = \underset{0 \leq k \leq k_{\max}}{\arg\min} \left[\ln \left(\sum_{i=k+1}^{N} \varrho_i^{(l,m)} \right) + k \times c \times p(N_l, T_m) \right].$$

This is the subsample, scaled analog of the Bai and Ng (2002) IC criterion. $\hat{k}_{BN}$ is a function of a positive constant c which controls the sensitivity of the estimator. When c is small, $\hat{k}_{BN}$ does not penalize extra factors, so the estimator finds $k_{\max}$ as the number of factors. When c is large, $\hat{k}_{BN}$ over-penalizes the factors, so $\hat{k}_{BN}$ finds zero factors in the residual. The $S(c)$ function is defined by

$$S(c) = \left(\frac{1}{LM} \sum_{l,m} \left(\hat{k}_{BN}(c, l, m) - \frac{1}{LM} \sum_{l,m} \hat{k}_{BN}(c, l, m) \right)^2 \right)^{1/2}.$$

Under Bai and Ng (2002) there is no control for the coefficient, c, before the penalty function. In other words, Bai and Ng (2002) choose the value: $\hat{k}_{BN} = \hat{k}_{BN}(1, L, M)$. Hallin and Liska (2007) use subsamples to choose $c = c_o$ in a region where S vanishes. Because of this control, the penalty function is less sensitive to the size of N and T. $\hat{k}_{HL}$ is equal to $\hat{k}_{BN}$ evaluated at c_o, that is to say $\hat{k}_{HL} = \hat{k}_{BN}(c_o, L, M)$.

1.4 Practical Suggestions and Monte Carlo Studies

This section provides the data generating process used in the Monte Carlo studies, summarizes the procedure we discussed in earlier sections, and reports the results of Monte Carlo simulations.

Data Generating Process

The data generating process is given by (1.1) where factor loadings and idiosyncratic errors are generated from

$$\alpha_{1i} \sim iidN(0, 1), \ \alpha_{2i} \sim iidN(0, 1), \ \text{and} \ y_{it}^o \sim iidN(0, 1). \tag{1.21}$$

The first two series, y_{1t} and y_{2t}, are the true factors G_{1t} and G_{2t}.

Meanwhile the statistic factors are generated from

$$
\begin{bmatrix} F_{1t} \\ F_{2t} \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} \begin{bmatrix} F_{1,t-1} \\ F_{2,t-1} \end{bmatrix} + \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}, \tag{1.22}
$$

where $f_{jt} \sim iidN\left(0, \sqrt{1-\rho_j}\right)$ for $j = 1, 2$ and ρ_1 and ρ_2 are distributed $iidU\left[0, 0.5\right]$. Hence the variances of F_{1t} and F_{2t} are unity and they are independent each other. The empirical factor are generated by multiplying a Cholesky decomposed matrix of A. That is,

$$
\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} G_{1t} \\ G_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} F_{1,t-1} \\ F_{2,t-1} \end{bmatrix}, \tag{1.23}
$$

where

$$
E \begin{bmatrix} G_{1t} \\ G_{2t} \end{bmatrix} \begin{bmatrix} G_{1t} & G_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12} & \Omega_{22} \end{bmatrix}. \tag{1.24}
$$

We consider four cases of different Ω: $\left[\Omega_{11}, \Omega_{12}, \Omega_{22}\right] = \left[1, 0.2, 1\right], \left[1, 0.5, 1\right], \left[2, 0.2, 1\right]$, and $\left[2, 0.5, 1\right]$. Meanwhile for all simulations, we set $N = \left[25, 50, 100, 200\right]$ and $T = \left[25, 50, 100, 200\right]$.

Identifying Procedures

Here we present a step-by-step procedure used in Monte Carlo simulation and empirical exercises.

Step 1: (Estimation of Factor Number and Common Factors)

The factor number, r, and the common factors should be estimated first. Before estimating the factor number and common factors, one should standardize each time series by its standard deviation. In our empirical examples, we usually take logs, difference, and then standardize the sample. For the factor number estimation, we use Bai and Ng (2002)'s

IC_2 and Hallin and Liska (2007)'s IC_2 criteria (hereafter IC_2 and HL_2, respectively). Note that other Bai and Ng's criteria perform worse than IC_2. It is worth noting that the HL criteria use sub-samples over N and T. Hence when N or T is small, HL criteria perform very poorly. To overcome this issue, we use whole N or T sample whenever N or T is small in our empirical example, especially the data of 21 exchange rates. Here we fix N but use sub-sample over T only.

Step 2: (Identifying Potential Leaders by using R^2)

First we select potential leaders by using the R^2 criterion. From (1.10), the following regressions can be performed.

$$\hat{F}_{1t} = c_{11}P_{jt} + c_{12}\hat{F}_{2t} + \varepsilon_{1t}^* \tag{1.25}$$

$$\hat{F}_{2t} = c_{21}P_{jt} + c_{22}\hat{F}_{1t} + \varepsilon_{2t}^* \tag{1.26}$$

If $P_{jt} = G_{1t}$, or $P_{jt} = G_{1t} + \zeta_{jt}$ for $\zeta_{jt} = \epsilon_{jt}/\sqrt{T}$, then as $N, T \to \infty$, the variance of ε_{1t}^* goes to zero. Alternatively, if $P_{jt} = G_{1t} + m_{jt}$ where m_{jt} has a finite variance, then the variance of ε_{1t}^* should not be close to zero. Similarly, if $P_{jt} = G_{2t}$ or $P_{jt} = G_{2t} + \zeta_{jt}$, then the variance of ε_{2t}^* also goes to zero as $N, T \to \infty$. By utilizing this fact, we select P_{jt} of which R^2 is highest. It will later be shown that this R^2 method detects the potential leaders very precisely even in small samples.

Step 3: (Identifying Approximate Dominant Leaders)

Run (1.13) or (1.14) and obtain the regression residuals. Check whether or not the estimated factor number becomes zero. Following Theorem 1.1, identify whether or not P_{jt} is an approximate dominant leader. Collect all dominant leaders.

Step 4: (Clustering Approximate Dominant Leaders)

Calculate the correlation matrix among approximate dominant leaders if they are many. Verify the membership by running (1.19).

See next section for the demonstration of the above four steps.

Monte Carlo Results

By means of Monte Carlo simulation, we verify our theoretical claims and investigate their finite sample performance.

The Factor Number Estimation of Asymptotically Weak Factors

First, we examine how well Lemma 1 works in finite sample environments. We consider three different data generating processes. The first DGP is given in (1.1) with $[\Omega_{11}, \Omega_{12}, \Omega_{22}] = [1, 0.2, 1]$. The second and third DGPs are the same – single factor with $F_t \sim iidN(0, 1)$ and $y_{it}^o \sim iidN(0, 1)$ – except for the variance of factor loadings. That is,

$$\begin{aligned} \text{DGP2} \quad & y_{it} = \lambda_i F_t + y_{it}^o, \quad \lambda_i \sim iidN\left(1, N^{-1}\right) \\ \text{DGP3} \quad & y_{it} = \lambda_i F_t + y_{it}^o, \quad \lambda_i \sim iidN\left(1, T^{-1}\right) \end{aligned}.$$

Hence the true number of factors are 2, 1 and 1 for DGP 1,2,3, respectively. According to Bai and Ng (2002), the estimated factor number must be r as $N, T \to \infty$.

The estimated idiosyncratic terms are defined as

$$\begin{aligned} \text{DGP1} \quad & \hat{y}_{it}^o = y_{it} - \hat{\lambda}_{1i}\hat{F}_{1t} - \hat{\lambda}_{2i}\hat{F}_{2t}, \\ \text{DGP2 \& 3} \quad & \hat{y}_{it}^o = y_{it} - \frac{1}{N}\sum_{i=1}^N y_{it}. \end{aligned}$$

Following from Lemma 1, we expect that $\Pr\left[\hat{\#}\left(\hat{y}_{it}^o\right) = 0\right]$ should be reasonably high.

Table 1.1 shows the results of Monte Carlo simulations of 2,000 replications. We set the maximum lag length to be 8. Both IC_2 and HL_2 detect the correct factor number with

Table 1.1. Detecting Asymptotically Weak Factors

IC_2		$\Pr\left[\hat{\#}\left(y_{it}\right)=r\right]$			$\Pr\left[\hat{\#}\left(\hat{y}_{it}^{o}\right)=0\right]$		
T	N	DGP1	DGP2	DGP3	DGP1	DGP2	DGP3
25	25	0.97	1.00	0.98	0.99	1.00	0.84
25	50	1.00	1.00	1.00	1.00	1.00	0.99
25	100	1.00	1.00	1.00	1.00	1.00	1.00
50	25	1.00	1.00	1.00	1.00	1.00	0.98
50	50	1.00	1.00	1.00	1.00	1.00	0.99
50	100	1.00	1.00	1.00	1.00	1.00	0.99
100	25	1.00	1.00	1.00	1.00	1.00	1.00
100	50	1.00	1.00	1.00	1.00	1.00	0.98
100	100	1.00	1.00	1.00	1.00	1.00	0.98
HL_2							
25	25	0.78	1.00	1.00	1.00	1.00	0.99
25	50	0.99	1.00	1.00	1.00	1.00	1.00
25	100	1.00	1.00	1.00	1.00	0.99	1.00
50	25	0.98	1.00	1.00	0.99	1.00	1.00
50	50	0.98	1.00	1.00	0.98	0.99	0.98
50	100	0.99	1.00	1.00	0.99	0.97	0.99
100	25	1.00	1.00	1.00	1.00	1.00	1.00
100	50	0.98	1.00	1.00	0.97	1.00	0.98
100	100	0.98	1.00	1.00	0.98	0.99	0.97

very sharp precision with y_{it}. However Bai and Ng's (2002) IC_2 performs slightly better than Hallin and Liska's (2007) HL_2 criterion. With the estimates of $\hat{y}_{it}^{o}$, IC_2 performs much better than HL_2 and provides extreme accuracy even when N or T are small. For the case of DGP 3, the probability of detecting zero factors does not become unity when N is large. This is a rather natural result since the maximum value of λ_i is not bounded over N. When λ_i is generated from the uniform distribution, this anomaly disappears completely. In sum, the results in Table 1.1 support our theoretical claim in Lemma 1 very well even with finite samples.

Identifying True Factors with given Potential Factors

Next, we simulate Theorem 1.1 by considering the case where the potential factors are given. The DGP is given in (1.1) with $[\Omega_{11}, \Omega_{12}, \Omega_{22}] = [2, 0.5, 1]$. Other cases are available online.[2]

Table 1.2 reports the frequencies at which the estimated factor number becomes zero. That is, for the first and second factor case, we report $\Pr\left[\hat{\#}\left(\hat{y}_{11,it}^{o}\right) = 0 \text{ or } \hat{\#}\left(\hat{y}_{21,it}^{o}\right) = 0\right]$ and $\Pr\left[\hat{\#}\left(\hat{y}_{12,it}^{o}\right) = 0 \text{ or } \hat{\#}\left(\hat{y}_{22,it}^{o}\right) = 0\right]$ over 2,000 replications. The first four columns show the probability of zero factor number when the potential factors are exact dominant leaders. Note that when N is small $(N = 25)$, the probabilities of detecting the true factors seem to be decreasing as T increases when IC_2 is used. It may be due to the inaccuracy of the estimation of the common factor. However when N is moderately large, such anomaly disappears completely. IC_2 always identifies the true factor with a great accuracy. Meanwhile HL_2 does not show this anomaly when N is small, but becomes less accurate compared to IC_2.

The last four columns in Table 1.2 report the case where the potential factors are not true factors. Evidently, when N or T is small, both IC_2 and HL_2 fail to identify that the potential factors are not true factors. However, as either T or N increases, the frequency of estimating zero factors decreases very quickly. Note that IC_2 detects much better when N or T is small compared to HL_2. Meanwhile HL_2 performs relatively well when N or T becomes large. In sum, the results in Table 1.2 support Theorem 1.1 strongly.

Table 1.3 reports the finite sample performance of Bai and Ng (2006)'s $\max_t \tau_t$ test. The underlying DGP is exactly same as that in Table 1.2. The nominal size is fixed to be 0.05. The size and size-adjusted power are reported. As we discussed before, as N increases with a fixed T, the estimated common factors become more accurate which leads the size distortion of the $\max_t \tau_t$ test to become lessened. However, as T increases, the size distortion increases

---

[2]The $[1, 0.5, 1]$, $[1, 0.2, 1]$, and the $[2, 0.2, 2]$ cases are reported online along with other calculations (more specific today).

Table 1.2. Estimated Probability of Detecting True Factors

| | | $P_{jt} = G_{jt}$ (Exact) | | | | $P_{jt} = G_{jt} + \varepsilon_{jt}$ (False) | | | |
| | | IC_2 | | HL_2 | | IC_2 | | HL_2 | |
T	N	1st	2nd	1st	2nd	1st	2nd	1st	2nd
25	25	0.97	0.96	1.00	1.00	0.21	0.23	0.54	0.53
25	50	1.00	1.00	1.00	1.00	0.16	0.15	0.25	0.23
25	100	1.00	1.00	1.00	1.00	0.12	0.12	0.13	0.13
25	200	1.00	1.00	1.00	1.00	0.09	0.10	0.07	0.07
50	25	0.95	0.95	0.99	1.00	0.12	0.11	0.31	0.31
50	50	1.00	1.00	0.98	0.98	0.11	0.11	0.07	0.07
50	100	1.00	1.00	0.99	0.99	0.05	0.05	0.01	0.02
50	200	1.00	1.00	1.00	1.00	0.03	0.04	0.00	0.00
100	25	0.91	0.91	0.99	1.00	0.07	0.07	0.23	0.23
100	50	1.00	1.00	0.98	0.98	0.05	0.05	0.03	0.03
100	100	1.00	1.00	0.98	0.98	0.03	0.03	0.00	0.00
100	200	1.00	1.00	0.99	1.00	0.01	0.01	0.00	0.00
200	25	0.90	0.90	1.00	0.99	0.06	0.05	0.21	0.19
200	50	1.00	1.00	0.98	0.98	0.03	0.03	0.02	0.02
200	100	1.00	1.00	0.98	0.98	0.01	0.01	0.00	0.00
200	200	1.00	1.00	0.99	0.99	0.00	0.00	0.00	0.00

significantly. Meanwhile as Bai and Ng (2006) reported, the power of the test becomes perfect.

Identifying True Factors when Potential Leaders are Unknown

When potential leaders are unknown, they must be estimated. We use the R^2 criterion discussed in Section 1.3. The DGP is given in (1.1) with $[\Omega_{11}, \Omega_{12}, \Omega_{22}] = [1, 0.2, 1]$. As we mentioned earlier, we report all other cases online. We generate two approximate dominant leaders for each factor as follows

$$P_{jt} = G_{1t} + \epsilon_{jt}/\sqrt{T} \text{ for } j = 1, 2, \text{ and } P_{jt} = G_{2t} + \epsilon_{jt}/\sqrt{T} \text{ for } j = 3, 4.$$

The first four approximate factors are included in the panel data y_{it}. That is, $y_{it} = P_{it}$ for $i = 1, 2, 3, 4$. For the rest of y_{it}, we impose the following restriction on the idiosyncratic

Table 1.3. Bai and Ng's $\max_t \tau_t$ Test Rejection Rates (Size: 5%)

T	N	Size		Power	
		1st	2nd	1st	2nd
25	25	0.40	0.44	0.99	0.99
25	50	0.32	0.39	1.00	1.00
25	100	0.25	0.31	1.00	1.00
25	200	0.18	0.26	1.00	1.00
50	25	0.49	0.52	1.00	0.99
50	50	0.41	0.48	1.00	1.00
50	100	0.33	0.42	1.00	1.00
50	200	0.22	0.32	1.00	1.00
100	25	0.55	0.60	1.00	1.00
100	50	0.50	0.58	1.00	1.00
100	100	0.44	0.51	1.00	1.00
100	200	0.33	0.45	1.00	1.00
200	25	0.63	0.65	1.00	1.00
200	50	0.59	0.64	1.00	1.00
200	100	0.52	0.61	1.00	1.00
200	200	0.43	0.53	1.00	1.00

variance.

$$y_{it}^o \sim iidN\left(0, \sigma_i^2\right), \text{ for } \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T} C_{it}^2, \ C_{it} = \lambda_i' F_t. \tag{1.27}$$

Without imposing this restriction, there is always a chance that some y_{it} for $i > 4$ becomes an approximate common factor with high λ_{1i} and λ_{2i}. Alternatively, we can generate λ_is from a uniform distribution without imposing the restriction in (1.27). The simulation results are similar each other, so only this case is reported.

We choose each four potential leaders by maximizing the R^2 statistic from (1.25) and (1.26). So the maximum number of potential leaders becomes eight. Next, we check whether or not each potential leader is truly a common factor by estimating the number of the common factors to the regression residuals in (1.13) and (1.14).

Table 1.4 reports the frequencies with which approximate dominant leaders are selected as the true factors by combining the sieve method with the R^2 criterion together. The

Table 1.4. Identifying Potential Leaders by using R^2

T	N	All P_{jt} Selected	Only P_{jt} Selected
25	25	0.98	0.48
25	50	0.94	0.58
25	100	0.93	0.68
25	200	0.94	0.66
50	25	0.99	0.97
50	50	0.99	0.97
50	100	0.97	1.00
50	200	0.98	1.00
100	25	1.00	1.00
100	50	1.00	1.00
100	100	0.99	1.00
100	200	0.99	1.00
200	25	1.00	1.00
200	50	1.00	1.00
200	100	1.00	1.00
200	200	1.00	1.00

first column reports the correct inclusion rate that all four approximate dominant leaders are selected as the true factors. The second column shows the frequency that only four dominant leaders are selected as the true factors. Evidently, the sieve method with the R^2 criterion suggested in

Section 1.3 works very well. When T is moderately large ($T \geq 50$), the suggested method selects only correct approximate dominant leaders as the true factors.

Finally, Table 1.5 reports the clustering results. As it shown in Table 1.4, the accuracy of identifying the true factor is fairly sharp. Hence we assume that the approximate dominant leaders are given. We don't use the correlation method but rather use the criteria in (1.19) since the selected leaders are only four. The first column in Table 1.5 reports the frequency that the clustering algorithm selects correct members. The second column shows the false inclusion rate. Obviously the criteria in (1.19) demonstrates the pinpoint accuracy.

Table 1.5. Clustering Frequency

T	N	Correctly Clustered	False Inclusion Rate
25	25	0.99	0.00
25	50	1.00	0.00
25	100	1.00	0.00
25	200	1.00	0.00
50	25	1.00	0.00
50	50	1.00	0.00
50	100	1.00	0.00
50	200	1.00	0.00
100	25	1.00	0.00
100	50	1.00	0.00
100	100	1.00	0.00
100	200	1.00	0.00
200	25	1.00	0.00
200	50	1.00	0.00
200	100	1.00	0.00
200	200	1.00	0.00

1.5 Empirical Examples

This section provides three empirical examples. The first example uses a state-level panel of burglary rates in the U.S. from the FBI Uniform Crime Reports. The second example uses Global Insight data on the nominal exchange rates of 21 currencies against the U.S. dollar. The last example considers the three Fama French (1993) factors as potential factor candidates for a panel of portfolios.

State Panel of Burglary Rates

The FBI Uniform Crime Reports contain state-level, annual burglaries per 100,000 persons from 1965 to 2010 for the 50 United States. Natural logs are taken before first-differencing; then the series are demeaned and standardized. In this example, both IC_2 and HL_2 estimate one factor when k_{max} is set to 10. This result does not change if the starting year is allowed to vary up to 1995. Hence, burglary rates are found to have one factor.

Since the number of common factors is one, the homogeneity test on the factor loading coefficients becomes our interest. Note that the underlying model can be re-expressed as,

$$y_{it} = \bar{\lambda}F_t + \left(\lambda_i - \bar{\lambda}\right)F_t + y_{it}^o, \tag{1.28}$$

where $\bar{\lambda}$ is the average of the factor loadings. When the cross-sectional average is subtracted from (1.28), $\bar{\lambda}F_t$ vanishes and the common component then becomes $\left(\lambda_i - \bar{\lambda}\right)F_t$. If $\lambda_i = \lambda + O_p\left(C_{NT}^{-1}\right)$, then the resulting panel data after taking off the cross-sectional averages will only have asymptotically weak factors.

Table 1.6 reports the result. Evidently, both IC_2 and HL_2 estimate zero factor number after taking off both time series and cross-sectional averages, which implies that the factor loadings are asymptotically homogeneous.

Table 1.6. Evidence of Homogeneity Factor Loadings in 50 State-level US Burglary Rates

Demeaning by	IC_2	HL_2
Time Series Mean Only	1	1
Time Series Mean and Cross-Sectional Average	0	0

Nonetheless, we proceed with leadership identification next, using the time series mean only.[3] First, we select 3 potential leaders by choosing the individuals with highest R^2 from:

$$\hat{F}_t = \alpha_{1i} + \beta_i y_{it} + m_{it} \text{ for } i = 1, \ldots, 50. \tag{1.29}$$

Including Connecticut as the regressor in (1.29) is found to maximize the R^2 value. The R^2 values do not significantly decline between candidates, indicating that burglary rates are seriously affected by the underlying factor.

Next, the following regression is performed for each leader candidate, l, to obtain the regression residuals:

$$y_{it} = \alpha_{2i} + \phi_i y_{l,t} + y_{it}^o \text{ for } i \neq l.$$

---

[3]It would be unreasonable to use the time series mean and cross-sectional average case because it has no common factors, so principal components estimation will perform quite poorly.

If the factor number of these residuals is estimated to be zero, then $y_{l,t}$ is estimated to be a leader. In both examples, every factor number estimation finds at least one common factor, so no particular state becomes a leader in either example. The results are provided in Table 1.7.

Table 1.7. Leadership Estimation for a Single Factor

| Potential Leader | R^2 | Estimation Methods | |
Candidate	Statistics	IC_2	HL_2
Connecticut	0.783	1	1
Virginia	0.771	1	1
South Carolina	0.759	1	1

Exchange Rates

GMSW identify Euro/USD, Swiss Franc/USD and Japanese Yen/USD as key exchange rates which explain the estimated common factors very well. GMSW justified their result in a number of ways, and gave a thorough interpretation of the results. We re-estimate the empirical common factors. There is one minor difference from GMSW. We neutralize the effect of the numeraire currency by taking off the common time effects. Note that the bilateral exchange rates with USD numeraire can be changed to those with Euro by subtracting the log Euro/USD. To see this, consider the following simple example. The log spot Yen/USD rate can be converted to the log Yen/Euro rate by subtracting the log Euro/USD rate.

$$\ln \text{Spot} \left(\frac{\text{Yen}}{\text{USD}} \right) - \ln \text{Spot} \left(\frac{\text{Euro}}{\text{USD}} \right) = \ln \text{Spot} \left(\frac{\text{Yen}}{\text{Euro}} \right)$$

This currency conversion makes the new bilateral exchange rates with Euro numeraire included a common factor of the log Euro/USD. Hence by taking off the cross-sectional average (or common time effects), we can eliminate the effects of the numeraire currency. We update

21 bilateral exchange rates to 2012.M12 and estimated the factor number with the maximum lag length of 8. To be specific, we take natural log, take first difference, subtract the cross-sectional average, and standardize each series before estimating the factor number.

Table 1.8 demonstrates how we select Euro/USD and Swiss Franc/USD as the determinant of the first factor and Yen/USD as the determinant of the second factor. The first step is the estimation of the factor number. Both IC_2 and HL_2 criteria estimate two factors and this result does not change at all regardless of the different starting samples. It is worthy noting that the estimation result based on HL_2 criterion may change the order of the cross-sectional units since the HL criterion is based on stability of subsample analyses both over N and T. For this example, we allocate three empirical factors, Euro/USD, Yen/USD, and Swiss Franc/USD as $i = 2, 3, 4$ respectively. Note that GMSW estimate three factors.

As the second step, we ran (1.25) and (1.26) and selected four exchange rates of which R^2 are highest. For the first factor, Euro, Taiwan, Thailand, and Philippine currencies against USD are selected. For the second factor, Euro, Swiss, Japan and Australia currencies against USD are selected. We ran (1.13) and (1.14), obtained the regression residuals, and checked whether or not the estimated factor number becomes zero. From the first regression, Euro/USD is identified as the approximate factor and from the second regression, Swiss Franc and Japanese Yen against USD are identified as the approximate factor. Hence by the end of Step 3, we have the three approximate factors in our hands.

As the final step of clustering, we check the correlation matrix among three factors. Euro/USD is highly correlated with Swiss Franc/USD meanwhile Japanese Yen/USD seems to be isolated from the other two. From this result, we hypothesize that Euro/USD and Swiss Franc/USD are the approximate factor for the first empirical factor, and the Yen/USD becomes the second empirical factor. We verify our result by running (1.19) and find that IC_2 pairs Euro and Swiss Franc as the same factor. Note that HL_2 gives the different result.

Table 1.8. Identifying Common Factors in 21 Spot Rates

Step 1: Estimation of the Factor Numbers		
	IC_2	HL_2
21 Spot Rates	2	2

Step 2 & 3: Selecting Potential Leaders by R^2 and Identifying Leaders

Augmented with $\hat{F}_{1t}$	IC_2	HL_2	IC_2	HL_2	R^2
Euro	0	0	0	0	0.624
Taiwan	1	1	1	0	0.604
Thailand	1	1	1	1	0.572
Philippine	1	1	1	1	0.566
Augmented with $\hat{F}_{2t}$					
Euro	0	0	0	0	0.608
Swiss	1	1	0	0	0.582
Japan	1	1	0	0	0.453
Australia	1	1	1	1	0.447

Step 4: Clustering Leaders
Correlation Matrix

	Japan	Swiss
Euro	0.047	0.664
Japan		0.291

Verifying Membership by using (1.19)

Regressors	IC_2	HL_2
Euro &Swiss	1	1
Euro &Japan	0	1

Fama-French Three Factor Model

One of the most popular examples in factor analysis is the Fama-French portfolio theory. Fama and French (1993) consider three factors for portfolio returns, denoted as follows: 'Market', 'SMB', and 'HML'. We are interested in testing if these empirical factors are actually the underlying unobserved. This is not a leadership model; rather, this is an exogenous factor test. The famous Fama-French three factor model is given by

$$y_{it} = r_{ft} + \alpha_{1i}(\text{Mkt}_t - r_{ft}) + \alpha_{2i}\text{SMB}_t + \alpha_{3i}\text{HML}_t + y_{it}^o,$$

where r_{ft} is the risk free return rate. For our analysis, we use annual average value portfolio returns for 96 portfolios plus the three Fama-French factors from 1964 to 2008. The data available on Kenneth French's website begins in 1927 and ends in 2012. Our sample is chosen to avoid missing data before 1964 and structural market changes from the 2008 financial collapse. The returns are demeaned (by time series averages) and standardized. The maximum factor number is set to be 10. According to this theory, y_{it} shares the same common factor of r_{ft}. Hence we take off the cross-sectional average first, and then choose Mkt, SMB and HML as the known potential factors. We estimate the number of common factors after standardizing the sample. We denote this sample as $\tilde{y}_{it}$ to distinguish the sample y_{it} where the cross sectional averages are not taken off. Table 1.9 reports the summary of the results. First, both IC_2 and HL_2 estimate 2 factors surprisingly. This result does not change at all depending on the different ending years. This result implies that if Fama-French three factor model were indeed correct, then one of the three factor loadings must be homogeneous. Moreover if we don't take off the cross-sectional average, then we find three factors.

To identify this, we ran (1.19). That is, the panel is regressed on each hypothesized factor, Mkt, SMB, and HML:

$$\tilde{y}_{it} = \gamma_{j,i} P_{j,t} + e_{j,it},$$

where $j = Mkt, SMB$, and HML separately, one at a time. The factor number is estimated for each of the residual panels from these regressions. SMB and HML are found to be leaders. 'Market' is not found to be a leader. Next, the following regression is performed to see if SMB and HML become the same or different factors:

$$\tilde{y}_{it} = \gamma_{1,i}^* P_{SMB,t} + \gamma_{2,i}^* P_{HML,t} + e_{it}^*.$$

The estimated factor number of the residual is found to be zero. Hence, SMB and HML must account for different factors.

Table 1.9. Fama French Subsample Analysis for Leadership Estimation by Ending Year (Starting Year: 1964; $N = 99$)

Ending Year	Estimated Factor Number (IC_2)					
	with	with	Regressand: $\tilde{y}_{it}$, Regressor			
	y_{it}	$\tilde{y}_{it}$	Mkt	SMB	HML	SMB & HML
2008	3	2	2	1	1	0
2007	3	2	2	1	1	0
2006	3	2	2	1	1	0
2005	3	2	2	1	1	0
2004	3	2	2	1	1	0
2003	3	2	2	1	1	0
2002	3	2	2	1	1	0
2001	3	2	2	1	1	0
2000	3	2	2	1	1	0

To see whether or not the factor loading coefficients on Mkt are homogeneous, we estimate the first common factor without taking off the cross-sectional average, y_{it}. And then we regress $\hat{F}_{1t}$ on the risk free rate, r_{ft}, and $Mkt_t - r_{ft}$. This regression identifies the weighting coefficients between r_{ft} and $Mkt_t - r_{ft}$. The fitted value, $\tilde{F}_{1t}$, is obtained approximately the following weights.

$$\tilde{F}_{1t} = -0.0227 r_{ft} - 0.0032 \left(Mkt_t - r_{ft} \right).$$

Next, in Figure 1.1 we plot the estimated first common factor, the fitted value of $\tilde{F}_{1t}$, and $Mkt_t - r_{ft}$ together after standardization of each series. Evidently, the fitted value, $\tilde{F}_{1t}$, is more similar to the estimated factor than the fit by only market. After 1972, Mkt-Rft beats the fitted values by Rft and Mkr-Rf in only 3 time periods. From this empirical evidence, we conclude that the factor loadings on Mkt are almost homogeneous across each portfolio.

1.6 Conclusion

Factor analysis has become an increasingly popular tool in empirical research. Because there is no well-defined strategy for factor identification, researchers have been forced to choose
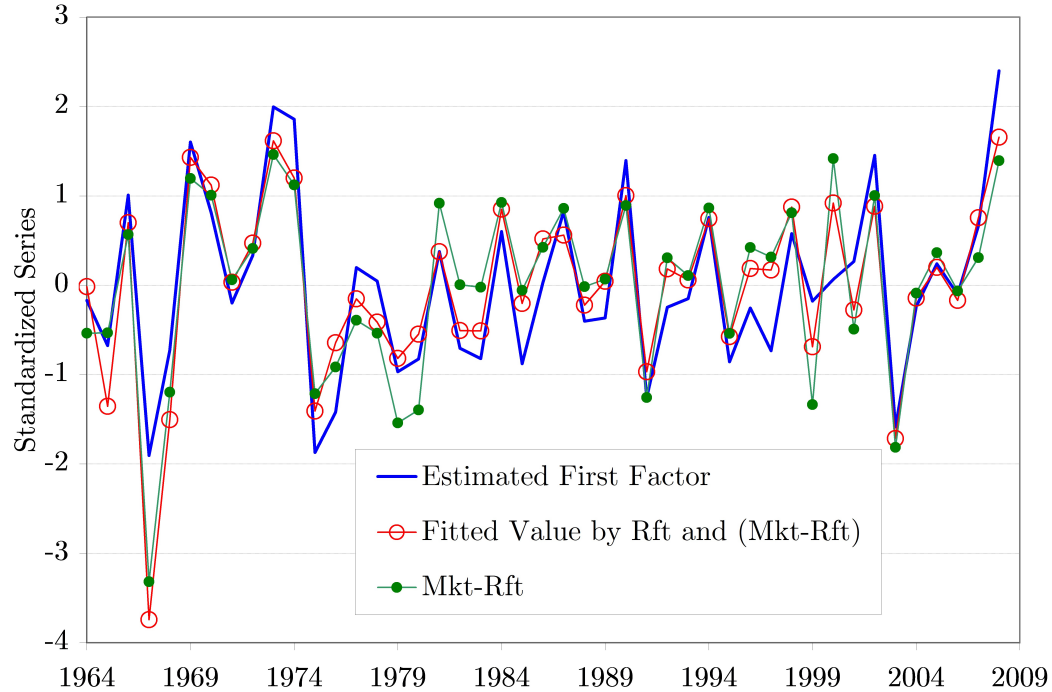
Figure 1.1. Missing Factor by Taking off Cross-Sectional Averages

between two outcomes. A researcher can either ignore any economic interpretation of the factor (perhaps the most important part of their model), or the researcher can speculate about the determinant without any concrete justification.

While this testing issue is a central problem, the thorny issue is identifying which particular time series to claim is the determinant. In many contexts, after estimating a factor, the investigator is left with a time series which could be any marco-economic variable. A strategy is needed for selecting which variables could be an underlying factor. This issue becomes even more complicated when there are multiple factors because PC estimation yields a rotation of the underlying factors. Any particular variable could be a determinant even when such a variable has somewhat low correlation with each estimated factor.

This paper provides simple and effective solutions to these problems. First, a new method is described for testing if a particular variable is the common factor. Second, by modeling

endogenous common factors, a strategy is developed for picking which variables could be determinants without requiring researchers to pore over the universe of exogenous variables. The performance of these methods is studied both in theory and in practice. Theoretically, the developed procedure correctly identifies the leader when N and T jointly approach infinity under the minimal assumptions of Bai and Ng (2002). Monte Carlo simulation shows that the procedure performs quite well in the finite sample. The procedure is then applied to three empirical examples. The resulting estimation performs very well in practice.

CHAPTER 2

HEAVY TRAFFIC: DETERMINATION OF HOMICIDE RATES ACROSS THE 50 UNITED STATES

Author – Jason Parker

The Department of Economics, GR31

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

ABSTRACT

This paper argues that murder is fundamentally different from other crime rates by using an approximate factor structure estimated by principal components. Using factor number estimation from Bai and Ng (2002), all state-level crime rate panels (e.g.,, murder, burglary) in the U.S. are found to have single common factors after serially demeaning. However, cross-sectionally demeaning reveals zero common factors in every panel except murder implying that the national average is an excellent proxy for the common factors apart from murder. Common factor identification developed in Parker and Sul (2013) reveals that the determinants for the murder rate factor are the murder rates of California and Texas. This is taken as evidence that drug-trafficking is the source of cross-sectional dependence of state-level murder rates.

2.1 Introduction

U.S. crime rates exhibit serious cross-sectional dependence. In fact, apart from murder, this paper finds that all U.S. state-level crime rate series are correlated with each other. When the burglary rate in California increases, the burglary rate in Maine – which is 3,000 miles away – is also likely to increase. To some extent, this result is not puzzling at all. Many empirical studies have tried to explain state-level crime rates using common causes. If state-level crime rates can be well-understood, it is assumed that the overall national crime rate will be well-explained. For example, Levitt (2004) identifies four factors that explain the decline: the number of police, rising prison populations, the receding crack epidemic, and the legalization of abortion. If all states share these four trends, then it is obvious that crime rates between states should be correlated. The puzzle then becomes why murder rates are not highly correlated between states. In other words, the conventional factors may not effectively explain the movement of murder rates across states.

To investigate the source of the puzzle, this paper takes a different empirical strategy from the conventional regression method. In these conventional regression models, crime rates become dependent variables while economic, social, demographic, and political variables become potential causes of crime (e.g., Moody and Marvell 2010). The conventional methods treat the cross-sectional dependence as nuisance information which may hamper statistical inference rather than as important, additional information. This paper considers cross-sectional dependence to be the most important ingredient of crime panel data. Under the approximate common factor representation, an unknown common factor becomes the source of this cross-sectional dependence. By estimating the unknown common factors and their influences (factor loading coefficients), the variation of the panel data can be decomposed into the variation due to common components and that due to purely individual-specific (or idiosyncratic) components. When the common components are more dominant, the degree of the cross-sectional dependence in the panel data becomes more serious.

By analyzing cross-sectional dependence in crime rates, this paper finds the following important economic implications: First, all crime rates including the murder rate have single common factors. Second, except for the murder rate, the impact of the unknown common factor on the state-level crime rates is extremely similar between states, after controlling for fixed effects. In other words, the national average of the crime rates except for the murder rate can explain the state-level crime rates fairly well. As the crime rate at the national level increases, the state-level crime rate in every state is likely to increase as well. However, for the murder rate case, such homogeneous influence by the national average does not apply. That is, an increase in the national average murder rate does not necessarily mean the increase in the murder rate in a particular state.

The approximate common factor representation has been exponentially used mainly in macroeconomic and financial studies where the common factors usually need to be identified. A few examples follow. Stock and Watson (2005) uses a dynamic factor structure to explain common shocks in 132 macroeconomic time series. To improve forecasting, Bernanke, Boivin, and Eliasz (2005) uses factor analysis as a data reduction technique to model how central banks read market signals when making policy choices. However, as long as the common factors are unknown, the source of the cross-sectional dependence is also unknown. A recent paper by Parker and Sul (2013) overcomes this issue and suggests a simple and novel identification strategy for the unknown common factors. This paper adopts their method to identify the unknown common factors to the murder rates.

This paper finds that, surprisingly, the murder rates in California and Texas explain the factor estimated by principal components very well. Meanwhile, in other crime rates, there are no such endogenous factors. To understand the common factor for murder, the homicide rate in California and Texas must be analyzed and understood. There is something special in Texas and California, which makes these states particularly interesting. These states are both known for being major points of illegal crossing from Mexico to the U.S. for many

people and goods, most notably, for drug smuggling operations (Andreas 2009). There are many other locations where smuggling could occur, e.g., Florida. Yet, because these routes are quite direct from South America to drug markets in the U.S. and because these routes are not heavily policed (in number and in punishment), Texas and California have become extremely popular points-of-entry for drug-traffickers.

More interestingly but shockingly, the influence of the murder rate factor on each state is very similar to the drug-trafficking routes. The murder rates in those states on the routes are highly correlated with the murder rates in Texas and California. As many empirical studies found, drugs influence many other types of crime. Directly and indirectly, using drugs causes crime. However, this paper considers the impact of drugs more widely, particularly on the murder rate. That is, the murder rates are high wherever drugs are carried through a state. Due to the difficulty of obtaining quantitative data for drug-trafficking, this paper uses the rather crude, inter-ocular trauma test. To support such crude empirical evidence, this paper investigates further to find the source of the murder rates inside cities in Texas and California. If the conjecture is correct, then the well-known centers for drug distribution – Dallas/Fort Worth and San Diego – should be responsible for the murder rates of Texas and California cities, respectively. The empirical results in this paper pinpoint that the murder rate in Fort Worth and the murder rate in San Diego are equivalent to the murder factors in Texas and California.

The remainder of the paper is organized as follows: Section 2.2 provides initial evidence for a single common factor in each crime panel. Also, for crime rates besides murder, the common factor is shown to be well-explained by the national average or cross-sectional average of all the states in the U.S., indicating that murder is fundamentally different from other types of crime. Section 2.3 uses leadership estimation developed in Parker and Sul (2013) to pin-down the factors endogenously, finding that Texas and California are the leaders for murder rates. Section 2.4 shows how drug-trafficking can potentially explain the leadership of these key states. Section 2.5 concludes.

2.2   Anatomy of Cross-Sectional Dependence in Crimes

This section consists of three parts: First, this section describes the structure and economic meaning behind common factor modeling. Next, factor number estimation is used to show that all crime rates have one common factor. Also, factor number estimation can be used to show that other crime rates besides murder can be well-explained by their national averages. Lastly, the murder and burglary factors are distinguished by use of variance decomposition to show further evidence that murder is different from other crime rates.

Approximate Common Factor Model

The classical common factor or principal component (PC) analysis has been popularly used in psychology where the independent variables are usually unknown, and in social science where data reduction decreases the number of dependent variables used to specify particular models. Recently theoretical development in panel data econometrics allows researchers to overcome some technical difficulties when estimating the unknown number of common factors and to perform statistical inference on the estimated common factors. These rapid developments have enabled empirical scientists to apply the approximate common factor model to various areas including finance, macroeconomics and public economics.

An explanation of the common factor model and the notation used in this paper follows. Let y_{it} be a crime rate of the ith state at time t where $i = 1, ..., N$ and $t = 1, ..., T$. The cross-sectional dependence among y_{it} can be modeled by the following approximate common factor structure:

$$y_{it} = \lambda_i F_t + \varepsilon_{it}, \tag{2.1}$$

where the number of common factors is assumed to be one for simplicity. Note that the individual fixed effects are not expressed in (2.1) for notational convenience. The inclusion of fixed effects does not alter the results at all. The common factor, F_t, is the source of

the cross-sectional dependence. The factor loading coefficients, λ_i, represent the economic distance between the common factor F_t and y_{it}. The product of the two, $\lambda_i F_t$, is called the common component. The left-over term becomes the individual specific or idiosyncratic term, ε_{it}. It is typically assumed to be independent of the common component and exhibits, at most, weak cross-sectional and serial dependence. The idiosyncratic term represents state-specific crime variation, i.e., variation which happens in one state and none of the others. In contrast, the common component represents variation which happens in all states together. The idiosyncratic component and the common component together represent all of the variation in crime rates.

The common factor affects each state but to a different degree. A rise in the national average has a large impact on states with high λ_i values and a smaller impact on states with λ_i close to zero. The loading may be related to some important characteristics of the state, for instance whether the state is more or less rural. The degree of correlation between any two states is determined by their $\lambda_i F_t$ values. That is, the correlation between any two states (i and j) becomes

$$E\left[y_{it}y_{jt}\right] = E\left(\lambda_i \lambda_j F_t^2\right) + E\left[\varepsilon_{it}\varepsilon_{jt}\right] = E\left(\lambda_i \lambda_j F_t^2\right),$$

since usually the cross-sectional dependence among idiosyncratic components can be ignored or assumed to be zero.

The influence of the common factor on each individual state can be measured by the following variance decomposition.

$$V\left(y_{it}\right) = V\left(\lambda_i F_t\right) + V\left(\varepsilon_{it}\right), \tag{2.2}$$

since typically the correlation between the common and idiosyncratic components is small enough to be ignored. This variance decomposition explicitly shows how variation in the factor explains variation in the state-level crime rate.

Because F_t is unknown, (2.1) cannot be estimated by least squares regression. Hence, F_t is typically estimated by the PC of y_{it}, then (2.1) is estimated by ordinary least squares substituting $\hat{F}_t$ for F_t. The PC of y_{it} are the largest eigenvectors of the estimated correlation matrix of y_{it}. The PC are standardized so that F_t does not become unworkably small as T becomes large.

Evidence from Factor Number Estimation

The data consists of state-level crime rates from the FBI's Uniform Crime Reports (UCR hereafter) from the years 1971 to 2010 for the 50 United States.[1] The District of Columbia is not included in the panel as is the prevailing standard in the literature.[2] The crime rates (defined as number of crimes per 1000 persons) are considered for the following types of crime: murder and non-negligent manslaughter, forcible rape, aggravated assault, robbery, burglary, larceny theft, and motor vehicle theft.

The factor number is estimated by using two of the most popular methods: Bai and Ng (2002)'s IC_2 and Hallin and Liska (2007)'s IC_2 criteria. They will be called BN and HL, respectively, throughout the paper. BN usually performs better than HL with a small N and T sample, while HL performs relatively better than BN in samples with large N and T values (Parker and Sul 2013). For almost all cases, the BN and HL criteria estimate the same factor number. However, whenever BN and HL estimate different numbers of factors, the largest result is used in this paper. Later, the Parker and Sul (2013) estimation will be discussed, which relies heavily on factor number estimation. Using the larger factor number

---

[1]All types of the crime data are available from 1965. However the first 6 years of the data were not included in the sample in order to obtain the robust empirical results. The inclusion of the first six years affects the factor number estimation. For example, in the case of robbery, the estimated factor number becomes heavily dependent on the choice of the starting year when the data starts from 1965. To avoid such sensitive results, we do not include the first six years.

[2]D.C. crime rates behave much like a city or metropolitan area. Crime is significantly higher in D.C. than any other 'state' series. In 2010, the murder rate for D.C. was .22 murders per 1000 people, almost double that of any state.

naturally reduces the false rejection rate when the null is true. Hence, this cautious decision rule is very conservative.

For factor number estimation, the natural logarithm of each time series of data is first differenced then standardized by its variance. Table 2.1 shows the estimation result. Surprisingly, all crime rates have single common factors.

Table 2.1. Factor Number Estimation for Various State-level Crime Rate Panels

Crime	Factor Number Estimation	
	Before taking off cross-sectional averages	After taking off cross-sectional average
Murder	1	1
Rape	1	0
Robbery	1	0
Assault	1	0
Burglary	1	0
Larceny	1	0
Auto-theft	1	0

While common factors are rarely included in crime regressions, many applied papers have included common time effects (e.g., Donohue and Levitt 2001; Moody and Marvell 2010). Common time effects are common factors where the factor loadings are perfectly homogeneous. It is natural to wonder whether or not common time effects can account for the single common factor present in all these crime data sets. This can be established by testing for homogeneous factor loadings. If the factor loadings are homogeneous, then the following holds:

$$\tilde{y}_{it} = y_{it} - \frac{1}{N} \sum_{i=1}^{N} y_{it} = (\lambda_i - \lambda) F_t + \varepsilon_{it} - \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{it} = \varepsilon_{it} - \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{it} \text{ if } \lambda_i = \lambda \text{ for all } i.$$

That is, removing the cross-sectional average from the original series eliminates any cross-sectional dependence fairly well. Hence, the factor number in $\tilde{y}_{it}$ naturally becomes zero. The results are also shown in Table 2.1.

Evidently, after controlling for common time effects, all crime rates have zero common factors except for murder. This result implies that simple national averages of crime rates (besides murder) represent the common behavior of these rates very well, even at the state-level. To confirm whether or not the results in Table 2.1 are robust, the same estimation procedures are repeated by changing the starting sample point from 1971 up to 1974 and the same results are obtained.

Evidence from Variance Decomposition

Variance decomposition shows how the factor affects each state series in a manner which is easily interpretable. Here, the common and idiosyncratic variances are reported as ratios to the total variance in (2.2). Because the correlation between the factor and the idiosyncratic component is assumed to be small, the common and idiosyncratic variance ratios sum to one. Note that the variance decomposition is different for each state series because each state has different loadings and idiosyncratic errors. The common variance ratio indicates how well the factor accounts for variation in a particular state. The reader may find it easier to interpret the common variance ratio as an R^2 in the regression of the state crime series on the estimated factor.

Table 2.2 reports selected cases of the variance decomposition for murder rates. The average common variance decomposition ratio (Average Value) for murder rates is relatively small, only 16.8%. This indicates that most murders are likely to be uncorrelated. Cross-sectional dependence probably arises from one specific 'sub-genre' of murder. Later, this paper argues that the source of cross-sectional dependence in murder rates is drug-trafficking.[3] The three highest, lowest, and middle common component variance decomposition ratios are also reported in Table 2.2, ranked by their common variance ratios. The highest states have

[3]Note that in Levitt (2004), Steven Levitt states that he believes that 17% of murders are caused by drug-trafficking. While Levitt does not explain his exact reasoning (as the paper is a perspectives piece), the estimated variance decomposition here would seem to verify his estimate if the factor is indeed drug-related.

Table 2.2. Variance Decomposition for State-Level Murder Rates

State	Rank	Variance Decomposition	
		Common	Idiosyncratic
Average Value	n.a.	16.8%	83.2%
California	1	59.1%	40.9%
Texas	2	58.6%	41.4%
Pennsylvania	3	46.5%	53.5%
Washington	24	13.5%	86.5%
Massachusetts	25	12.7%	87.3%
Arizona	26	11.9%	88.1%
Montana	48	0.2%	99.8%
Idaho	49	0.1%	99.9%
Alaska	50	0.0%	100.0%

The data consists of a UCR sample from 1971-2010. A single factor is estimated by PC. The sample is ranked by correlation to the factor or common component value.

common variance of almost 60% while the lowest states have common variance of essentially 0%. This pattern is consistent with the findings in Section 2.2.2 regarding factor number estimation. The murder factor affects each state heterogeneously, indicating a strong factor structure whether or not cross-sectional averages are removed. Further, because the low loading states are so close to zero, it is clear that the murder factor only affects a few key states. States with small loadings (e.g., Idaho and Alaska) are virtually unaffected by the murder rate factor. States like California and Texas drive much of the movement in the national murder rate. The importance of these states may relate to the presence of large cities, since California has many large cities (e.g., Los Angeles and San Francisco) and Texas has two of the five largest metropolitan statistical areas (Dallas and Houston).

Table 2.3 details another variance decomposition but for burglary instead of murder. Table 2.3 is organized similarly to Table 2.2. Burglary is chosen arbitrarily, as all the crime rates besides murder have similar patterns. The complete results for every crime

rate and every state are available on the author's website. With burglary, the average common variance is much higher than that for murder rates. Further, in keeping with Section 2.2.2, all states but Alaska are affected by the burglary factor somewhat similarly. There may be some between-state variation, but no state seems completely uncorrelated with the estimated factor. The factor affects each state similarly, indicating that no particular state is responsible for national movement in the burglary rate.

Table 2.3. Variance Decomposition for State-Level Burglary Rates

State	Rank	Variance Decomposition	
		Common	Idiosyncratic
Average Value	n.a.	49.1%	50.9%
Illinois	1	76.7%	23.3%
Connecticut	2	72.6%	27.4%
South Carolina	3	72.2%	27.8%
California	24	51.1%	48.9%
Nebraska	25	50.2%	49.8%
Maine	26	49.2%	50.8%
North Dakota	48	17.6%	82.4%
Montana	49	11.5%	88.5%
Alaska	50	6.9%	93.1%

The data consists of a UCR sample from 1971-2010. A single factor is estimated by PC. The sample is ranked by correlation to the factor or common component value.

Graphically, the patterns described in Tables 2.2 and 2.3 can be visualized by averaging. When cross-sectional averages (one average per year) are taken of single-factor panels, the resulting time series is quite similar to the common factor (Pesaran 2006):

$$\frac{1}{N} \sum_{i=1}^{N} y_{it} = \left[\frac{1}{N} \sum_{i=1}^{N} \lambda_i \right] F_t + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{it} = \bar{\lambda} F_t + \bar{\varepsilon}_t \tag{2.3}$$

where $\bar{\lambda}$ and $\bar{\varepsilon}_t$ are the averages of the factor loadings and ε_{it} errors, respectively. Because $\bar{\varepsilon}_t$, the average of mean zero errors, is small, the cross-sectional average is quite close to the

factor multiplied by a constant. This will be true regardless of the size of the factor loadings unless the factor loadings add to zero. Figure 2.1 plots the cross-sectional averages for the three states with largest loadings (High Loading), the three states with smallest loadings (Low Loading), and all 50 states (All States). The cross-sectional averages are natural-logged before being demeaned and standardized. Panel A shows this plot for murder rates; Panel B shows this plot for burglary rates.

Panel A: Murder Rates   Panel B: Burglary Rates
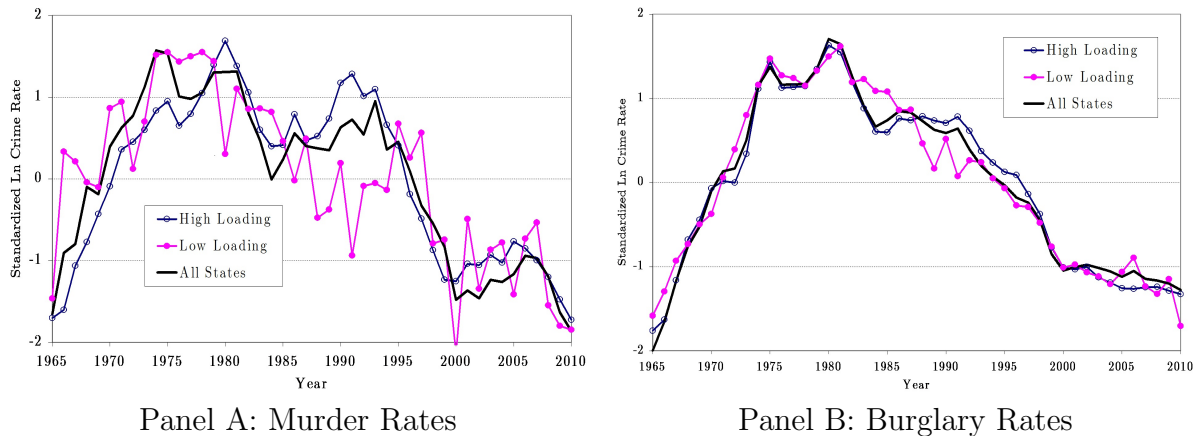
Figure 2.1. Log Cross-Sectional Averages for High Loading, Low Loading, and All States

In Panel B, the cross-sectional averages are all quite similar to each other because the factor affects all 50 states. However, in Panel A, the Low Loading states are extremely different from the High Loading states and All States because the factor does not affect the Low Loading states. The Low Loading series looks like a white noise process because it is an average of a mean zero process. The process seems different from absolute zero because the series is standardized. Therefore, while other crime factors can be well-explained by their national averages (a weighted cross-sectional average), the factor for murder cannot be.

2.3   Common Factor Identification

Parker and Sul (2013) develops a way to test if a particular time series, G_t, is equivalent to the underlying factor, F_t. This section briefly explains this method then applies it to the

crime data. Begin with the usual factor model in (2.1). Now consider running the least squares regression:

$$y_{it} = \beta_i G_t + v_{it}. \tag{2.4}$$

How many factors are in the $\{\hat{v}_{it}\}$ panel? If G_t is equivalent to F_t, Parker and Sul (2013) explains that the usual factor number estimator from Bai and Ng (2002) and the subsample version from Hallin and Liska (2009) consistently estimate that $\{\hat{v}_{it}\}$ has zero common factors. When G_t is substantially different from F_t, $\{\hat{v}_{it}\}$ will have one common factor. Essentially, factor number estimation can be used to test if a given time series is the underlying common factor. This is a type of 'common factor identification.'

This is an interesting result, but the question remains: Where can a researcher find the appropriate time series to test? Parker and Sul (2013) also provides a way to check to see if a particular individual in the panel (time series) is equivalent to the common factor (time series). The method involves regressing the estimated factor, F_t, on each individual, y_{it}, and obtaining the R^2 value. The individuals with high R^2 values are typically called 'candidates.' Then these individuals can be used one at a time in (2.4). That is,

$$y_{it} = b_{c,i} y_{ct} + e_{c,it} \text{ for all } i \neq c \tag{2.5}$$

for each leader candidate $i = c$. The factor number of $\{\hat{e}_{c,it}\}$ is now an estimator for whether or not state c is the factor. If c is the factor, it is called a 'leader.'

The leadership model provides a powerful way to identify common factors so that the cross-sectional dependence has a specific economic interpretation. The leader affects all the individuals in the panel. The amount which it affects them, $b_{c,i}$, can be thought of as an economic distance. Individuals with large $b_{c,i}$ values follow the leader closely. Individuals with small $b_{c,i}$ values are virtually unaffected by the leader.

For the crime data, leadership is only possible for the murder panel. For crimes besides murder, each state series is similar to the others. Section 2.2 shows that every state has the

same factor loading for crimes other than murder. That is,

$$\lambda_i = \lambda \text{ for each state, } i,$$

where λ is an arbitrary constant value. Hence, if a leader is present, each state will be affected by the leader equally. Every state will have the same economic distance to the leader. This is highly unlikely. Suppose that the burglary rate in New York is the leader. New Jersey will have the same distance to New York as Alaska. The leadership model of cross-sectional dependence is incompatible with the pattern present in these other crimes.

For each type of crime, three potential leaders are selected using the highest R^2 value described above. The factor number is estimated as before, i.e., the maximum of the BN and HL estimators. Only the murder panel is found to exhibit leadership (Table 2.4). The leaders for murder are Texas and California, where the residual factor number becomes zero. All the other states and crimes are found to have a single residual factor. Texas and California have R^2 values close to 0.6. The R^2 quickly diminishes for other states. Other crimes show no stark contrasts in R^2 values. These patterns in R^2 validate the leadership findings. The findings confirm the aforementioned theory, that no crime besides murder can have a leader.

As before, subsample analysis of the factor number is performed using five subsamples of varying starting year. The subsample analysis verifies the findings in Table 2.4 in all cases, but only the murder rate residuals for California and Texas are reported below (Figure 2.2). Clearly, California and Texas are leaders in this data.

When two series become leaders in the data, they are called 'approximate dominant leaders.' The leaders are both quite close to the factor, but the factor is a combination of the estimated leaders. Therefore, the estimated murder factor is regressed (least squares) on the murder rates of California and Texas:

Table 2.4. Leadership Estimation for Crime Panels (Bold: Leader; Parentheses: R^2 value)

Crime	Candidates for Leadership		
	1st Candidate	2nd Candidate	3rd Candidate
Murder	**California** (0.606)	**Texas** (0.601)	Pennsylvania (0.477)
Rape	Ohio (0.687)	Texas (0.647)	Florida (0.642)
Assault	Florida (0.724)	Illinois (0.642)	Washington (0.626)
Robbery	Alabama (0.725)	Illinois (0.724)	Virginia (0.646)
Burglary	Illinois (0.786)	Connecticut (0.745)	South Carolina (0.740)
Larceny	New Jersey (0.845)	North Carolina (0.831)	Illinois (0.794)
Auto-theft	North Carolina (0.680)	Arizona (0.647)	Pennsylvania (0.622)

Leadership estimation performed as in Parker and Sul (2013). Bold indicates that zero factors are found in the residual from (2.5). R^2 values reported are the R^2 from the regression of the estimated factor on the individual state-level time series in question.

Panel A: Removing California Result   Panel B: Removing Texas Result
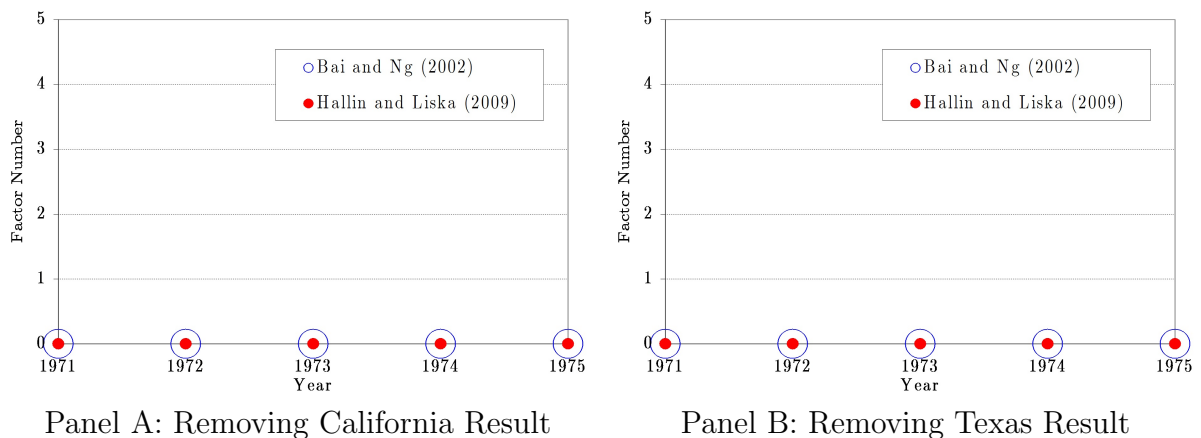
Figure 2.2. Factor Number Subsample Analysis for Leadership in Murder Rates

$$\tilde{F}_t = 0.5002 \times \Delta \ln(\text{California}_t) + 0.4925 \times \Delta \ln(\text{Texas}_t).[4]$$

Essentially, the murder factor is half of the murder rate in California and half of the murder rate in Texas. These series fit the estimated murder factor very well considering that the factor is estimated and the crime rates themselves suffer from some measurement error (Figure 2.3).
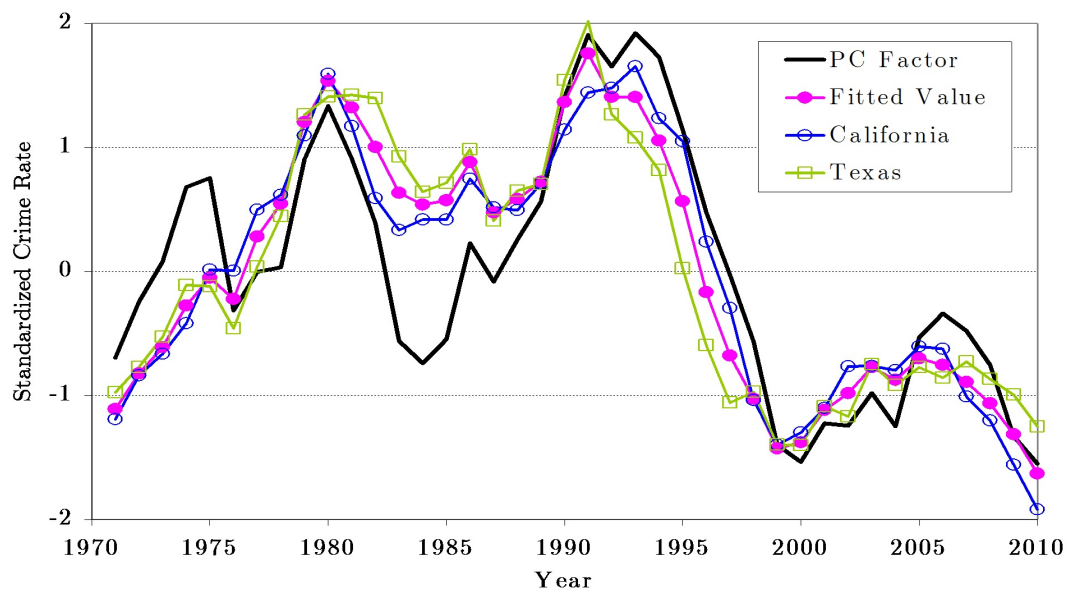
Figure 2.3. Leaders for Homicide Rates and Estimated Factors

2.4 Drug-Trafficking in the United States

Factor identification is useful because it provides an economic interpretation of the underlying factor. This section explains how the leadership of California and Texas can be interpreted as drug-trafficking.

---

[4]Because these series are quite similar to each other, there may be concerns about colinearity. Note that the correlation between the murder rate in California and the murder rate in Texas is only 0.585.

Cross-sectional dependence in murder is confusing in itself. Are extramarital affairs in California somehow related to those in South Carolina? It's hard to believe that many homicides are somehow correlated. Evidence from variance decomposition reveals that only 16.8% of the variation in homicide rates is due to the common factor. This implies that most murders are unrelated, while a few are essentially correlated. Also many of the state series have loadings near zero, which indicates that the cross-sectional dependence is confined to only specific places or specific murders. The question then becomes: Which murders are cross-sectionally dependent?

To pin down an economic explanation for cross-sectional dependence in murder, it is useful to consider what California and Texas have in common. Texas and California are the second and third largest states in the U.S. by land-area. As mentioned before, these two states have many large cities. This could imply that Texas and California are leaders by aggregation. This case is further explained and disproved below. It seems that besides being large, the leadership of Texas and California can only be explained by the border these states share with Mexico. Many illegal aliens cross this border each year. The U.S. Government Accountability Office estimates that 450,000 immigrants per year crossed the border from Mexico to the U.S. illegally in the period from 1998 to 2004 (Government Accountability Office 2006). Because this illegal crossing is so common, it is relatively easy for traffickers to smuggle drugs into the U.S. using these same routes (Andreas 2009). Each state has four of the twenty-eight High Intensity Drug Trafficking Regions as defined by the U.S. Office of National Drug Control policy, and both states are well-known points of entry for drug smugglers.

Drugs can be smuggled into the U.S. in a number of ways. Marijuana is sometimes smuggled by train and ultralight aircraft (Department of Justice 2011). Tunnels connecting Juarez to El Paso are well-known, difficult to detect passages into the country. MDMA is commonly brought into the U.S. using the border with Canada. Some smugglers also bring

drugs into the U.S. by sea and plane, although these methods are becoming less-popular as detection technology improves (Department of Justice 2011). The primary entrance to the U.S. is through the Southwest border where patrol is relatively minimal. Only two states besides Texas and California share a border with Mexico, Arizona and New Mexico. The most direct land-route to the U.S. from Central America is through Mexico and then Texas. California has a large market for drugs and can be easily accessed by taking a boat to Baja California and traveling north on land.

It is difficult to find other evidence for drug-trafficking due to the scarcity of available data. The data which do exist are riddled with measurement error problems because drug-trafficking is an illegal activity (Fryer, Heaton, Levitt, and Murphy 2006). Some data exists for drug usage and the relationship between drug usage and crime. However, it is a mistake to rely on usage data because drug-traffickers are not necessarily users and vice versa. Further, arrests for drug-trafficking are useless because they depend heavily on enforcement levels, police technology, and the transportation method employed by the trafficker. Drug sales are inappropriate because final consumers are not located where trafficking occurs. Essentially, there is no convenient way to correctly pin down an exogenous measure of trafficking.

However, further analysis can be performed to check the leadership of Texas and California and rule out aggregation as a possible explanation for the leadership of Texas and California. Potentially, these states could be leaders because, when only a single factor is present in the data, cross-sectional averages tend to estimate the factor as in (2.3). Because California and Texas have many large metropolitan statistical areas (e.g., Los Angeles, San Francisco, Dallas, Houston), these state-level crime series could be interpreted as cross-sectional averages of the underlying city-level crime series. Therefore, California and Texas might appear to be the factor because of state-level aggregation. To rule out this possibility, the two leader states are disaggregated into the city-level crime rates (one panel per state) and leadership is again estimated as before (Table 2.5). The data are natural logged, de-meaned by the time series average, and standardized so that one common factor is present

in both data sets. UCR data is only readily available and accurate from 1985 to 2010 for city-level crime rates. Fort Worth and San Diego are found to be leaders for within-state murder rates. San Diego is quite close to Mexico (less than 21 mi.) and well-known for its relationship to drug movement (Andreas 2009). Fort Worth is a part of the Dallas Fort-Worth Denton MSA (henceforth DFW), a known nexus of drug distribution. Because these data are disaggregated, this result verifies the claim that leadership is related to drug-trafficking and not within-state aggregation.

Table 2.5. Within-State Murder Leadership

State Panel	Sample Size ($N =$)	Candidate for Leader	Residual Factor Number
Texas	23	Fort Worth	0
California	39	San Diego	0

It may seem logical to use city-level data throughout the U.S. as a further explanation. However, it is impossible to perform this leadership identification correctly because cities around the U.S. have extremely correlated crime rates. Table 2.6 shows the correlation between the four largest cities in the U.S. and previously determined leaders. The data is obtained as before, from the FBI Uniform Crime Reports from 1985 to 2010. Presently, there is simply not enough data for this city-level analysis throughout the U.S. Cross-country leadership analysis is also impossible because of data availability. However, evidence that city-level murders are highly correlated further indicates that drug-trafficking is the source of cross-sectional dependence in murder rates in the U.S. City-level murders could be correlated because trade and distribution of cocaine and heroin are more common in urban areas.

To provide more evidence for the drug-trafficking theory, the estimated factor loadings from the regression in (2.5) are plotted in Figure 2.4. As stated above, the factor loading represent economic distances to particular sources of cross-sectional dependence (here California and Texas). Clearly, the factor has a large influence on more urban states. Drugs

Table 2.6. Correlations to Known Leaders

City-Level	Correlation to	
Murder Rate	Fort Worth	San Diego
New York	0.891	0.969
Los Angeles	0.889	0.912
Chicago	0.703	0.693
Houston	0.813	0.926
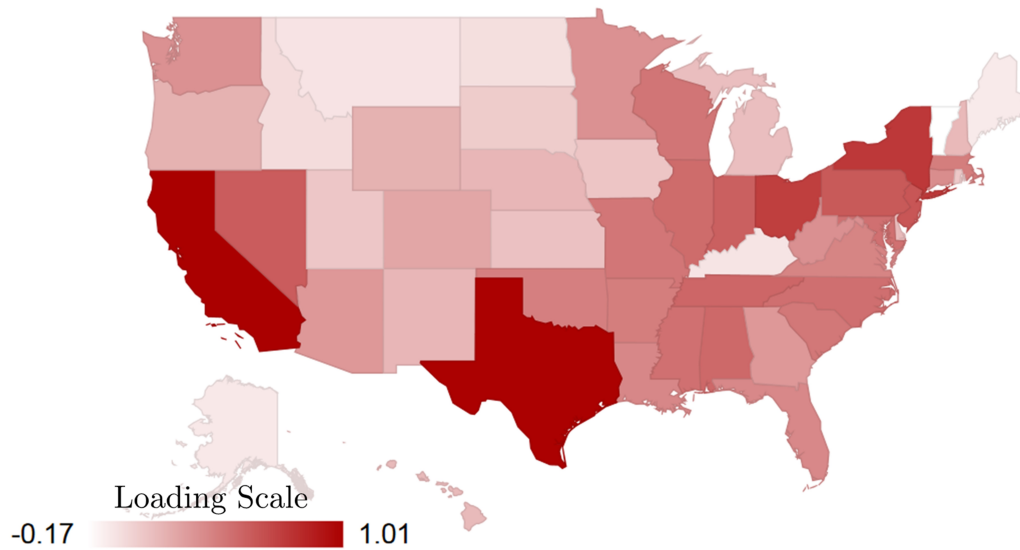
Figure 2.4. Map of Leadership in the U.S.

enter the U.S. from Mexico through the Southwest and travel to the Midwest and Northeast. From Figure 2.4, it is also possible to assess which drug is the source of cross-sectional dependence in states. The factor does not load heavily on rural states, where methamphetamines are typically produced. Further, the factor does not load heavily on all states near Canada which are highly influenced by the trafficking of MDMA.

Figure 2.5 shows a map of marijuana trafficking reproduced from the 2011 Drug Threat Assessment. The Drug Threat Assessment is a source of information which was published annually by the U.S. Department of Justice until 2011. The data for this map are collected from local law enforcement agencies concerning arrests and other information on major

routes drug-traffickers used to move product to market from 2008 to 2010. From eyeball estimation (also known as the inter-ocular trauma test in some literatures), it is possible to discern that the primary routes seem to correspond with high loading areas from Figure 2.4. Low loading areas such as Kentucky, Montana, and Vermont do not have significant drug trafficking. Further, DFW and San Diego seem to be the most important cities in Figure 2.5, cities which were found to be important in Table 2.5. Therefore, it seems quite plausible that cross-sectional dependence in murder is strongly related to marijuana trafficking.
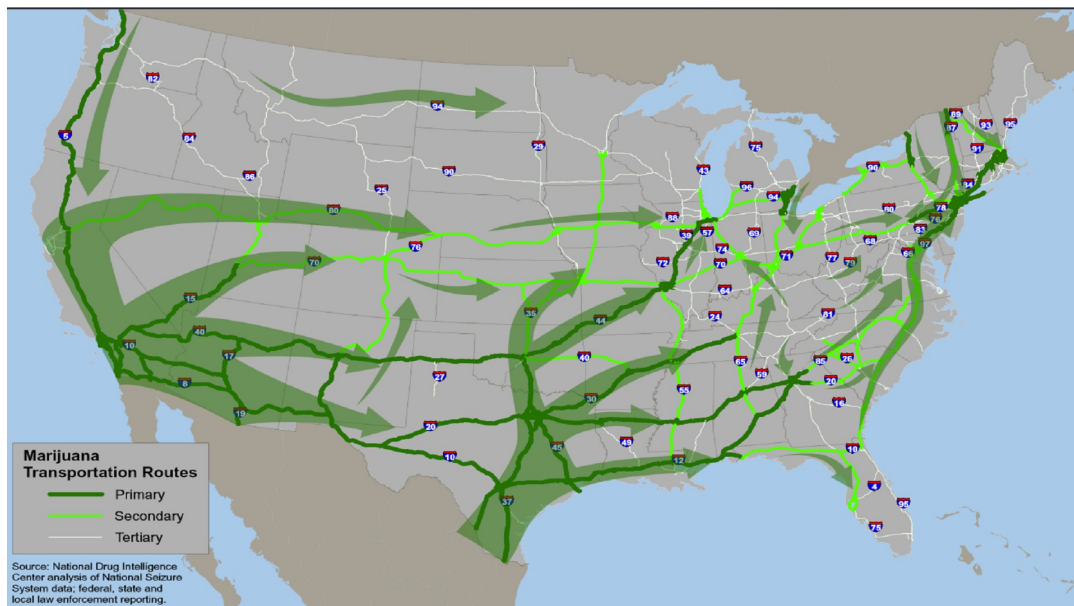
Figure 2.5. Map of Marijuana Trafficking in the U.S. (Reproduced from the 2011 Drug Threat Assessment)

Other drug-trafficking routes are plotted in Figure 2.6. Panel A shows primary routes for heroin, which upon careful examination are somewhat different from the factor loadings. For instance, heroin does not have many primary routes through the U.S. Upland South. However, the estimated factor loadings are somewhat large in the Upland South, with the exception of Kentucky. Panel B shows cocaine routes which are strong around Atlanta and weak around DFW. This pattern is dissimilar to the results from Table 2.5. It must be said that the factor could potentially be related to some weighted average of particular drugs.

This possibility cannot be excluded. However, from Figures 2.5 and 2.6, it appears likely that cross-sectional dependence in murder rates is likely to be caused by marijuana trafficking specifically.
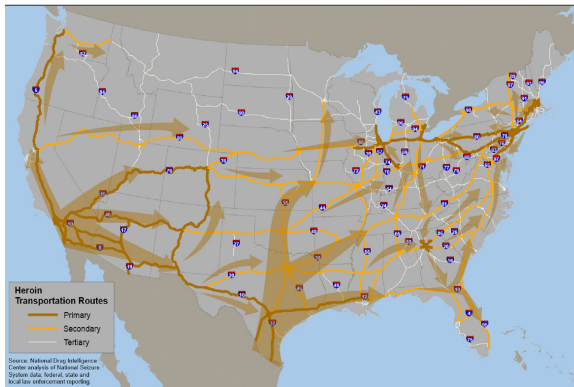
Panel A: Heroin Trafficking   Panel B: Cocaine Trafficking
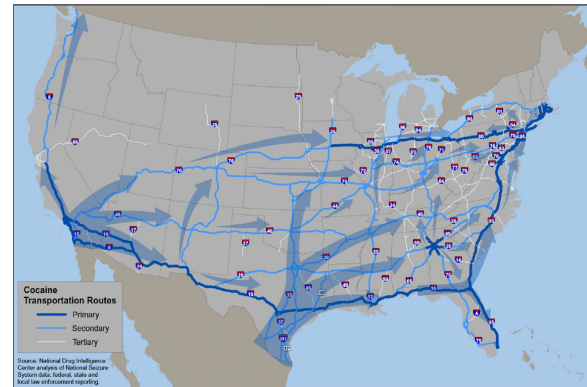
Figure 2.6. Maps of Drug Trafficking in the U.S. (Reproduced from the 2011 Drug Threat Assessment)

As explained above, in single-factor panels, the national average is closely related to the common factor. Therefore, all this evidence shows not only that the common factor for murder is related to drug-trafficking, but also that movement in the national average is caused by fluctuation in drug-trafficking. Further it is likely that the most important drug to murder is marijuana. From all the evidence presented, particularly Figure 2.3, the decline in the national murder rate during the 1990s was likely caused by declining levels of marijuana smuggling.

2.5 Conclusion

This paper models crime rates using the common factor structure and provides a simple explanation for cross-sectional dependence in the murder rate. Other violent and property crimes can easily be explained by their respective national averages. Variation in the national averages is left open for debate. On the other hand, homicide and non-negligent manslaughter are very different. In particular, leadership identification using Parker and Sul (2013)

indicates that Texas and California are the sources of cross-sectional dependence in murder rates. These states share large borders with Mexico and are common sources of illegal entry. This coupled with evidence that city-level murders are extremely correlated indicates that the cause of cross-sectionally dependent murder is drug-trafficking.

The policy implications of drug-trafficking explaining the common factor are enormous. 17% of the variation in homicide rates is explained by the common factor. However, since the common factor affects most of the co-movement, back-of-the-envelope calculations reveal that the common factor accounts for 94.9% of the decline in the national murder rate from 1991 to 2000. Declining murder rates are directly related to declining amounts of drug-trafficking. Through the 90s, agencies like the DEA and PFM battled drug-trafficking in the U.S. and in Mexico by making more arrests and by targeting key players in the drug trade. Famous kingpins such as Pablo Escobar were killed in the 90s by various police forces. The evidence presented here indicates that drug enforcement agencies were extremely successful during the 90s. Their efforts caused serious changes in the murder rate in all major cities across the U.S., even though these changes seemed unrelated at the time. Assuming that drugs remain generally prohibited in the U.S., further funding of the war on drugs (specifically targeting drug-trafficking) may help to prevent future murders in the U.S.

CHAPTER 3

IMPROVED TWO-SAMPLE COMPARISONS FOR LABORATORY DATA

Author – Jason Parker

The Department of Economics, GR31

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

ABSTRACT

This paper details the problems of the Wilcoxon-Mann-Whitney (WMW) rank-sum test for comparing two independent samples and proposes a new testing procedure for comparing distributions in laboratory experimental games. The WMW test is commonly misinterpreted as a test of means, when the test actually distinguishes higher-moment differences as evidenced by Monte Carlo simulation. As a distribution test, WMW performs quite poorly in terms of power of the test. Further, because of this poor finite-sample performance, many researchers use unconventional methods, such as pooling rounds of repeated games which is shown to be inconsistent. A new testing procedure is developed using the beta distribution which has better power than WMW. This sequential testing procedure has the added advantage that it enables researcher to not only show that the distributions are different, but also explain how the distributions have changed between samples, using the new definition of distributional shifting. This new procedure is proven consistent theoretically and the finite-sample performance is shown to be better through Monte Carlo simulation and application to a canonical empirical example.

3.1 Introduction

Comparing two independent samples is straightforward in theory. Many test statistics have been developed and used in practice. When the sample size is moderately large, a central limit theorem can be employed to compare the central tendencies of the two samples. Such procedures do not require normality assumptions and are simple enough to be taught in undergraduate statistics courses. These results work quite well when the cost of collecting data is relatively cheap. However, in lab experiments, collecting data is expensive and collected data is often asymmetrically distributed. Limited budgets lead to small sample sizes, where parametric form becomes important. Hence, more powerful tests are needed.

Wilcoxon (1945) develops two statistical tests for two-sample comparisons, the rank-sum test and the signed-rank test. The limiting distribution for the rank-sum test is derived in Mann and Whitney (1947). This Wilcoxon-Mann-Whitney rank-sum test (hereafter referred to as WMW) is the most common statistical test used for comparing the distributions of two *independent* samples in the experimental literature. Despite its prominence, WMW has many problems, or pitfalls, which are investigated in Section 2.2 of this paper. First, WMW has low power – meaning that it struggles to reject false nulls unless the sample size is large. WMW particularly struggles to detect differences in second or higher moments as seen by Marsh (2010) and others. Second but related, WMW is commonly misinterpreted. WMW should only be interpreted as a test of whether or not two samples are drawn from the same distribution. The null is the same distribution; the alternative is that the distributions are different. Rejection of the null could result from the distributions having different means, variances, or other higher moments. However, because Marsh (2010) and others have found that WMW is poor at detecting higher-than-first moment differences, WMW has commonly been interpreted as a test of means. The test is treated like a general z-score test, when reality paints a very different picture. Also, because WMW has low power, many researchers apply the test in uncommon ways, e.g., pooling many rounds of repeated games before

testing, which usually results in inconsistent estimation. Altogether, the most important issue is interpretation. If WMW detects higher moment differences, identifying an economic meaning for the test becomes challenging. Findings from WMW that one distribution is somehow "right or left" of another distribution could be quite misleading.

To overcome these issues, this paper provides an alternative statistical procedure for experimentalists which is robust, simple, less restrictive, and more powerful. For the goal of detecting whether or not two samples are derived from the same distribution, this paper contains information about a new distribution test which is quite powerful. However, when the goal is more general (i.e., to understand *how* two distributions are different), a new testing procedure is needed for comparing a few different characteristics of the distributions.

This 'sequential testing procedure' consists of the following three steps: The first step is to pre-test if the distributions are different between samples. WMW can be used for testing the first null hypothesis, yet because the power of WMW is low, this paper recommends using the new distribution test. If the first null of equal distributions is rejected, then there is evidence that the two samples are drawn from different distributions – differences which can be investigated. The next suggested step in testing for shifting is to test the null hypotheses that each distribution is symmetric. This constitutes two tests, testing symmetry for each sample separately. If the distributions are symmetric, then the mean, median, and mode (if it exists) are all the same. If one distribution is symmetric and the other is not, this implies a specific change in distribution which can be interpreted depending on the context. Rejection of the first null but non-rejection of the second null (in both cases) implies that the samples have heterogeneous means and/or variances. In practice this second step can be omitted entirely because many distributions are clearly asymmetric. For example, in ultimatum games, offers are usually right skewed – more offers are in the right tail. Similarly in public goods games, contributions to the public account are usually right skewed. The third null hypothesis is designed to test whether or not one distribution "moves" to the left or right

relative to the other distribution. This notion of movement is defined here as 'shifting' – a change in skewness – and the corresponding null hypothesis is that both distributions have equal skewness (non-shifting). If the equal distribution null is rejected but the non-shifting null is not rejected, then the primary difference between the two samples is heterogeneous variance. When both nulls are rejected, then one distribution has shifted to the left or to the right of the other distribution.

To perform the sequential tests, estimation is performed using the beta distribution. This 'quasi'-parametric approach is appropriate for a variety of reasons. First, almost all lab data are bounded and can be transformed to lie between 0 and 1. For example, in ultimatum and dictator games, subjects bargain over a fixed pie. If the size of the pie is normalized to 1, then all of the offers are bounded between 0 and 1. Also, in public goods games, the contributions to the public account can be normalized to be a percentage of the income or endowment. Of course there are many distributions which support the domain between 0 and 1 such as the Kumaraswamy and uniform distributions. However, the beta distribution is most commonly used in practice. Second, nested within the beta distribution are many other common statistical distributions such as the chi-squared, uniform, and Kumaraswamy distributions. Thus, the beta distribution can take a wide variety of density shapes. In this sense, this is labeled the 'quasi'-parametric approach. Lastly, the beta distribution is characterized by only two unknown parameters, namely α and β. Thus, building various testing statistics for the sequential procedure becomes straightforward. That is, the limiting distributions of all of the statistics proposed in this paper are based on the limiting distributions of the method of moments estimator for the two key parameters.

Under the assumption that two samples are generated from two independent beta distributions, three statistics are provided to test the sequential null hypotheses. To be specific, to pre-test that the distributions are the same, a simple χ^2 test is constructed for testing if the latent parameters from the beta distribution, α and β, are the same. To test the other

hypotheses, simple z-score tests are constructed for testing the symmetry of each empirical distribution and the degree of skewness between the two distributions. Later, it is shown that the sequential beta tests provide powerful tools to compare two random samples in practice. Note that the independence assumption can be relaxed by constructing a bivariate beta distribution but at this stage, a general bivariate beta distribution is not available. If these assumptions are not appropriate, remarks in Section 3.3 detail nonparametric and paired sample alternatives which can be applied.

The rest of the paper proceeds as follows. The next section details the motivation for this paper: the problems of WMW and some important empirical examples where the WMW test is not ideal. In Section 3.3, the new sequential testing procedure is developed theoretically with detailed explanations on how to apply the method. Section 3.4 studies the finite sample performance of the various tests to show that the sequential testing procedure is very accurate in small samples. Returning to the canonical empirical examples, Section 3.5 details how the sequential testing procedure can be used to better explain the data. Section 3.6 concludes.

3.2 Motivation and a Canonical Empirical Example

This section explores an empirical example where the standard WMW test is not effective. Then, theoretical criticisms of WMW are developed.

Canonical Empirical Example

The canonical example uses data from Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991 RPOZ hereafter), which is perhaps the most famous and exemplary article from the ulti-matum game literature. The experiment is a traditional ultimatum game where 2 players bargain over a fixed pie. First, one player (the proposer) makes an offer (s) to a responder. The responder observes the offer then chooses whether to accept or reject it. If the responder rejects the offer, both players earn nothing. If the responder accepts, she earns s and

the proposer earns $1 - s$. This game was repeated over 10 rounds. The game was played in 12 sessions of 10 players each in different countries: Israel, Japan, Yugoslavia, and the U.S., three sessions per country.[1] For each country, the sample sizes, means and standard deviations for the first round (First), last round (Last), and time series average (TS Avg) are reported in Table 3.1.

Table 3.1. Sample Statistics for RPOZ (1991)

Country	Sample Size	Mean			Variance x 10^2		
		First	Last	TS Avg	First	Last	TS Avg
Israel	30	0.363	0.331	0.342	2.465	0.846	0.738
Japan	29	0.446	0.402	0.410	4.452	0.518	1.145
Yugoslavia	30	0.442	0.439	0.425	0.738	0.336	0.281
US	27	0.447	0.444	0.430	0.922	0.518	0.303

From Table 3.1, the sample mean offers are smallest in Israel and largest in the US. In the time series average data, Japan has the largest sample variance and Yugoslavia has the smallest sample variance. In every round, Yugoslavia and the US seem to have the similar sample means and variances. The main interest in RPOZ was to compare the offers between pairs of countries. Due to the serial dependence of the offers over rounds, RPOZ emphasized that the WMW test should be performed round-by-round, but particularly the last round sample was used. Meanwhile others typically use the time series averages over rounds (for example, Ho and Su, 2012).Thus, the results of the WMW tests using two different sampling data – the last round and TS average samples – are reported in Table 3.2. Note that the p-values are in parentheses. The first column shows the replication results of RPOZ, which confirms their conclusion. Yet interestingly, in the TS Avg case, the results are much weaker. The offers in Israel do not share the same distribution with offers from Japan, Yugoslavia

---

[1]RPOZ (1991) has 4 sessions for the US. In the data obtained from RPOZ, two of the US sessions are found to be exactly the same. Therefore, one sample was dropped from the subsequent analysis. This may make our results slightly different from an analysis performed with the entire data.

Table 3.2. Between-Country Comparisons for the WMW and Z-Score Tests (p-value)

Samples	WMW		Z-Score	
	Last	TS Avg	Last	TS Avg
Israel v Japan	-3.183 (0.001)	-2.593 (0.005)	-3.315 (0.000)	-2.692 (0.004)
Israel v Yugoslavia	-4.711 (0.000)	-3.823 (0.000)	-5.437 (0.000)	-4.464 (0.000)
Israel v US	-4.578 (0.000)	-3.860 (0.000)	-5.232 (0.000)	-4.625 (0.000)
Japan v Yugoslavia	-2.012 (0.022)	-0.796 (0.213)	-2.150 (0.016)	-0.636 (0.262)
Japan v US	-2.354 (0.009)	-1.246 (0.106)	-2.204 (0.014)	-0.859 (0.195)
Yugoslavia v US	-0.661 (0.254)	-0.496 (0.310)	-0.323 (0.373)	-0.369 (0.356)

and the US, but among these three countries the distributions do not seem to be different according to the WMW test. Such different results would have a strange interpretation: a difference in the last round which doesn't appear 'overall.' The correct interpretation is that WMW cannot distinguish between TS Avg samples from Japan and Yugoslavia or Japan and the US because WMW lacks power, even though the samples are obviously different.

Notice that the traditional two sample z-score test performs similarly to WMW in all of these cases. The z-score test is a test for different means. Since the means of offers from Japan, Yugoslavia, and the US are roughly equal in the TS Avg case, the z-score test is not expected to reject the null. Many papers have shown that WMW struggles (in terms of power) to distinguish higher-than-first-moment differences (e.g., differences in variance, skewness, kurtosis, etc.). Since WMW performs similarly to the z-score test, WMW is commonly interpreted as a test of means. This paper shows below that this interpretation is mistaken. Here, in the TS Avg case, offers from Japan have different variance than offers from Yugoslavia and the US, but WMW and the z-score test both cannot reject their null hypotheses. There could be differences in skewness as well (even though WMW does not reject); this possibility is explored below.

Pitfalls of the WMW Test

WMW is used to test if two independent, random samples, $\mathbf{x} = (x_1, \ldots, x_{n_x})'$ and

$\mathbf{y} = \left(y_1, \ldots, y_{n_y}\right)'$ are drawn from the same distribution, i.e.,

$$\mathcal{H}_0 \ : \ F_x\left(x\right) = F_y\left(x\right) \ \text{for all } x \in \mathbb{R}, \tag{3.1}$$

$$\mathcal{H}_1 \ : \ F_x\left(x\right) \neq F_y\left(x\right) \ \text{for all } x \in \mathbb{R}, \tag{3.2}$$

where $F_x : \mathbb{R} \to [0,1]$ and $F_y : \mathbb{R} \to [0,1]$ are cumulative distribution functions (CDFs) for X and Y respectively. Accordingly, the rejection of this null hypothesis does not imply that X and Y have different means. However, the converse is true. That is, the rejection of equal means always implies rejection of the null hypothesis in (3.1) since the cumulative distribution function must be different if the means are different. Similarly, the rejection of the equal variances also always implies the rejection of the null in (3.1). Therefore the WMW test is very conservative in the sense that this test examines whether or not two independent samples share the same probability distribution.

Monte Carlo studies by Marsh (2010) and Fay and Proschan (2010) showed, that the WMW test struggles to detect differences in variance even when the sample size n is very large. Such results lead Fay and Proschan (2010) to conclude that the WMW test is only useful for detecting mean-differences, which is the equivalent to the standard z-score test. In practice, many researchers state that they use the WMW statistic to test for equal means. Thus, statistically, according to Marsh (2010) and Fay and Proschan (2010), the pitfall of the WMW test is the lack of power with heterogeneous variances. Is it true that the WMW test does not detect the differences in variance? This paper re-examines this issue by means of Monte Carlo simulation. In contrast to Marsh (2010), a bounded but asymmetric distribution is considered as the data generating process. That is, two independent samples from the beta distributions are generated by,

$$x_i \sim Beta\left(\alpha_x, \beta_x\right), \ y_i \sim Beta(\alpha_y, \beta_y).$$

For a beta distributed sample with parameters α and β, the mean and variance become $\alpha/\left(\alpha + \beta\right)$ and $\alpha\beta/\left[\left(\alpha + \beta\right)^2 \left(\alpha + \beta + 1\right)\right]$, respectively. In every case, $E\left(x_i\right) = E\left(y_i\right) = \mu$.

Also note that $\alpha_k = (1/2)\beta_k$, $\alpha_k = (2/3)\beta_k$ and $\alpha_k = \beta_k$ with $\mu = 1/3$, $\mu = 0.4$ and $\mu = 0.5$ for $k = x$ and y, respectively. When μ moves away from 0.5, the distributions become more asymmetric. When $\mu = 0.5$, the distribution becomes symmetric. We let $\alpha_x = 1$ but consider $\alpha_y \in [1, 2, 5, 10]$. The number of replications is 10,000.

Table 3.3 reports the results of the Monte Carlo simulations. Evidently when $\alpha_x = \alpha_y$ and $\beta_x = \beta_y$ (the same distribution), the null of the same CDF rejects at the nominal size regardless of the sample size, n. As the variance of y_i is getting smaller than that of x_i, the rejection rate increases as n increases. Also as the mean is farther from 0.5 or equivalently, when the distribution becomes more skewed, the power of the test becomes strengthened. The symmetric case where $\mu = 0.5$ is first reported where the power of the test becomes very poor as revealed by Marsh (2010). Even overall, the power of the WMW test in small samples is generally not overwhelming. Accordingly, in agreement with Marsh (2010) and Fay and Proschan (2010) but not for the same reason, this paper finds that the WMW test has low power in the finite sample.

Table 3.3. Impact of Heterogeneous Variances on WMW (Nominal Size = 5%, $\alpha_x = 1$)

	μ	α_y	$V(y) \times 10^2$	$n = 25$	50	100	250	500
Size	0.5	1	8.33	0.048	0.051	0.051	0.054	0.049
	0.4	1	6.86	0.048	0.051	0.051	0.054	0.049
	1/3	1	5.56	0.048	0.051	0.051	0.054	0.049
Power	0.5	2	5.00	0.051	0.055	0.054	0.054	0.055
	0.5	5	2.27	0.061	0.062	0.063	0.067	0.066
	0.5	10	1.19	0.072	0.071	0.073	0.074	0.077
	0.4	2	4.00	0.058	0.063	0.071	0.099	0.139
	0.4	5	1.78	0.072	0.088	0.113	0.192	0.318
	0.4	10	0.92	0.088	0.107	0.137	0.235	0.400
	1/3	2	3.18	0.061	0.071	0.092	0.158	0.261
	1/3	5	1.39	0.092	0.127	0.193	0.380	0.620
	1/3	10	0.72	0.112	0.160	0.244	0.472	0.738

Being overly conservative is not a problem if WMW rejects the null. Yet because of this lack of power, the test is commonly performed in non-standard ways to find significant

results. For instance, in many experiments data is pooled over time. Instead of applying the test to one round of data, researchers stack all the rounds together, which is shown here to be inappropriate. Looking to assess the validity of a few different procedures, the following data generating process is developed,

$$x_{it} = \rho x_{i,t-1} + u_{it}, \qquad y_{jt} = \rho y_{j,t-1} + e_{jt},$$

where x_{it} is the offer of player $i \in \{1, \ldots, n\}$ in round $t \in \{1, \ldots, T\}$ and y_{jt} is the offer of player $j \in \{n+1, \ldots, 2n\}$ in round $t \in \{1, \ldots, T\}$, so that the data is serially dependent which is common in repeated experimental data. The variances of u_{it} and e_{jt} are defined as follows:

$$u_{it} \sim \mathcal{N}\left(0, 1 - \rho^2\right), \text{ and } e_{jt} \sim \mathcal{N}\left(0, 1 - \rho^2\right),$$

so that the variances of x_{it} and y_{jt} are always equal to unity. The simulations were replicated 10,000 times and the results are shown in Table 3.4. The last round and time series average cases do not depend upon ρ so it is not included in the tables (every result is the same).

Table 3.4. Impact of Serial Dependence on the WMW Test (Nominal Size: 5%)

T	n	Last Round	TS Avg	Pooled Sample		
				$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.95$
10	25	0.05	0.05	0.22	0.40	0.50
10	50	0.05	0.05	0.21	0.39	0.51
10	100	0.05	0.05	0.21	0.40	0.51
25	25	0.05	0.05	0.24	0.47	0.64
25	50	0.05	0.05	0.23	0.46	0.63
25	100	0.05	0.05	0.23	0.46	0.63

Notice that x_{it} and y_{jt} have the exact same distributions. Therefore, WMW should accept the null in 5% of the cases (i.e., size $< 5\%$). Using only the first round of data and taking time series averages are both well-behaved. In contrast, pooling has massive size distortion, even when the number of rounds is only 10. As the number of rounds increases,

this size distortion continues to increase to one. Similarly, as the data becomes more serially correlated (ρ increases), size distortion increases. Laboratory data typically has serious serial correlation because each player is correlated with their 'future-selves.'

In summary, the WMW test was designed to test whether or not two independent samples share the same CDF. The use of dependent samples invites false rejection very often even when the null is true. Also, the WMW test suffers from lack of power in the finite sample. In the next section, an alternative testing procedure is proposed, which is more powerful that the WMW test.

3.3   Alternative Sequential Tests

Experimentalists are interested in testing if the distribution has "moved to the right or left." At the present time, this notion of "movement" has not been defined. It could be that the mean, median, or mode has shifted. The mean can move to the right while the mode moves to the left. Each of these definitions has its own problems. Someone could also argue that "movement" is defined by stochastic dominance. However, when the variance of the distributions has changed, this definition becomes problematic. To identify whether or not the distribution has moved either to the right or left, the null hypothesis in (3.1) should be examined first. As discussed before, the WMW test suffers from a serious lack of power in the finite sample. This section aims to design better tests for which the power of the tests is much better than that of the WMW test. Throughout the paper, we assume that $\mathbf{x}$ and $\mathbf{y}$ are independent, random samples from unknown but bounded distributions.[2]

Pre-Test for the Equal Distributions   The beta distribution has a variety of useful properties which make it ideal for application to bargaining data. Beta random variables

---

[2]Assuming a bounded distribution is reasonable in a laboratory context because subjects are not able to play outside of a fixed range. For instance, in the canonical example of ultimatum games, the proposed offer must be between nothing and the whole pie (inclusive).

are bounded between zero and one. This is appropriate for most laboratory data which has a bounded structure by experimental construction. This method is called 'quasi'-parametric because the beta distribution is quite general. The density can be convex or concave. It can have a U-shape, L-shape or bell-shape, and the density can be unimodal or bimodal. The density function is given by,

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \qquad \text{for } x \in [0, 1], \alpha > 0, \beta > 0;$$

where $B(\alpha, \beta)$ is the beta function. There are many interesting properties of the beta distribution. First mean, variance and skewness can be expressed as functions of α and β. That is, $E(x) = \alpha / (\alpha + \beta)$, $V(x) = \alpha\beta \left[(\alpha + \beta)^2 (\alpha + \beta + 1) \right]^{-1}$, the median of x, $\text{Med}(x)$, becomes $(3\alpha - 1) / (3\alpha + 3\beta - 2)$, and the skewness of x, $\text{Skew}(x)$, becomes $2 (\beta - \alpha) \sqrt{1 + \alpha + \beta} \times \left[\sqrt{\alpha\beta} (2 + \alpha + \beta) \right]^{-1}$. Hence when $\alpha = \beta$, the distribution becomes symmetric as we showed before. When $\alpha = \beta = 1$, the distribution becomes uniform. Since all of the moments can be expressed as function of α and β, the null hypothesis of the equal distributions becomes equivalent to the null hypothesis of the equal αs and βs. That is,

$$\boxed{\begin{array}{l} \mathcal{H}_{1O} : F_x = F_y, \\ \mathcal{H}_{1A} : F_x \neq F_y \end{array}} \iff \boxed{\begin{array}{l} \mathcal{H}_{1O} : \alpha_x = \alpha_y \ \& \ \beta_x = \beta_y \\ \mathcal{H}_{1A} : \alpha_x \neq \alpha_y \text{ or } \beta_x \neq \beta_y \end{array}} \qquad (3.3)$$

To perform the distributional test, the beta parameters must be estimated first. The method of moments estimator is employed:

$$\hat{\alpha}_x = (\bar{x}) \frac{\bar{x}(1 - \bar{x}) - s_x^2}{s_x^2}, \quad \hat{\beta}_x = (1 - \bar{x}) \frac{\bar{x}(1 - \bar{x}) - s_x^2}{s_x^2}, \qquad (3.4)$$

where $\bar{x}$ and s_x^2 are the sample mean and variance of $\{x_i\}$, respectively. In practice, the maximum likelihood estimation has not commonly been used for two fundamental problems. First, the maximum likelihood estimators have no closed form. In general, some maximum likelihood approximation algorithms commonly fail for the beta distribution due to the presence of local maxima. This makes the maximum likelihood solution difficult (though not

impossible) to implement.[3] Second, bargaining data typically has many individuals who play absolute zero or one. Playing absolute zero or one should not be a serious problem because players may only be playing approximately beta. Nonetheless, any x value of zero or one makes the likelihood function uniformly zero for all choices of α and β, so maximum likelihood estimation becomes impossible.

Define the beta distribution test statistic, $\mathcal{B}$, as,

$$\mathcal{B} = \left[\hat{\alpha}_x - \hat{\alpha}_y, \hat{\beta}_x - \hat{\beta}_y\right] \left(\hat{\boldsymbol{\Omega}}_x/n_x + \hat{\boldsymbol{\Omega}}_y/n_y\right)^{-1} \left[\hat{\alpha}_x - \hat{\alpha}_y, \hat{\beta}_x - \hat{\beta}_y\right]',$$

where $\hat{\boldsymbol{\Omega}}_x$ is defined in Appendix C. Then by using the delta method, it is straightforward to show that as $n_x, n_y \to \infty$ the limiting distribution of $\mathcal{B}$ becomes the χ^2 distribution with two degrees of freedom. It is important to note that the testing of the null hypothesis, $\mathcal{H}_{1O}$, becomes a pre-testing for the next two tests which we will discuss shortly. Usually the 20% significance level is used for pre-testing. In other words, we recommend using 3.22 as the critical value for the beta distribution statistic $\mathcal{B}$, which is the 20% significance level for the χ^2_2 distribution.

Test for Shifting Distributions   When the null of the equal distribution is rejected, the interesting and important research question becomes how distributions are different from one another. Note that the null of equal distributions can be rejected when any moments of two distributions are different. However researchers are usually interested in comparing two empirical distributions, particularly the location of the distributions. When both distributions are symmetric, the mean, mode and median are all same so a mean comparison is adequate to find how the location of the two distributions are changing. Meanwhile, when the distributions is asymmetric, the mean, mode, and median are each possibly different from each other. Of course, when a distribution is not symmetric, the median becomes a better measure for central tendency but depending on the nature of the experimental game, sometimes

---

[3]For details, see Gupta and Nadarajah p. 230.

the mode of the distribution could be the parameter of interest. But usually, the researcher does not know which measure they want to use for comparison. In such circumstances, the following definition may be helpful.

Definition (Shifted)   *The distribution of Y has shifted to the right of the distribution of X, if and only if,*

$$\text{Skew}\left(X\right) > \text{Skew}\left(Y\right).$$

Skewness delivers information about the shape of this distribution that does not depend on center or spread. It follows that: it is rather useful to measure how a distribution moves left or right relative to the other distribution, especially when the distribution is not symmetric. For example consider the following case where $\alpha_x = 2$, $\beta_x = 4$, $\alpha_y = 3$, and $\beta_y = 6$. Then both samples share the same mean of 0.33 but don't share either same variance or skewness. The skewness for $\{y_i\}$ becomes 0.41 but that for $\{x_i\}$ is 0.47. Hence the series $\{y_i\}$ is less right skewed compared to the series $\{x_i\}$. According to our new definition, we can say that the empirical distribution of $\{y_i\}$ shifts to the right of the empirical distribution of $\{x_i\}$. Figure 3.1 shows this case explicitly.Notice that the majority of the mass of the distribution seems to move to the right, but the tail moves left. This is consistent with the new definition (and the reason for skewness occurring in the opposite direction from shifting). Therefore, when a test of means cannot reject the null hypothesis, testing for shifting could still produce statistically significant results. Another example is where $\alpha_x = 1.5$, $\beta_x = 3.833$, $\alpha_y = 2$, and $\beta_y = 5.333$. Here, both samples have the same median of 0.25 but the skewness, mean, and variance are all different. The skewness for $\{y_i\}$ becomes 0.67, but that for $\{x_i\}$ is 0.63. The distribution of $\{y_i\}$ shifts to the right of that of $\{x_i\}$ here as well. Figure 3.2 shows this case. Likewise, both sample can have the same mode when shifting occurs. Consider when $\alpha_x = 2$, $\beta_x = 3$, $\alpha_y = 3$, and $\beta_y = 4.333$. Here, the modes for both distributions are the
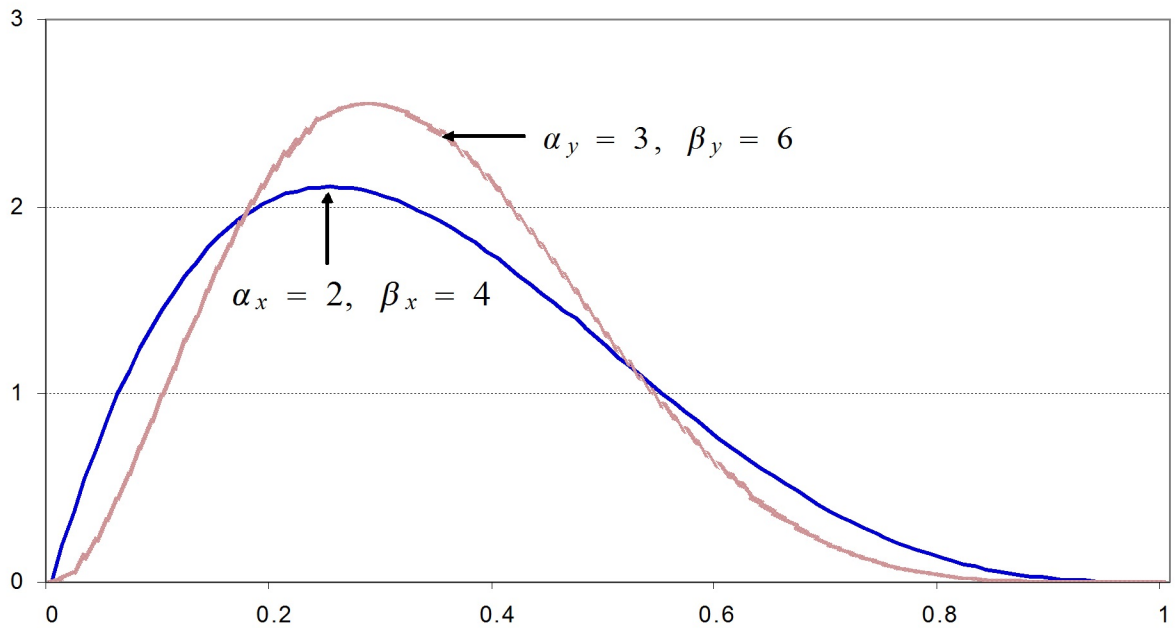
Figure 3.1. Shifting Densities with Equal Means
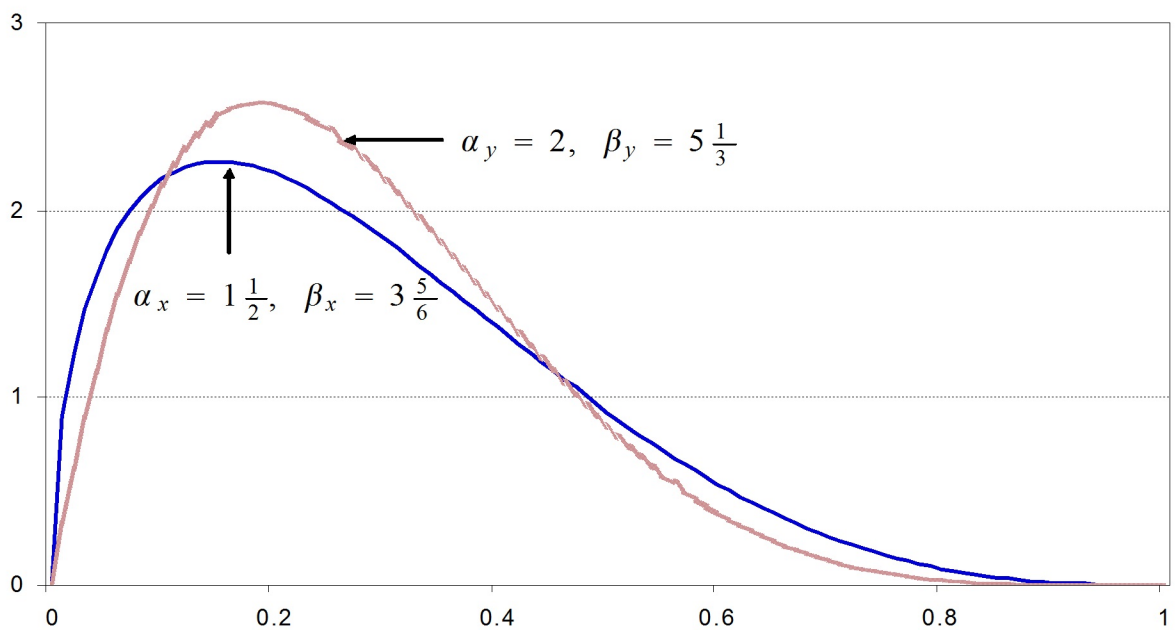
Figure 3.2. Shifting Densities with Equal Medians

same, occurring at 0.333, but $\mathrm{Skew}(X) = 0.286$ while $\mathrm{Skew}(Y) = 0.306$. Figure 3.3 shows this case explicitly. Notice that in this case, the tail is far more important than than the highest mass point. Around the highest mass point, both distributions are very similar, so the tail becomes the only important difference. $\{y_i\}$ has a smaller right tail, so $\{y_i\}$ has shifted left of the distribution of $\{x_i\}$ (opposite direction of other cases).
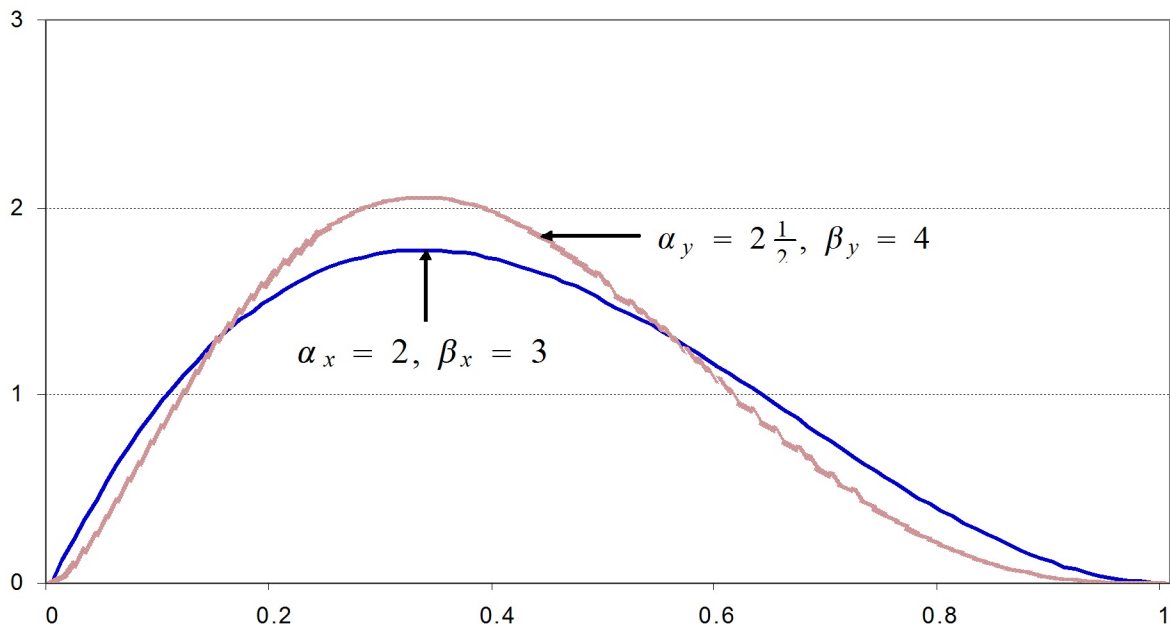
Figure 3.3. Shifting Densities with Equal Modes

Similarly, two distributions can have the same variances when shifting occurs. Therefore, shifting provides new information about how the density changes. This new information is fundamentally important because of its economic intuition: Suppose there are two groups, a treatment and a control for ultimatum game offers. As stated above, in ultimatum games, distributions of offers are typically right skewed. If the empirical distribution from the treatment group has shifted to the left of that from the control group, then most individuals in the treatment group are making lower offers, while a small number of treated individuals are making much larger offers. Thus, shifting compares how treated individuals play relative to their own group using the control group as a 'base-line' for this comparison.

Note that before estimating the skewness, it is useful to test whether or not either distribution is symmetric. Since both samples are bounded between 0 and 1, the non-rejection of the null hypothesis of a symmetric distribution implies that the means of both distributions are the same and become 0.5. Thus, the following null hypothesis of interest becomes,

$$\boxed{\mathcal{H}_{2O} : f_x\left(c_x + x\right) = f_x\left(c_x - x\right) \text{ for some } c_x \in \mathbb{R} \text{ and all } x \in \mathbb{R},}$$

$$\Updownarrow \tag{3.5}$$

$$\boxed{\mathcal{H}_{2O} : \alpha_x = \beta_x} \iff \boxed{\mathcal{H}_{2O} : \mu_x = 0.5}$$

The test statistic can then be defined as,

$$z_{\text{Sym}(X)} = \sqrt{n_x}\frac{\hat{\alpha}_x - \hat{\beta}_x}{\left(\mathbf{J}'\hat{\boldsymbol{\Omega}}_x\mathbf{J}\right)^{1/2}},$$

where $\mathbf{J} = [1, -1]'$. Straightforward application of the delta method to Lemma 3 shows that this test is consistent as $n_x \to \infty$. The result for $z_{\text{Sym}(Y)}$ is obviously the same. For testing the alternate formulation of $\mu_x = 0.5$, the test statistic becomes,

$$z^o_{\text{Sym}(X)} = \sqrt{n_x}\frac{\hat{\mu}_x - 0.5}{\left(\mathbf{J}^{*\prime}\hat{\boldsymbol{\Omega}}_x\mathbf{J}^*\right)^{1/2}}$$

where $\hat{\mu}_x = \hat{\alpha}_x/\left(\hat{\alpha}_x + \hat{\beta}_x\right)$ and $\mathbf{J}^{*\prime} = \left(\hat{\alpha}_x + \hat{\beta}_x\right)^{-2}\left[\hat{\beta}_x, -\hat{\alpha}_x\right]$. Testing using either method has three possible outcomes: First, tests for both samples could reject the null. In this case, both distributions are potentially symmetric and shifting has not conclusively occurred. Here, testing for shifting separately is redundant and should not be performed. Second, a test for one sample could reject and the other could accept. In this case, the distribution which rejects has shifted away from symmetry while the other has not, so shifting has occurred. Here also, there is no need to test for shifting. Third and finally, both tests can reject. In this case, shifting must be tested directly.

To test if shifting has occurred, the following hypotheses are defined.

$$\mathcal{H}_{3O} : \text{Skew}\,(X) = \text{Skew}\,(Y), \quad \mathcal{H}_{3A} : \text{Skew}\,(X) \neq \text{Skew}\,(Y). \tag{3.6}$$

If the null is rejected, the distributions have shifted. If the null cannot be rejected, then there is no conclusive evidence of shifting. The following statistic is used to test $\mathcal{H}_{3O}$,

$$z_{\text{Skew}} = \frac{\hat{\kappa}_{3x}^{(\text{Beta})} - \hat{\kappa}_{3y}^{(\text{Beta})}}{\left(\mathbf{J}_x'\mathbf{\Omega}_x\mathbf{J}_x/n_x + \mathbf{J}_y'\mathbf{\Omega}_y\mathbf{J}_y/n_y\right)^{1/2}},$$

where

$$\mathbf{J}_x = \frac{\left(\hat{\alpha}_x + \hat{\beta}_x\right)\sqrt{\hat{\alpha}_x\hat{\beta}_x}}{\left(\hat{\alpha}_x + \hat{\beta}_x + 2\right)^2\sqrt{\hat{\alpha}_x + \hat{\beta}_x + 1}}\left[-\frac{\left(\hat{\beta}_x + 1\right)\left(3\hat{\alpha}_x + \hat{\beta}_x + 2\right)}{\hat{\alpha}_x}, \frac{\left(\hat{\alpha}_x + 1\right)\left(\hat{\alpha}_x + 3\hat{\beta}_x + 2\right)}{\hat{\beta}_x}\right]'.$$

Straightforward application of the delta method to Lemma 3 shows that this test is consistent as $n_x, n_y \to \infty$. When the distribution pre-test is combined with the symmetry and skewness tests, the method is referred to as the 'sequential testing procedure.'

Theorem 1 proves that this method is consistent as $n_x, n_y \to \infty$.

Theorem 1 (Consistency of the Beta Sequential Procedure for Independent Samples) *Given two independent, random samples,* $\mathbf{x}$ *and* $\mathbf{y}$, *the following cases apply:*

(i) Under the null that $\alpha_x = \alpha_y$ *and* $\beta_x = \beta_y$,

$$\lim_{n_x n_y \to \infty} \Pr\left[\mathcal{B} < F_{\chi_2^2}^{-1}(1 - \alpha)\right] = 1 - \alpha,$$

where $F_{\chi_2^2}^{-1}(p)$ *is the quantile function of the* χ_2^2 *distribution.*

(ii) When $F_X \neq F_Y$, *under the null that* $\alpha_x = \beta_x$,

$$\lim_{n_x n_y \to \infty} \Pr\left[\left|z_{\text{Sym}(X)}\right| < F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)\right] = 1 - \alpha,$$

where $F_Z^{-1}(p)$ *is the quantile function of the* $\mathcal{N}(0,1)$ *distribution.*

(iii) When $F_X \neq F_Y$ and X and Y are not symmetric, under the null that $\kappa^*_{3x} = \kappa^*_{3y}$,

$$\lim_{n_x n_y \to \infty} \Pr\left[|z_{\text{Skew}}| < F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)\right] = 1 - \alpha.$$

The alternatives are obviously consistent as well. See Appendix C for detailed proof.

The following remarks concern sequential testing in greater generality. Sometimes, researchers are interested in evaluating the distributions of paired (or dependent) samples. For these comparisons, experimentalists have commonly used the Wilcoxon signed-rank test. The following remark concerns paired samples and the beta testing procedure.

Remark 1 (Beta Sequential Procedure for Dependent Samples) In order to analyze paired data, $(x_1, y_1), \ldots, (x_n, y_n)$, using the beta approach, a bivariate beta distribution must be assumed. This distribution is still an active area of research. Some of these distributions will lend themselves to this estimation better than others. For instance, El-Bassiouny and Jones (2005) provide a distribution in which many other bivariate beta distributions are nested, with joint density:

$$f(x,y) = C\frac{x^{a/2-1}(1-x)^{(b+d)/2-1}y^{b/2-1}(1-y)^{(a+d)/2-1}}{(1-xy)^{(a+b)/2}}F\left(\frac{a+b}{2}, \frac{d-c}{2}; \frac{a+d}{2}; \frac{x(1-y)}{1-xy}\right),$$

where C is a constant defined so that the double integral equals unity and $0 < x, y < 1$. The moments of this distribution involve the generalized hypergeometric function, so the method of moments estimator will not have a closed-form. Similarly, the five parameter bivariate beta distribution in Gupta and Wong (1985) has moments which involve the generalized hypergeometric function. Other choices include the three parameter bivariate beta distribution from Gupta and Wong (1985) with joint density:

$$f(x,y) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}x^{a-1}y^{b-1}(1-x-y)^{c-1},$$

where $x + y \leq 1$ and $x, y > 0$ and where $\Gamma(x)$ is the gamma function. Because this density has only three parameters and because the inequality is restrictive, this density is not very general. Nadarajah and Kotz (2005) define the density

$$f(x,y) = \frac{x^{c-1}(y-x)^{b-1}y^{a_1-c-b}(1-y)^{b_1-1}}{B(a_1, b_1)B(c, b)}$$

where $0 \leq x \leq y \leq 1$. Because of the inequality, this distribution is also quite restrictive. If a general bivariate beta distribution with closed-form method of moments estimators cannot be developed, then approximation methods can be used with some existing bivariate beta distributions.

Remark 1 states that this paper does not provide a beta test for paired samples. Development of this test will be straightforward (although potentially tedious) once a bivariate beta distribution is generally accepted in the statistical literature.

When the beta distribution cannot be assumed (even approximately), nonparametric statistics may be more appropriate. In order to perform the sequential testing procedure nonparametrically, the null hypotheses must be somewhat relaxed. The following statistics assume independent samples. Consider the following alternative to the distribution statistic outlined above,

$$\chi^2_{\text{Nonp}} = \left[\bar{x} - \bar{y}, s^2_x - s^2_y\right]\left(\hat{\Xi}_x/n_x + \hat{\Xi}_y/n_y\right)^{-1}\left[\bar{x} - \bar{y}, s^2_x - s^2_y\right]',$$

where,

$$\Xi_x = \begin{bmatrix} \bar{\mu}_{2x} & \bar{\mu}_{3x} \\ \bar{\mu}_{3x} & \bar{\mu}_{4x} - \bar{\mu}^2_{2x} \end{bmatrix},$$

and $\bar{\mu}_{jx}$ refers to the jth central moment of X. Under the null that X and Y have the same mean and variance, $\chi^2_{\text{Nonp}} \xrightarrow{d} \chi^2_2$. For the second test, symmetry is a stronger condition than is needed in practice. As shown in MacGillivray (1981), for many distributions (Pearson family), zero skewness implies that the mean, median, and mode (if it exists) all occur at

the same point. If these measures of central-tendency are all the same, then testing any one of them will be equivalent to testing them all. Furthermore, if X has positive skewness, then $\text{Mode}(X) < \text{Median}(X) < \text{E}(X)$ and *vice versa* for negative skewness. These facts establish that testing if $\text{Skew}(X) = 0$ is sufficient for the purposes of this test. The skewness coefficient and its estimator are defined by,

$$\kappa_{3x} = \frac{\bar{\mu}_{3x}}{\sigma_x^3}, \qquad \hat{\kappa}_{3x} = \frac{n\sqrt{n-1}}{n-2} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{3/2}}.$$

The test statistic for zero skewness is defined by,

$$z_{\text{Sym}(X)}^* = \sqrt{n_x} \frac{\hat{\kappa}_{3x}}{\hat{\omega}_x^{1/2}}.$$

where ω_x is defined in Appendix C. Under the null that X has zero skewness, $z_{\text{Sym}(X)} \xrightarrow{d} \mathcal{N}(0,1)$. Because the samples are assumed to be independent, the skewness test statistic is simply defined by,

$$z_{\text{Skew}}^* = \frac{\hat{\kappa}_{3x} - \hat{\kappa}_{3y}}{(\hat{\omega}_x/n_x + \hat{\omega}_y/n_y)^{1/2}}.$$

Under the null that no shifting has occurred, $z_{\text{Skew}}^* \xrightarrow{d} \mathcal{N}(0,1)$.

Remark 2 (Nonparametric Sequential Procedure) Given two independent, random samples, $\mathbf{x}$ and $\mathbf{y}$, the following cases apply:

(i) Under the null that $\text{E}[X] = \text{E}[Y]$ and $\text{Var}[X] = \text{Var}[Y]$,

$$\lim_{n_x n_y \to \infty} \text{Pr}\left[\chi_{\text{Nonp}}^2 > F_{\chi_2^2}^{-1}(1-\alpha)\right] < \alpha,$$

where $F_{\chi_2^2}^{-1}(p)$ is the quantile function of the χ_2^2 distribution.

(ii) When $\text{E}[X] \neq \text{E}[Y]$ or $\text{Var}[X] \neq \text{Var}[Y]$, under the null that X has zero skewness,

$$\lim_{n_x n_y \to \infty} \text{Pr}\left[\left|z_{\text{Sym}(X)}^*\right| > F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)\right] < \alpha,$$

where $F_Z^{-1}(p)$ is the quantile function of the $\mathcal{N}(0,1)$ distribution.

(iii) When $\mathrm{E}[X] \neq \mathrm{E}[Y]$ or $\mathrm{Var}[X] \neq \mathrm{Var}[Y]$, and when X and Y have non-zero skewness, under the null that $\mathrm{Skew}(X) = \mathrm{Skew}(Y)$,

$$\lim_{n_x n_y \to \infty} \mathrm{Pr}\left[|z^*_{\mathrm{Skew}}| > F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)\right] < \alpha.$$

Remark 2 proves that the nonparametric sequential testing method is consistent as $n_x, n_y \to \infty$. Remark 2 follows as an indirect consequence of Lemmas 4 and 5, which must are easily adjusted to account for the independence assumptions. The alternatives are obviously consistent as well.

For the nonparametric sequential procedure, paired samples are allowed by using,

$$\chi^2_{\mathrm{Pair}} = n\left[\bar{x} - \bar{y}, s_x^2 - s_y^2\right] \hat{\Xi}^{-1} \left[\bar{x} - \bar{y}, s_x^2 - s_y^2\right]',$$

where Ξ is stated in Appendix C. Testing zero skewness is the same as the nonparametric independent sample case. The test statistic for equal skewness must be adjusted by an ω_{xy} term (defined in Appendix C) so that z^{**}_{Skew} becomes,

$$z^{**}_{\mathrm{Skew}} = \sqrt{n}\frac{\hat{\kappa}_{3x} - \hat{\kappa}_{3y}}{(\hat{\omega}_x + \hat{\omega}_y - 2\hat{\omega}_{xy})^{1/2}}$$

Remark 3 (Nonparametric Paired-Sample Sequential Procedure)   Given a paired (potentially-dependent) sample, $(x_1, y_1), \ldots, (x_n, y_n)$, the following cases apply:

(i) Under the null that $\mathrm{E}[X] = \mathrm{E}[Y]$ and $\mathrm{Var}[X] = \mathrm{Var}[Y]$,

$$\lim_{n_x n_y \to \infty} \mathrm{Pr}\left[\chi^2_{\mathrm{Pair}} > F^{-1}_{\chi^2_2}(1 - \alpha)\right] < \alpha,$$

where $F^{-1}_{\chi^2_2}(p)$ is the quantile function of the χ^2_2 distribution.

(ii) When $\mathrm{E}[X] \neq \mathrm{E}[Y]$ or $\mathrm{Var}[X] \neq \mathrm{Var}[Y]$, under the null that X has zero skewness,

$$\lim_{n_x n_y \to \infty} \mathrm{Pr}\left[|z^*_{\mathrm{Sym}(X)}| > F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)\right] < \alpha,$$

where $F_Z^{-1}(p)$ is the quantile function of the $\mathcal{N}(0, 1)$ distribution.

(iii) When $\mathrm{E}[X] \neq \mathrm{E}[Y]$ or $\mathrm{Var}[X] \neq \mathrm{Var}[Y]$, and when X and Y have non-zero skewness, under the null that $\mathrm{Skew}(X) = \mathrm{Skew}(Y)$,

$$\lim_{n_x n_y \to \infty} \mathrm{Pr}\left[|z_{\mathrm{Skew}}^{**}| > F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)\right] < \alpha.$$

Remark 3 states that the nonparametric paired-sample sequential procedure is consistent as $n \to \infty$. The paired-sample case is a direct consequence of Lemmas 4 and 5 in Appendix C.

3.4   Monte Carlo Simulation Results

This section details the finite sample performance of the WMW test and sequential testing procedure. First, the independent sample tests are found to all have the correct size (prob. of rejecting a true null). Second, the power of the independent sample tests are investigated. The newly developed tests are found to have excellent power in the finite sample. Lastly, the two symmetry tests are compared. The tests are performed at the 5% nominal level, so the tests would ideally reject a true null once per twenty replications, on average. In each case, 10,000 replications are used.[4]

For all the following simulations, the data generating process (DGP) is given by,

$$x_i \sim Beta\left(\alpha_x, \beta_x\right), \; y_i \sim Beta(\alpha_y, \beta_y).$$

Under the null of the same distribution in (3.3), $\alpha_x = \alpha_y$ and $\beta_x = \beta_y$. Under the null of the same skewness in (3.6), $\mathrm{Skew}(X) = \mathrm{Skew}(Y)$. In both processes, the samples are independent, identically distributed, and they are independent of one another. In practice, the sample size for each of the two samples could be different. For simplicity, the sample size for both samples will be fixed to one number, n. Various cases are considered but here only the following four cases are reported to save space.

---

[4]The tests described in Remarks 1-3 are similarly analyzed in the online appendix provided on the author's website. Stata code for the testing procedure is also provided with the same input format as the WMW test for convenience.

Table 3.5. Data Generating Processes for Monte Carlo Simulations

	DGP1		DGP2		DGP3		DGP4	
	x	**y**	**x**	**y**	**x**	**y**	**x**	**y**
α	35.81	35.81	9	3	6	3	3	6
β	47.50	47.50	3	1.35	2	1	3	5
Mean	0.43	0.43	0.75	0.69	3/4	3/4	1/2	6/11
Variance	0.05^2	0.05^2	0.12^2	0.20^2	0.14^2	0.19^2	0.19^2	0.14^2
Skewness	0.06	0.06	-0.60	-0.60	-0.69	-0.86	0	-0.10

DGP 1 is used to demonstrate the size of the distribution test; the distribution is constructed to resemble the US TS average sample from RPOZ. In DGP 2, the two distributions are different, yet both have equal skewness. Hence when the distribution test is used on DGP 2, it demonstrates the power of the test, but when skewness is tested, DGP 2 is used to show the size. DGP 3 shows the power of the test when the samples have equal means although the variances and skewness are different. DGP 4 is used to demonstrate the power of the test when one distribution is symmetric.

Table 3.6. Rejection Rates of the WMW and Beta Distribution Tests (Nominal: 5%)

Test	DGP 1 (Size)				DGP 2 (Power)			
	$n = 25$	50	100	250	$n = 25$	50	100	250
WMW	0.05	0.05	0.05	0.05	0.14	0.23	0.41	0.77
Beta Distribution Test	0.02	0.04	0.04	0.04	0.35	0.87	1.00	1.00

Table 3.6 reports the probability of rejecting the null of the equal distributions. With DGP 1, where the distributions of the two samples are the same, the WMW test provides very accurate the size of the test, while the beta distribution test is mildly conservative when n is small. In terms of power of the tests demonstrated using DGP 2, the beta distribution test significantly dominates WMW, rejecting perfectly when n is greater than 100.

Table 3.7 reports the probability of rejecting the null for the two symmetry tests described in (3.5) ($\alpha = \beta$ and $\mu = 0.5$) under DGP 4. DGP 4 is chosen so that the symmetry test does not perform perfectly, so comparisons can be made. Observing the **x** sample, where

Table 3.7. Rejection Rates for Two Symmetry Tests in DGP 4 (Nominal Size: 5%)

	$\mathcal{H}_{2O} : \alpha_x = \beta_x$				$\mathcal{H}_{2O} : \mu_x = 0.5$			
n	25	50	100	250	25	50	100	250
Size	0.03	0.04	0.04	0.04	0.07	0.05	0.05	0.05
Power	0.31	0.58	0.87	1.00	0.45	0.65	0.89	1.00

the null is true, the $\alpha_x = \beta_x$ test is rather conservative, rejecting in less than 5% of cases. The $\mu_x = 0.5$ has slight oversize distortion when n is quite small, yet this distortion vanishes quickly as n increases. In terms of power, the $\mu = 0.5$ test uniformly dominates the $\alpha = \beta$ test. Therefore, since the $\mu = 0.5$ test shows little size distortion and beats the $\alpha = \beta$ test in terms of power, the $\mu = 0.5$ test is recommended. This test was used in the following simulation and example.

Table 3.8. Rejection Rate for Sequential Skewness Test (Nominal Size: 5% for the Skewness Test, 20% for the distribution test)

Test	DGP 1 & 3				DGP 2 & 4			
	$n = 25$	50	100	250	$n = 25$	50	100	250
Size (DGP 1&2)	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.05
Power (DGP 3&4)	0.12	0.21	0.37	0.74	0.30	0.56	0.86	0.97

Table 3.8 shows the size and power of the sequential testing procedure for testing the null of equal skewness. Here, the distribution pre-test is performed at the 20% level, while the other tests are performed at the 5% level. In terms of size, the sequential testing procedure is quite accurate even when n is small. Also, the skewness test is quite powerful. Although results for WMW are not shown alongside the skewness tests (since they test different null hypotheses), the sequential testing procedure dominates such methods uniformly.

3.5 Return to the Empirical Example

The importance of the sequential testing procedure is illustrated through application to the canonical empirical example. RPOZ studies whether there are differences in ultimatum game offers between Israel, Japan, Yugoslavia and the US. Using WMW and the z-score test on TS average data, offers from Israel are found to be different than the others, but differences could not be detected between Japan and Yugoslavia and Japan and the US, even though the variances are radically different. Table 3.9 reports p-values from the beta sequential testing procedure. Interestingly, all distributions are found to be different using the beta sequential method, except for the US and Yugoslavia (which is consistent with results from other, round samples using this test as well as others) either with TS average or with the last round sample. Furthermore, the sequential procedure provides information about shifting behavior. Note that all beta-skewness test statistics are significantly greater than zero except for the pair of Yugoslavia and the US. This implies that the offers from the US has shifted right from those from Japan and Israel, and the offers from Japan has shifted right from that from Israel also. Results regarding Yugoslavia are similar to those from the US.

Table 3.9. Sequential Testing p-Values for Between Country Comparisons

Samples (**x** v **y**)	TS Average			Last Round		
	WMW	Beta Sequential		WMW	Beta Sequential	
		Distn̄	Skew		Distn̄	Skew
Israel v Japan	0.005	0.043	0.055	0.001	0.011	0.000
Israel v Yugo	0.000	0.002	0.000	0.000	0.001	0.000
Israel v US	0.000	0.002	0.000	0.000	0.001	0.000
Japan v Yugo	0.213	0.013	0.013	0.022	0.074	0.008
Japan v US	0.106	0.023	0.011	0.009	0.118	0.027
Yugo v US	0.310	0.916	n.a.	0.254	0.527	n.a

Figure 3.4 shows the estimated densities for US and Yugoslavia offers. As shown in Table 3.9, the estimated two distributions are almost identical each other. The estimated

parameters for the US are $\hat{\alpha}_{\text{US}} = 35.8058$ and $\hat{\beta}_{\text{US}} = 47.4994$; for Yugoslavia, $\hat{\alpha}_{\text{Yugo}} = 37.9801$ and $\hat{\beta}_{\text{Yugo}} = 51.4831$. Notice that, as expected, the estimated densities for the US and Yugoslavia offers are almost exactly the same. This is consistent with the estimated means and variances above. Indeed, from the beta sequential test, offers from the US and Yugoslavia are drawn from the same distribution. The mode of Yugoslavian offers is slightly lower than that of the US offers graphically, but the difference is not statistically significant.
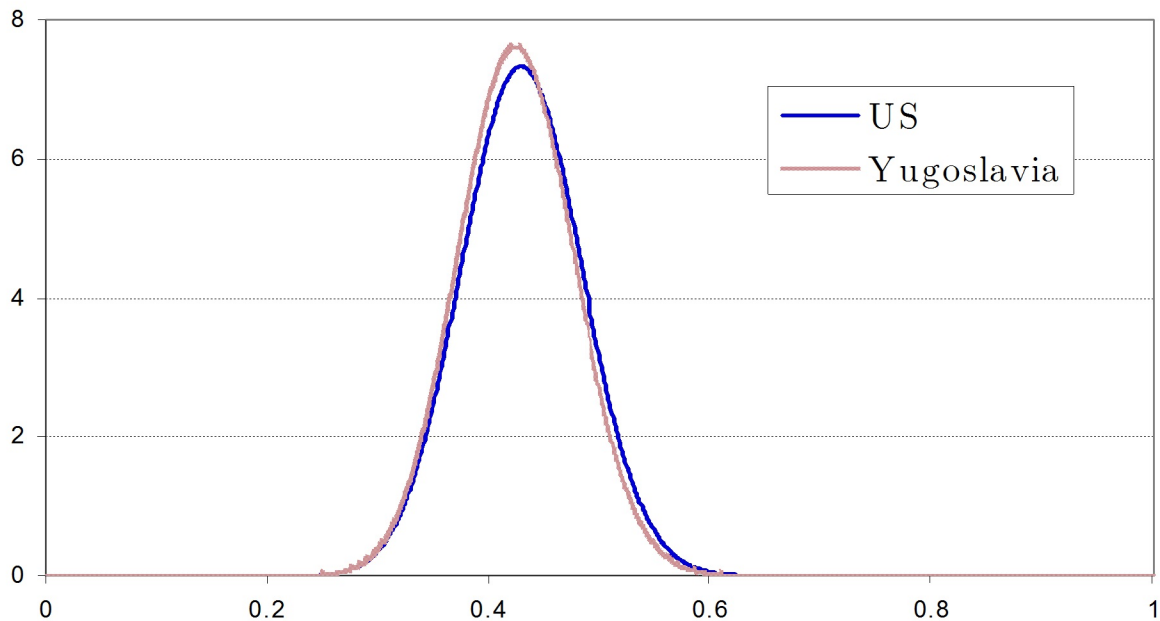
Figure 3.4. Estimated Densities For the US and Yugoslavia Data (TS Average)

Figure 3.5 shows the estimated densities from the US, Israel, and Japan. The shifting results are clearly visible. The Japan density has shifted to the right of that from Israel, and the US density right of those from Japan and Israel. The estimated parameters for the Israel sample are $\hat{\alpha}_{\text{Isr}} = 10.3546$ and $\hat{\beta}_{\text{Isr}} = 19.9160$; those from Japan are $\hat{\alpha}_{\text{Japan}} = 8.5263$ and $\hat{\beta}_{\text{Japan}} = 12.2469$. The low variance of the US is clearly visible.
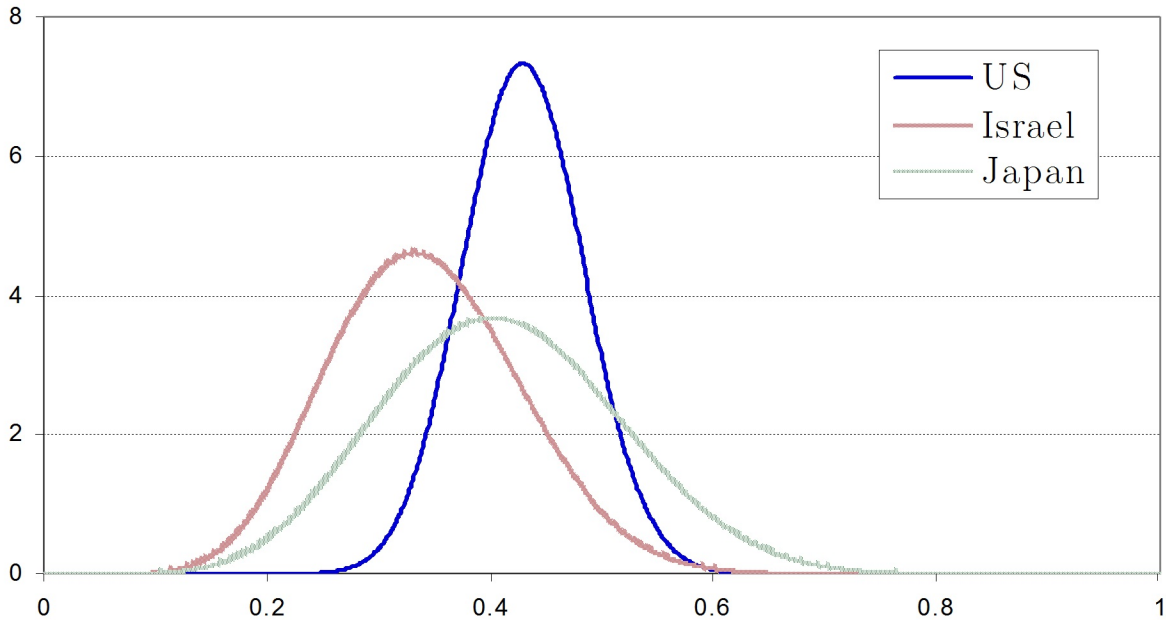
Figure 3.5. Estimated Densities Between Countries (TS Average)

3.6 Conclusion

This paper pointed out some statistical issues of the WMW test, which has been popularly used for the comparison of a pair of samples. First, even though WMW was designed for the comparison of two random distributions, it has been misinterpreted as a test of means or central-tendency, partly because WMW struggles to distinguish differences in variance. Second, WMW has low power. Third, because WMW has low power, some experimentalists have pooled their repeated data together, which causes inconsistency. Fourth, and most importantly, even when WMW rejects the null of the same distribution, WMW does not explain how the distributions are different.

This paper explains these criticisms and provides new tests. The first test is a general way to compare distributions. The paper also provides a test of skewness which may be more appropriate for deciding if a distribution has "shifted to the right or left." The skewness test should be performed as a part of the 'sequential testing method' for correct interpretation.

These tests are performed 'quasi'-parametrically using the very general beta distribution. The tests are proven to work in the large sample theoretically. The finite sample performance of the tests are analyzed by means of Monte Carlo simulations and through application to three relevant empirical examples. The new testing procedure dominates WMW in terms of power.

APPENDIX A

PROOFS FOR CHAPTER 1

Proof of Lemma 1 (Asymptotic Factor Number for Weak Factors) The only requirement is to show that the difference in residual sum of squares between using no factors and using k factors is small as $N, T \to \infty$. i.e.,

$$V(0) - V(k) = O_p\left(C_{NT}^{-2}\right), \tag{A.1}$$

for the following reason. The criterion function, IC_p, is defined by: $IC_p(k) = \ln[V(k)] + kp_{N,T}$, where $p_{N,T} \to 0$ and $C_{NT}^2 p_{N,T} \to \infty$ as $N, T \to \infty$. Hence,

$$IC_p(0) - IC_p(k) = \ln\left[\frac{V(0)}{V(k)}\right] - kp_{N,T}.$$

From Bai and Ng (2002), $V(0) - V(k) = O_p\left(C_{NT}^{-2}\right)$ implies $\ln[V(0)/V(k)] = O_p\left(C_{NT}^{-2}\right)$ whereas the penalty goes to infinity when multiplied by C_{NT}^2. Therefore, as $N, T \to \infty$, the penalty dominates $\ln[V(0)/V(k)]$ no matter which $k > 0$ is chosen. Hence, it is only necessary to show (A.1).

The eigenvalues of a rank k matrix A are denoted as $\varrho_1(A), \ldots, \varrho_k(A)$, ordered from largest to smallest. Proceed by expressing the difference in eigenvalue form,

$$
\begin{aligned}
V(0) - V(k) &= \sum_{l=1}^{N} \rho_l\left(\frac{x'x}{NT}\right) - \sum_{l=k+1}^{N} \rho_l\left(\frac{x'x}{NT}\right) \\
&= \sum_{l=1}^{k} \rho_l\left(\frac{x'x}{NT}\right) \le k\rho_1\left(\frac{x'x}{NT}\right).
\end{aligned} \tag{A.2}
$$

Now use the model of x_{it} to show,

$$
\begin{aligned}
\rho_1\left(\frac{x'x}{NT}\right) &= \rho_1\left(\frac{(\Psi Z' + x^{o\prime})(Z\Psi' + x^o)}{NT}\right) \\
&= \rho_1\left(\frac{\Psi Z'Z\Psi'}{NT} + \frac{\Psi Z'x^o + x^{o\prime}Z\Psi'}{NT} + \frac{x^{o\prime}x^o}{NT}\right).
\end{aligned}
$$

From the Hermitian matrix eigenvalue inequality,

$$\rho_1\left(\frac{x'x}{NT}\right) \le \rho_1\left(\frac{\Psi Z'Z\Psi'}{NT}\right) + \rho_1\left(\frac{\Psi Z'x^o + x^{o\prime}Z\Psi'}{NT}\right) + \rho_1\left(\frac{x^{o\prime}x^o}{NT}\right) \qquad \text{(A.3)}$$
$$= I + II + III.$$

Note that $III = \rho_1\left(x^{o\prime}x^o/NT\right)$ follows $O_p\left(C_{NT}^{-2}\right)$ from the regularity conditions described above.

To bound the II term, re-express the quantity using the L_2-norm[1]:

$$\rho_1\left(\frac{\Psi Z'x^o + x^{o\prime}Z\Psi'}{NT}\right) = \left\|\frac{\Psi Z'x^o + x^{o\prime}Z\Psi'}{NT}\right\|_2.$$

Next the triangle inequality can be applied:

$$\rho_1\left(\frac{\Psi Z'x^o + x^{o\prime}Z\Psi'}{NT}\right) \le \left\|\frac{\Psi Z'x^o}{NT}\right\|_2 + \left\|\frac{x^{o\prime}Z\Psi'}{NT}\right\|_2.$$

First consider:

$$\left\|\frac{\Psi Z'x^o}{NT}\right\|_2 = \sqrt{\rho_1\left((NT)^{-2}x^{o\prime}Z\Psi'\Psi Z'x^o\right)}.$$

Then,

$$\rho_1\left((NT)^{-2}x^{o\prime}Z\Psi'\Psi Z'x^o\right) \le \operatorname{tr}\left[(NT)^{-2}x^{o\prime}Z\Psi'\Psi Z'x^o\right] = \operatorname{tr}\left[N^{-1}T^{-2}x^{o\prime}Z\left(\frac{1}{N}\sum_{i=1}^{T}\psi_i\psi_i'\right)Z'x^o\right]$$

Without loss of generality, assume that the loadings follow $\psi_i = O_p\left(C_{NT}^{-1}\right)$.[2] Hence,

$$\rho_1\left((NT)^{-2}x^{o\prime}Z\Psi'\Psi Z'x^o\right) = O_p\left(C_{NT}^{-2}\right)\operatorname{tr}\left[N^{-1}T^{-2}x^{o\prime}ZZ'x^o\right]$$
$$= O_p\left(\frac{1}{TC_{NT}^2}\right)\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t x_{it}^o\right\|_2^2,$$

---

[1]In Bai and Ng (2002), the commonly used norm is the Frobenius norm, $\|A\|_F = \operatorname{tr}[A'A]^{1/2}$. In our proof, we use the L_2-norm, $\|A\|_2 = \rho_1\left(A'A\right)^{1/2}$. For any matrix A, $\|A\|_2 \le \|A\|_F$ with equality if and only if $\operatorname{rank}(A) = 1$. Hence, for any vector, $\|A\|_2 = \|A\|_F$. We occasionally state a Bai and Ng (2002) assumption in terms of the L_2-norm, which is valid because of this equality for vectors.

[2]We know that $\psi_i'Z_t = O_p\left(C_{NT}^{-1}\right)$ so we can always rewrite $\psi_i'Z_t = \psi_i'I_rZ_t = \psi_i'A'A^{-1\prime}Z_t$ so that $\|A\psi_i\|_2 = O_p\left(C_{NT}^{-1}\right)$ and $\left\|A^{-1\prime}Z_t\right\|_2 = O_p(1)$.

so that,

$$\left\| \frac{\Psi Z' x^o}{NT} \right\|_2 = O_p \left(\frac{1}{T^{1/2} C_{NT}} \right) \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} Z_t x_{it}^o \right\|_2^2}.$$

This is straightforward to bound because Bai and Ng (2002) makes the following exogeneity assumption regarding the factors and errors:

$$E \left[\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} Z_t x_{it}^o \right\|_2^2 \right] \le M.$$

Therefore, $\left\| (NT)^{-1} \Psi Z' x^o \right\|_2 = O_p \left(T^{-1/2} \times C_{NT}^{-1} \right) = O_p \left(C_{NT}^{-2} \right).$ $\left\| (NT)^{-1} x^{o\prime} Z \Psi' \right\|_2$ can be similarly bounded. Hence, $II = O_p \left(C_{NT}^{-2} \right).$

Now, it is enough to show that $I = O_p \left(C_{NT}^{-2} \right)$. If $\psi_i' Z_t = O_p \left(C_{NT}^{-1} \right)$, we can assume without loss of generality that $Z_t = O_p \left(C_{NT}^{-1} \right)$ while $\psi_i = O_p (1)$. From here, it is straightforward to bound I:

$$\rho_1 \left(\frac{\Psi Z' Z \Psi'}{NT} \right) = \rho_1 \left(\frac{1}{N} \Psi \left(\frac{1}{T} \sum_{t=1}^{T} Z_t Z_t' \right) \Psi' \right)$$
$$= \rho_1 \left(O_p \left(C_{NT}^{-2} \right) \frac{\Psi \Psi'}{N} \right)$$
$$= O_p \left(C_{NT}^{-2} \right) \rho_1 \left(\frac{\Psi \Psi'}{N} \right),$$

and the loadings can be bounded by,

$$\rho_1 \left(\frac{\Psi \Psi'}{N} \right) \le \mathrm{tr} \left[\frac{\Psi \Psi'}{N} \right] = \frac{1}{N} \sum_{i=1}^{N} \psi_i' \psi_i = \frac{1}{N} \sum_{i=1}^{N} \| \psi_i \|_2^2.$$

Since the loadings are absolutely bounded $(\max_i \| \psi_i \|_2 = \bar{\psi})$,

$$\rho_1 \left(\frac{\Psi \Psi'}{N} \right) \le \frac{1}{N} \sum_{i=1}^{N} \bar{\psi}^2 = \bar{\psi}^2,$$

and because any bounded variable is $O_p (1)$,

$$\rho_1 \left(\frac{\Psi Z' Z \Psi'}{NT} \right) = O_p \left(C_{NT}^{-2} \right) \rho_1 \left(\frac{\Psi \Psi'}{N} \right) = O_p \left(C_{NT}^{-2} \right).$$

Therefore, from (A.2) and (A.3), (A.1) is clear, and the lemma is proved. $\square$

Proof of Theorem 1.1 (Identification of Estimated Factors: Known Potential Leaders): Begin under the conditions for *(i)*. P_{jt} must be correlated with at least one of the latent factors, F_{st}. It is only necessary to show that $\hat{y}^o_{sj,it}$ has a weak factor structure. Express $\hat{y}^o_{sj,it}$ as

$$\hat{y}^o_{sj,it} = y_{it} - \hat{\beta}_{s,ji} P_{jt} - \hat{\lambda}'_{i,-s} \hat{F}_{-s,t} = \lambda'_i F_t - \hat{\beta}_{s,ji} P_{jt} - \hat{\lambda}'_{i,-s} \hat{F}_{-s,t} + y^o_{sj,it}.$$

If P_{jt} is a leader, then there exists a rotation of the latent factors, H^*, which aligns $\left[P_{jt}, \hat{F}'_{-s,t} \right]'$ with the latent factors,

$$H^* = \left(F'F \right)^{-1} F' \left[P_j, \hat{F}_{-s} \right],$$

There is a 'better' rotation (in terms of minimizing the sum of squared residuals), but $\hat{y}^o_{sj,it}$ has a weak factor structure using only the 'poor' rotation H^*, as is shown here. $\|H^*\|_2 = O_p(1)$ and $\|H^{*-1}\|_2 = O_p(1)$ follows from Stock and Watson (1998) and Bai and Ng (2002). Thus,

$$\hat{y}^o_{sj,it} = \lambda'_i F_t - \left[\hat{\beta}_{s,ji}, \hat{\lambda}'_{i,-s} \right] \left(\left[P_{jt}, \hat{F}'_{-s,t} \right]' - H^{*\prime} F_t \right) - \left[\hat{\beta}_{s,ji}, \hat{\lambda}'_{i,-s} \right] H^{*\prime} F_t + y^o_{sj,it}$$
$$= \left(\lambda'_i H^{*-1\prime} - \left[\hat{\beta}_{s,ji}, \hat{\lambda}'_{i,-s} \right] \right) H^{*\prime} F_t - \left[\hat{\beta}_{s,ji}, \hat{\lambda}'_{i,-s} \right] \left(\left[G_t, \hat{F}'_{-s,t} \right]' - H^{*\prime} F_t \right) + y^o_{sj,it}$$

Using the proper alignment H^*,

$$\left\| \left[P_{jt}, \hat{F}'_{-s,t} \right]' - H^{*\prime} F_t \right\|_2 = O_p \left(N^{-1/2} \right)$$

follows from Bai and Ng (2003). Since, the factor is accurately estimated,

$$\left\| \lambda'_i H^{*-1\prime} - \left[\hat{\beta}_{s,ji}, \hat{\lambda}'_{i,-s} \right] \right\|_2 = O_p \left(T^{-1/2} \right),$$

follows from a simple least-squares analysis. Recognizing that $\left\| \left[\hat{\beta}_{s,ji}, \hat{\lambda}'_{i,-s} \right] \right\|_2 = O_p(1)$ and $\|H^{*\prime} F_t\|_2 = O_p(1)$, it is evident that $\hat{y}^o_{sj,it}$ has a weak factor structure. Therefore, $\hat{\#} \left(\hat{y}^o_{sj,it} \right) \to 0$ as $N, T \to \infty$. Under the conditions for *(ii)*, it is clear that the factors in $\hat{y}^o_{sj,it}$ do not vanish as $N, T \to \infty$ (the proof is straightforward).$\square$

APPENDIX B

TECHNICAL APPENDIX FOR CHAPTER 2

Common factor identification in this paper is much easier than in Parker and Sul (2013) because this is the single factor case. Consider the following single factor model defined in (2.1) where F_t is a univariate time series of latent factors observed at time t, λ_i is a collection of loadings to the latent factor for the ith individual, and y_{it}^o is the idiosyncratic error. Assume, without loss of generality, that y_{it} and F_t have been standardized.[1] $\hat{F}_t$ is the principal component estimator of F_t. Note that $\hat{F}_t$ may be different from F_t in sign. $\#\left(y_{it}\right)$ is the true number of common factors of y_{it} and $\hat{\#}\left(y_{it}\right)$ is the estimated number of common factors of y_{it}. Factor number estimation is defined explicitly in Bai and Ng (2002).

Assuming that F_t is observed, consider performing a seemingly unrelated regression of y_{it} on F_t, that is regress each individual time series in the panel, $y_{i\cdot} = (y_{i1}, \ldots, y_{iT})$, on the latent factor time series, $F = (F_1, \ldots, F_T)$. The residuals become:

$$\hat{y}_{it}^o = y_{it} - \hat{\lambda}_i F_t = y_{it}^o + (\lambda_i - \hat{\lambda}_i)F_t = y_{it}^o + O_p\left(T^{-1/2}\right).$$

Note that while $\hat{y}_{it}^o$ may have a common factor in the finite sample, as $N, T \to \infty$ any common factors in $\hat{y}_{it}^o$ must vanish because y_{it}^o has no common factors.

The natural question follows: Does factor number estimation detect those factors which asymptotically vanish? In order to answer this question, Parker and Sul (2013) defines the asymptotically weak factor as a factor structure which vanishes at the rate C_{NT}^{-1} where $C_{NT} = \min\left\{N^{1/2}, T^{1/2}\right\}$. This is formally expressed as the factor model: $x_{it} = \psi_i' Z_t + x_{it}^o$

---

[1]Since there is only one factor, the empirical true factors and statistical true factors defined in Parker and Sul (2013) coincide. Since this is the single factor case, there is no reason for these definitions here, so the true factor is simply referred to as the latent factor.

95

where $\psi_i' Z_t = O_p \left(C_{NT}^{-1} \right)$. Lemma 1 in Parker and Sul (2013) shows that as $N, T \to \infty$ jointly, $\hat{\#} \left(x_{it} \right) \to^p 0$. Therefore, as $N, T \to \infty$ jointly, $\hat{\#} \left(\hat{y}_{it}^o \right) \to^p 0$.

Alternatively, consider regressing the panel on some other time series, $L_t = \delta F_t + v_t$ where $v_t = o_p \left(1 \right)$:

$$y_{it} = \gamma_i L_t + u_{it}.$$

Now, the residual becomes

$$\hat{u}_{it} = y_{it} - \hat{\gamma}_i L_t = u_{it} + \left(\gamma_i - \hat{\gamma}_i \right) L_t.$$

However, because $\hat{\gamma}_i$ is inconsistent, $\hat{\gamma}_i - \gamma_i = o_p \left(1 \right)$ and $\hat{u}_{it}$ has a significant common factor. Hence is easy to show that as $N, T \to \infty$ jointly, $\Pr \left[\hat{\#} \left(\hat{u}_{it} \right) = 0 \right] = 0$.

For more information on common factor identification, including identification in the r factor case, see Parker and Sul (2013).

APPENDIX C

LEMMAS AND PROOFS FOR CHAPTER 3

Define μ_k as the non-central moments of beta random variable x, which is given by

$$\mu_k = \mathrm{E}\left[x^k\right] = \prod_{i=0}^{k-1} \frac{\alpha_x + i}{\alpha_x + \beta_x + i}.$$

Next, define

$$\boldsymbol{\Omega}_x = \mathbf{J}^o \boldsymbol{\Omega}_x^o \mathbf{J}^{o\prime},$$

where

$$\mathbf{J}^o = \left[\begin{array}{ccc} \left(\mu_2 - \mu_1^2\right)^{-1} & 0 & -\left(\mu_2 - \mu_1^2\right)^{-2} \mu_1 \left(\mu_1 - \mu_2\right) \\ 0 & \left(\mu_2 - \mu_1^2\right)^{-1} & -\left(\mu_2 - \mu_1^2\right)^{-2} \left(1 - \mu_1\right) \left(\mu_1 - \mu_2\right) \end{array} \right],$$

and

$$\boldsymbol{\Omega}_x^o = \left[\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{array} \right],$$

with

$$\sigma_{11} = \tau_4 + 4\left(\mu_2 - \mu_3\right)\mu_1^2 - 4\mu_1\mu_2^2,$$

$$\sigma_{12} = \tau_1 - \tau_4 - \tau_2 + 6\mu_1^2\mu_3 + \left(1 + 2\mu_2\right)\mu_1\mu_2 - \mu_2^2,$$

$$\sigma_{13} = -\tau_1 - 2\tau_3 - 2\left(1 + 2\mu_1\right)\mu_1^2\mu_2 + 2\mu_1\mu_3,$$

$$\sigma_{22} = 2\tau_1 + \tau_4 + 2\tau_2 - \left(1 + 4\mu_2\right)\mu_1^2 + \left(1 + \mu_2\right)\mu_2 - 2\mu_3 + \mu_4,$$

$$\sigma_{23} = \tau_1 + 2\tau_3 + 2\left(1 + 2\mu_2\right)\mu_1^3 + \left(3\mu_2 - 5\mu_1\right)\mu_2 + \mu_3 - \mu_4,$$

$$\sigma_{33} = 4\left(3\mu_2 - \mu_1^2\right)\mu_1^2 - 4\mu_1\mu_3 - 5\mu_2^2 + \mu_4,$$

97

where $\tau_1 = \mu_1\mu_4 + \mu_2\mu_3 - 2\mu_3\mu_1^2$, $\tau_2 = 3\mu_1\mu_3 + 2\mu_1^3 - \mu_1\mu_2$, $\tau_3 = 3\mu_1^2\mu_2 - 3\mu_1\mu_2^2 - 2\mu_1^4$, $\tau_4 = \mu_2^3 - 4\mu_1^4 - 4\mu_1^2\mu_2^2 + 8\mu_1^3\mu_2 + \mu_1^2\mu_4 + 2\mu_1\mu_2\mu_3$. $\mathbf{\Omega}_y$ is similarly defined as Ω_x so the formula for $\mathbf{\Omega}_y$ is omitted to save space.

For Lemma 1, we further define $n_s = n - s$, e.g., $n_1 = n - 1$, and $\bar{x}^k = \left(n^{-1}\sum_{i=1}^{n} x_i\right)^k$.

Lemma 1 (Expected Values of Beta Family Sample Moments) *If $x_1, \ldots, x_n$ are drawn from a random sample beta distribution, then the expectations of the following powers are:*

$$
\begin{aligned}
\mathrm{E}\left[\bar{x}\right] &= \mu_1 \\
n\mathrm{E}\left[\bar{x}^2\right] &= n_1\mu_1^2 + \mu_2 \\
n^2\mathrm{E}\left[\bar{x}^3\right] &= n_1 n_2 \mu_1^3 + 3n_1\mu_1\mu_2 + \mu_3 \\
n^3\mathrm{E}\left[\bar{x}^4\right] &= n_1 n_2 n_3 \mu_1^4 + 6n_1 n_2\mu_1^2\mu_2 + n_1\left(3\mu_2^2 + 4\mu_1\mu_3\right) + \mu_4
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left[\left(s_x^2 + \bar{x}^2\right)\right] &= \mu_2 \\
n\mathrm{E}\left[\bar{x}\left(s_x^2 + \bar{x}^2\right)\right] &= n_1\mu_1\mu_2 + \mu_3 \\
n^2\mathrm{E}\left[\bar{x}^2\left(s_x^2 + \bar{x}^2\right)\right] &= n_1 n_2\mu_1^2\mu_2 + n_1\left(2\mu_1\mu_3 + \mu_2^2\right) + \mu_4 \\
n^3\mathrm{E}\left[\bar{x}^3\left(s_x^2 + \bar{x}^2\right)\right] &= n_1 n_2 n_3\mu_1^3\mu_2 + 3n_1 n_2\left(\mu_1^2\mu_3 + \mu_1\mu_2^2\right) + n_1\left(4\mu_2\mu_3 + 3\mu_1\mu_4\right) + \mu_5 \\
n\mathrm{E}\left[\left(s_x^2 + \bar{x}^2\right)^2\right] &= n_1\mu_2^2 + \mu_4 \\
n^2\mathrm{E}\left[\bar{x}\left(s_x^2 + \bar{x}^2\right)^2\right] &= n_1 n_2\mu_1\mu_2^2 + n_1\left(\mu_1\mu_4 + 2\mu_2\mu_3\right) + \mu_5 \\
n^3\mathrm{E}\left[\bar{x}^2\left(s_x^2 + \bar{x}^2\right)^2\right] &= n_1 n_2\left[\mu_1^2\left(n_3\mu_2^2 + \mu_4\right) + \mu_2\left(\mu_2^2 + 4\mu_1\mu_3\right)\right] + n_1\left(3\mu_2\mu_4 + 2\mu_3^2 + 2\mu_1\mu_5\right) + \mu_6
\end{aligned}
$$

The proofs of Lemma 1 are straightforward hence omitted.

Lemma 2 (Limiting Distribution of Method of Moments Components) *If $x_1, \ldots, x_n$ are drawn from a random sample beta distribution, then*

$$
\sqrt{n}\left(\left[\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\right]' - [\theta_1, \theta_2, \theta_3]'\right) \xrightarrow{d} \mathcal{N}\left([0,0,0]', \mathbf{\Omega}_x^o\right).
$$

where

$$\theta_1 = \mu_1 \left(\mu_1 - \mu_2 \right), \ \theta_2 = \left(1 - \mu_1 \right) \left(\mu_1 - \mu_2 \right), \ \theta_3 = \mu_2 - \mu_1^2,$$

and

$$\hat{\theta}_1 = \left(\bar{x} \right) \left(\bar{x} \left(1 - \bar{x} \right) - s_x^2 \right), \ \hat{\theta}_2 = \left(1 - \bar{x} \right) \left(\bar{x} \left(1 - \bar{x} \right) - s_x^2 \right), \ \hat{\theta}_3 = s_x^2.$$

Proof of Lemma 2

By using Lemma 1, it is straightforward to show that:

$$
\begin{aligned}
\mathrm{E}\left[\hat{\theta}_1 \right] &= \frac{n_1}{n} \left(\mu_1^2 - \mu_1 \mu_2 \right) + \frac{1}{n} \left(\mu_2 - \mu_3 \right), \\
\mathrm{E}\left[\hat{\theta}_2 \right] &= \left(\mu_1 - \mu_2 \right) + \frac{n_1}{n} \left(\mu_1 \mu_2 - \mu_1^2 \right) + \frac{1}{n} \left(\mu_3 - \mu_2 \right), \\
\mathrm{E}\left[\hat{\theta}_3 \right] &= \frac{(n+1)}{n} \mu_2 - \frac{n_1}{n} \mu_1^2,
\end{aligned}
$$

and

$$
\begin{aligned}
n\mathrm{Var}\left[\hat{\theta}_1 \right] &= \mu_2^3 - 4\mu_1^4 - 4\mu_1^2 \mu_2^2 - 4\mu_1 \mu_2^2 + 4\mu_1^2 \mu_2 - 4\mu_1^2 \mu_3 + 8\mu_1^3 \mu_2 \\
&\quad + \mu_1^2 \mu_4 + 2\mu_1 \mu_2 \mu_3 + O\left(n^{-1} \right), \\
n\mathrm{Var}\left[\hat{\theta}_2 \right] &= \mu_2 - 2\mu_3 + \mu_4 - \mu_1^2 + 4\mu_1^3 + \mu_2^2 - 4\mu_1^4 + \mu_2^3 - 4\mu_1^2 \mu_2^2 \\
&\quad - 2\mu_1 \mu_2 + 6\mu_1 \mu_3 - 2\mu_1 \mu_4 - 2\mu_2 \mu_3 - 4\mu_1^2 \mu_2 - 4\mu_1^2 \mu_3 \\
&\quad + 8\mu_1^3 \mu_2 + \mu_1^2 \mu_4 + 2\mu_1 \mu_2 \mu_3 + O\left(n^{-1} \right), \\
n\mathrm{Var}\left[\hat{\theta}_3 \right] &= -4\mu_1^4 + 12\mu_1^2 \mu_2 - 4\mu_3 \mu_1 - 5\mu_2^2 + \mu_4 + O\left(n^{-1} \right), \\
n\mathrm{Cov}\left[\hat{\theta}_1, \hat{\theta}_2 \right] &= 4\mu_1^4 - \mu_2^2 - 2\mu_1^3 - \mu_2^3 + 4\mu_1^2 \mu_2^2 + 2\mu_1 \mu_2 - 3\mu_1 \mu_3 + \mu_1 \mu_4 \\
&\quad + \mu_2 \mu_3 + 2\mu_1 \mu_2^2 + 4\mu_1^2 \mu_3 - 8\mu_1^3 \mu_2 - \mu_1^2 \mu_4 - 2\mu_1 \mu_2 \mu_3 \\
&\quad + O\left(n^{-1} \right),
\end{aligned}
$$

$$
\begin{aligned}
n\mathrm{Cov}\left[\hat{\theta}_1, \hat{\theta}_3\right] &= 4\mu_1^4 + 2\mu_1\mu_3 - \mu_1\mu_4 - \mu_2\mu_3 + 6\mu_1\mu_2^2 - 8\mu_1^2\mu_2 + 2\mu_1^2\mu_3 \\
&\quad -4\mu_1^3\mu_2 + O\left(n^{-1}\right), \\
n\mathrm{Cov}\left[\hat{\theta}_2, \hat{\theta}_3\right] &= -4\mu_1^4 + 4\mu_1^3\mu_2 + 2\mu_1^3 + 6\mu_1^2\mu_2 - 2\mu_3\mu_1^2 - 6\mu_1\mu_2^2 - 5\mu_1\mu_2 \\
&\quad +\mu_4\mu_1 + 3\mu_2^2 + \mu_3\mu_2 + \mu_3 - \mu_4 + O\left(n^{-2}\right).
\end{aligned}
$$

Multiplying by n and taking the limit as $n \to \infty$ yields the variance-covariance matrix expressed above. By taking transformations, we can apply the Lindeberg-Levy central limit theorem and the statement is proved.□

Lemma 3 (Limiting Distribution of the Method of Moments Estimator)   *If* $x_1, \ldots, x_n$ *are drawn from a random sample beta distribution with parameters* α_x *and* β_x, *then*

$$
\sqrt{n}\left(\left[\hat{\alpha}_x, \hat{\beta}_x\right]' - [\alpha_x, \beta_x]'\right) \xrightarrow{d} \mathcal{N}\left([0,0]', \boldsymbol{\Omega}_x\right).
$$

Proof of Lemma 3

Let $g_1, g_2 : \mathbb{R}^3 \to \mathbb{R}$ be defined by:

$$
g_1\left(\theta_1, \theta_2, \theta_3\right) = \frac{\theta_1}{\theta_3}, \quad g_2\left(\theta_1, \theta_2, \theta_3\right) = \frac{\theta_2}{\theta_3}.
$$

Differentiating, we find that:

$$
\begin{aligned}
\frac{\partial g_1\left(\theta_1, \theta_2, \theta_3\right)}{\partial x} &= \frac{1}{\theta_3}, \quad \frac{\partial g_1\left(\theta_1, \theta_2, \theta_3\right)}{\partial y} = 0, \quad \frac{\partial g_1\left(\theta_1, \theta_2, \theta_3\right)}{\partial \theta_3} = -\frac{\theta_1}{\theta_3^2}, \\
\frac{\partial g_2\left(x, y, z\right)}{\partial x} &= 0, \quad \frac{\partial g_2\left(x, y, z\right)}{\partial y} = \frac{1}{\theta_3}, \quad \frac{\partial g_2\left(x, y, z\right)}{\partial z} = -\frac{\theta_2}{\theta_3^2}.
\end{aligned}
$$

The Jacobian, $\mathbf{J}_x^o$, consists of the above partial derivatives evaluated at the following values:

$$
\begin{aligned}
\theta_1 &= \lim_{n\to\infty} \mathrm{E}\left[m_1^2 - m_1 m_2\right] = \mu_1\left(\mu_1 - \mu_2\right), \\
\theta_2 &= \lim_{n\to\infty} \mathrm{E}\left[m_1\left(1 - m_1\right) - \left(1 - m_1\right)m_2\right] = \left(1 - \mu_1\right)\left(\mu_1 - \mu_2\right), \\
\theta_3 &= \lim_{n\to\infty} \mathrm{E}\left[m_2 - m_1^2\right] = \mu_2 - \mu_1^2.
\end{aligned}
$$

From here, we can use the delta method and Lemma 2 to find the limiting distribution of $\left[\hat{\alpha}_x, \hat{\beta}_x\right]'$, and the lemma is proved.□

Proof of Theorem 1

The result in Part *(i)* is a direct consequence of Lemma 5. Parts *(ii)* and *(iii)* are proved by applying the delta method. If a beta distribution is symmetric, then $\alpha = \beta$. Hence, we are interested in testing if $\alpha - \beta = 0$. Let $\mathbf{J}$ be the Jacobian of the transformation. So, $\mathbf{J} = [1, -1]'$. Therefore under the null of symmetry, by the delta method,

$$\sqrt{n} \left(\mathbf{J}' \left[\hat{\alpha}_x, \hat{\beta}_x \right]' \right) = \sqrt{n} \left(\hat{\alpha}_x - \hat{\beta}_x \right) \xrightarrow{d} \mathcal{N} \left(0, \mathbf{J}' \mathbf{\Omega}_x \mathbf{J} \right),$$

and

$$z_{\text{Sym}(X)} \xrightarrow{d} \mathcal{N} (0, 1).$$

The $\mathcal{H}_{2O} : \mu_x = 0.5$ test can be proved similarly.

Similarly, the Jacobian for the beta method of moments skewness is,

$$\mathbf{J}_x = \frac{(\alpha_x + \beta_x) \sqrt{\alpha_x \beta_x}}{(\alpha_x + \beta_x + 2)^2 \sqrt{\alpha_x + \beta_x + 1}} \left[-\frac{(\beta_x + 1)(3\alpha_x + \beta_x + 2)}{\alpha_x}, \frac{(\alpha_x + 1)(\alpha_x + 3\beta_x + 2)}{\beta_x} \right]'.$$

Hence, under the null that the distributions have equal skewness,

$$z_{\text{Skew}} \xrightarrow{d} \mathcal{N} (0, 1),$$

and the theorem is proved.$\square$

Notation

Denote $\mu_{i,j}$ and $\bar{\mu}_{i,j}$ as the (i, j)th non-central and central co-moment of (X, Y) respectively. So $\bar{\mu}_{i,j} = E \left[(x - \mu_x)^i (y - \mu_y)^j \right]$, and since the two samples are not assumed to be independent, this cannot be further reduced to a product of two expectations. $\bar{x}$ are s_x^2 are the usual sample mean and sample variance for the $\mathbf{x} = (x_1, \ldots, x_n)$.

Lemma 4 (General Limiting Distribution of the Mean and Variance) *Where $\{(x_i, y_i)\}$ is a size n random sample from a bivariate distribution with finite fourth moments, under the null that x and y have the same mean and variance,*

$$\sqrt{n}\left[\bar{x} - \bar{y}, s_x^2 - s_y^2\right]' \xrightarrow{d} \mathcal{N}\left([0, 0]', \Xi\right),$$

where

$$\Xi = \begin{bmatrix} \bar{\mu}_{2,0} + \bar{\mu}_{0,2} - 2\bar{\mu}_{1,1} & \bar{\mu}_{3,0} + \bar{\mu}_{0,3} - \bar{\mu}_{1,2} - \bar{\mu}_{2,1} \\ \bar{\mu}_{3,0} + \bar{\mu}_{0,3} - \bar{\mu}_{1,2} - \bar{\mu}_{2,1} & \bar{\mu}_{4,0} - \bar{\mu}_{2,0}^2 + \bar{\mu}_{0,4} - \bar{\mu}_{0,2}^2 - 2\bar{\mu}_{2,2} + 2\bar{\mu}_{0,2}\bar{\mu}_{2,0} \end{bmatrix}$$

Proof of Lemma 4

It is well-known that

$$\mathrm{E}\left[\bar{x}\right] = \mu_{1,0}, \quad \mathrm{E}\left[s_x^2\right] = \bar{\mu}_{2,0},$$

and

$$n\mathrm{Var}\left[\bar{x}\right] = \bar{\mu}_{2,0}, \quad n\mathrm{Var}\left[s_x^2\right] = \bar{\mu}_{4,0} - \frac{n-3}{n-1}\bar{\mu}_{2,0}^2.$$

It is also easy to derive:

$$n\mathrm{Cov}\left[\bar{x}, \bar{y}\right] = \bar{\mu}_{1,1}, \quad n\mathrm{Cov}\left[\bar{x}, s_x^2\right] = \bar{\mu}_{3,0}, \quad n\mathrm{Cov}\left[\bar{x}, s_y^2\right] = \bar{\mu}_{1,2}.$$

The only tedious term is

$$n\mathrm{Cov}\left[s_x^2, s_y^2\right] = \bar{\mu}_{2,2} - \bar{\mu}_{0,2}\bar{\mu}_{2,0} + \frac{2}{n-1}\bar{\mu}_{1,1}^2.$$

Using the symmetric form, we know all the components of:

$$\begin{aligned}
\Xi &= \lim_{n\to\infty} \begin{bmatrix} n\mathrm{Var}\left[\bar{x} - \bar{y}\right] & n\mathrm{Cov}\left[\bar{x} - \bar{y}, s_x^2 - s_y^2\right] \\ n\mathrm{Cov}\left[\bar{x} - \bar{y}, s_x^2 - s_y^2\right] & n\mathrm{Var}\left[s_x^2 - s_y^2\right] \end{bmatrix} \\
&= \begin{bmatrix} \bar{\mu}_{2,0} + \bar{\mu}_{0,2} - 2\bar{\mu}_{1,1} & \bar{\mu}_{3,0} + \bar{\mu}_{0,3} - \bar{\mu}_{1,2} - \bar{\mu}_{2,1} \\ \bar{\mu}_{3,0} + \bar{\mu}_{0,3} - \bar{\mu}_{1,2} - \bar{\mu}_{2,1} & \bar{\mu}_{4,0} - \bar{\mu}_{2,0}^2 + \bar{\mu}_{0,4} - \bar{\mu}_{0,2}^2 - 2\bar{\mu}_{2,2} + 2\bar{\mu}_{0,2}\bar{\mu}_{2,0} \end{bmatrix}
\end{aligned}$$

And so, by the Lindeberg-Levy central limit theorem,

$$\sqrt{n} \left[\bar{x} - \bar{y}, s_x^2 - s_y^2\right]' \xrightarrow{d} \mathcal{N}_2 \left([0, 0]', \Xi\right).$$

While the normality does not immediately follow as we have to apply tedious transformations to use the Lindeberg condition, such transformations are beyond the scope of this paper, and do not provide any useful intuition.□

Lemma 5 (General Limiting Distribution of the Skewness) *Where $\{(x_i, y_i)\}_{i=1,\ldots,n}$ is a random sample from a bivariate distribution with finite sixth moments, under the null that x and y have equal skewness,*

$$\sqrt{n} \left(\hat{\kappa}_{3x} - \hat{\kappa}_{3y}\right) \xrightarrow{d} \mathcal{N} \left(0, \omega_x + \omega_y - 2\omega_{xy}\right),$$

where

$$\omega_x = \frac{1}{4\bar{\mu}_{2,0}^5} \left(36\bar{\mu}_{2,0}^5 - 24\bar{\mu}_{4,0}\bar{\mu}_{2,0}^3 + 35\bar{\mu}_{2,0}^2\bar{\mu}_{3,0}^2 + 4\bar{\mu}_{6,0}\bar{\mu}_{2,0}^2 - 12\bar{\mu}_{5,0}\bar{\mu}_{2,0}\bar{\mu}_{3,0} + 9\bar{\mu}_{4,0}\bar{\mu}_{3,0}^2\right),$$

and

$$\begin{aligned}
\omega_{xy} &= \frac{1}{4\bar{\mu}_{0,2}^{5/2}\bar{\mu}_{2,0}^{5/2}} \left(12\bar{\mu}_{0,2}^2\bar{\mu}_{2,0}\bar{\mu}_{3,1} - 18\bar{\mu}_{0,2}^2\bar{\mu}_{2,1}\bar{\mu}_{3,0} - 18\bar{\mu}_{2,0}^2\bar{\mu}_{0,3}\bar{\mu}_{1,2} + 12\bar{\mu}_{0,2}\bar{\mu}_{2,0}^2\bar{\mu}_{1,3}\right) \\
&\quad + \frac{1}{4\bar{\mu}_{0,2}^{5/2}\bar{\mu}_{2,0}^{5/2}} \left(-36\bar{\mu}_{0,2}^2\bar{\mu}_{1,1}\bar{\mu}_{2,0}^2 - 4\bar{\mu}_{0,2}\bar{\mu}_{2,0}\bar{\mu}_{3,3} + 6\bar{\mu}_{0,2}\bar{\mu}_{3,0}\bar{\mu}_{2,3} + 6\bar{\mu}_{2,0}\bar{\mu}_{0,3}\bar{\mu}_{3,2}\right) \\
&\quad + \frac{1}{4\bar{\mu}_{0,2}^{5/2}\bar{\mu}_{2,0}^{5/2}} \left(-9\bar{\mu}_{0,3}\bar{\mu}_{3,0}\bar{\mu}_{2,2} + \bar{\mu}_{0,2}\bar{\mu}_{2,0}\bar{\mu}_{0,3}\bar{\mu}_{3,0}\right).
\end{aligned}$$

Proof of Lemma 5

By Taylor expanding around the expectations of the moment estimators, we obtain the approximation

$$\hat{\kappa}_{3x} = \frac{\bar{\mu}_{3,0}}{\bar{\mu}_{2,0}^{3/2}} + \frac{1}{\bar{\mu}_{2,0}^{3/2}} \left(\bar{m}_{3,0} - \bar{\mu}_{3,0}\right) - \frac{3}{2}\frac{\bar{\mu}_{3,0}}{\bar{\mu}_{2,0}^{5/2}} \left(s_x^2 - \bar{\mu}_{2,0}\right) + O_p\left(n^{-1}\right),$$

where $\bar{m}_{3,0} = n (n_1 n_2)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^3$. Notice that $\mathrm{E}\bar{m}_{3,0} = \bar{\mu}_{3,0}$. So it is obvious that $\mathrm{E}[\hat{\kappa}_{3x}] = \kappa_{3x} + O(n^{-1})$. So, the only thorny issue is deriving the variance-covariance matrix of $[\hat{\kappa}_{3x}, \hat{\kappa}_{3y}]'$,

$$\mathrm{Var}[\hat{\kappa}_{3x}] = \frac{1}{\bar{\mu}_{2,0}^3}\mathrm{Var}[\bar{m}_{3,0}] + \frac{9}{4}\frac{\bar{\mu}_{3,0}^2}{\bar{\mu}_{2,0}^5}\mathrm{Var}[s_x^2] - 3\frac{\bar{\mu}_{3,0}}{\bar{\mu}_{2,0}^4}\mathrm{Cov}[\bar{m}_{3,0}, s_x^2] + O(n^{-2}).$$

Although the calculation is rather tedious, it is straightforward to obtain the following,

$$\mathrm{Var}[\bar{m}_{3,0}] = n^{-1}(\bar{\mu}_{6,0} - \bar{\mu}_{3,0}^2 - 6\bar{\mu}_{2,0}\bar{\mu}_{4,0} + 9\bar{\mu}_{2,0}^3) + O(n^{-2}),$$
$$\mathrm{Cov}[\bar{m}_{3,0}, s_x^2] = n^{-1}(\bar{\mu}_{5,0} - 4\bar{\mu}_{2,0}\bar{\mu}_{3,0}) + O(n^{-2}).$$

Hence,

$$\mathrm{Var}[\hat{\kappa}_{3x}] = \frac{1}{4n\bar{\mu}_{2,0}^5}(36\bar{\mu}_{2,0}^5 - 24\bar{\mu}_{4,0}\bar{\mu}_{2,0}^3 + 35\bar{\mu}_{2,0}^2\bar{\mu}_{3,0}^2 + 4\bar{\mu}_{6,0}\bar{\mu}_{2,0}^2 - 12\bar{\mu}_{5,0}\bar{\mu}_{2,0}\bar{\mu}_{3,0} + 9\bar{\mu}_{4,0}\bar{\mu}_{3,0}^2)$$
$$+ \frac{1}{4n\bar{\mu}_{2,0}^5}(-12\bar{\mu}_{5,0}\bar{\mu}_{2,0}\bar{\mu}_{3,0} + 9\bar{\mu}_{4,0}\bar{\mu}_{3,0}^2) + O(n^{-2}).$$

To find the covariance, we compute:

$$\mathrm{Cov}[\bar{m}_{3,0}, \bar{m}_{0,3}] = n^{-1}(\bar{\mu}_{3,3} - \bar{\mu}_{0,3}\bar{\mu}_{3,0} - 3\bar{\mu}_{0,2}\bar{\mu}_{3,1} - 3\bar{\mu}_{1,3}\bar{\mu}_{2,0} + 9\bar{\mu}_{0,2}\bar{\mu}_{1,1}\bar{\mu}_{2,0}) + O(n^{-2}),$$
$$\mathrm{Cov}[\bar{m}_{3,0}, s_y^2] = n^{-1}(\bar{\mu}_{3,2} - \bar{\mu}_{0,2}\bar{\mu}_{3,0} - 3\bar{\mu}_{1,2}\bar{\mu}_{2,0}) + O(n^{-2}).$$

So,

$$\mathrm{Cov}[\hat{\kappa}_{3x}, \hat{\kappa}_{3y}] = \frac{1}{4n\bar{\mu}_{0,2}^{5/2}\bar{\mu}_{2,0}^{5/2}}(12\bar{\mu}_{0,2}^2\bar{\mu}_{2,0}\bar{\mu}_{3,1} - 18\bar{\mu}_{0,2}^2\bar{\mu}_{2,1}\bar{\mu}_{3,0} - 18\bar{\mu}_{2,0}^2\bar{\mu}_{0,3}\bar{\mu}_{1,2} + 12\bar{\mu}_{0,2}\bar{\mu}_{2,0}^2\bar{\mu}_{1,3}) +$$
$$+ \frac{1}{4n\bar{\mu}_{0,2}^{5/2}\bar{\mu}_{2,0}^{5/2}}(-36\bar{\mu}_{0,2}^2\bar{\mu}_{1,1}\bar{\mu}_{2,0}^2 - 4\bar{\mu}_{0,2}\bar{\mu}_{2,0}\bar{\mu}_{3,3} + 6\bar{\mu}_{0,2}\bar{\mu}_{3,0}\bar{\mu}_{2,3} + 6\bar{\mu}_{2,0}\bar{\mu}_{0,3}\bar{\mu}_{3,2})$$
$$+ \frac{1}{4n\bar{\mu}_{0,2}^{5/2}\bar{\mu}_{2,0}^{5/2}}(-9\bar{\mu}_{0,3}\bar{\mu}_{3,0}\bar{\mu}_{2,2} + \bar{\mu}_{0,2}\bar{\mu}_{2,0}\bar{\mu}_{0,3}\bar{\mu}_{3,0}) + O(n^{-2}).$$

Thus, under the null that X and Y have equal skewness,

$$\sqrt{n}(\hat{\kappa}_{3x} - \hat{\kappa}_{3y}) \xrightarrow{d} \mathcal{N}(0, \omega_x + \omega_y - 2\omega_{xy}),$$

and the lemma is proved.$\square$

REFERENCES

[1] Ahn, S.C. and A. R. Horenstein (2013), "Eigenvalue ratio test for the number of factors", *Econometrica*, 1203-1227.

[2] Amengual, D. and M. Watson (2007), "Consistent estimation of the number of dynamic factors in a large N and T panel", *Journal of Business & Economic Statistics* 25, 91-96.

[3] Andreas P. (2009) *Border Games: Policing the U.S.-Mexico Divide.* Cornell UP.

[4] Bai, J. (2003), "Inferential theory for factor models of large dimensions", *Econometrica* 71, 135-171.

[5] Bai, J. (2004), "Estimating cross-section common stochastic trends in nonstationary panel data", *Journal of Econometrics* 122, 137-183.

[6] Bai, J. and S. Ng (2002), "Determining the number of factors in approximate factor models", *Econometrica* 70, 191-221.

[7] Bai, J. and S. Ng (2006), "Evaluating latent and observed factors in macroeconomics and finance", *Journal of Econometrics* 131, 507-537.

[8] Bai, J. and S. Ng (2013), "Principal components estimation and identification of static factors", *Journal of Econometrics* 176, 18-29.

[9] Bates, B., M. Plagborg-Møller, J. Stock and M. Watson (2013), "Consistent factor estimation in dynamic factor models with structural instability" forthcoming in *Journal of Econometrics.*

[10] Choi, I. (2012). "Efficient estimation of factor models," *Econometric Theory*, 274–308.

[11] El-Bassiouny, A.H. and Jones, M.C. (2009) "A bivariate F distribution with marginals on arbitrary numerator and denominator degrees of freedom and related bivariate beta and t distributions." *Stat Methods Appl*, 18, 465-481.

[12] Fama, E. and K. French (1993), "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics.* 33, 3–56.

[13] Fay, Michael P. and Proschan, Michael A. (2010) "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules" *Statistics Survey*, 4, 1-39.

[14] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The generalized dynamic-factor model: identification and estimation", *Review of Economics & Statistics* 82, 540-554.

[15] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005), "The generalized dynamic-factor model: one-sided estimation and forecasting", *Journal of the American Statistical Association* 471, 830-840.

[16] Gaibulloev, K., T. Sandler and D. Sul, (2013), "Common drivers of transnational terrorism: principal component analysis", *Economic Inquiry. 707-721.*

[17] Greenaway-McGrevy, R., Mark, N., Sul, D., and Wu J. (2012), "Role of key currencies in the determination of exchange rates" mimeo, University of Texas at Dallas.

[18] Hallin, M. and R. Liska, (2007), "The generalized dynamic factor model determining the number of factors", *Journal of the American Statistical Association*, 603–617.

[19] Harding, M. (2013), "Estimating the number of factors in large dimensional factor models" mimeo, Stanford University.

[20] Kapetanios, G. (2010), "A testing procedure for determining the number of factors in approximate factor models with large datasets", *Journal of Business & Economic Statistics*, 28, 397–409.

[21] Ludwigson, S. and S. Ng (2007), "The empirical risk–return relation: a factor analysis approach", *Journal of Financial Economics* 83, 171-222.

[22] Onatski, A. (2009), "Testing hypotheses about the number of factors in large factor models", *Econometrica* 77, 1447-1479.

[23] Perez, M. and S.C. Ahn (2007), "GMM estimation of the number of latent factors " mimeo, Arizona State University.

[24] Phillips, P.C.B and D. Sul (2007), "Transition modeling and econometric convergence tests", *Econometrica*, 75, 1771-1855.

[25] Reis, R. and M. Watson (2010), "Relative goods' prices, pure inflation, and the phillips correlation", *American Economic Journal: Macroeconomics* 2, 128-157.

[26] Stock J. and M. Watson (2002a), "Forecasting using principal components from a large number of predictors", *Journal of American Statistical Association* 97, 1167-1179.

[27] Stock J. and M. Watson (2002b), "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Statistics* 20, 147-162.

[28] Roth, Alvin E., Prasnikar, Vesna, Okuno-Fujiwara, Masahiro, and Zamir, Shmuel (1991) "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study." *American Economic Review*, 81 (5), 1068-1095.

[29] Bai J. (2003) "Inferential Theory for Factor Models of Large Dimensions," *Econometrica* 71: 135-171.

[30] Bai, J. and S. Ng (2002) "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70: 191-221.

[31] Bai, J. and S. Ng (2006) "Evaluating Latent and Unobserved Factors in Macroeconomics and Finance," *Journal of Econometrics* 131: 507-537.

[32] Bernanke, B., J. Boivin and P. Eliasz (2005) "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics* 120: 387-422.

[33] Department of Justice, U.S. (2011) "2011 Drug Threat Assessment."

[34] Donohue, J. and S. Levitt (2001) "The Impact of Legalized Abortion on Crime," *Quarterly Journal of Economics* 116: 379-420.

[35] Fama, E. and K. French (1992) "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.

[36] Fryer R., P. Heaton, S. Levitt, and K. Murphy (2013) "Measuring Crack Cocaine and Its Impact," *Economic Inquiry* 51: 1651-1681.

[37] Gupta, Arjun K. and Nadarajah, Saralees. (2004) *Handbook of Beta Distribution and Its Applications.* Marcel Dekker, Inc.

[38] Gupta, Arjun K. and Wong, C.F. (1985) "On three and five parameter bivariate beta distributions." *Metrika*, 32, 85-91.

[39] Government Accountability Office, U.S. (2006) "Border-Crossing Deaths Have Doubled Since 1995; Border Patrol's Efforts to Prevent Deaths Have Not Been Fully Evaluated," p.42.

[40] Hallin, M. and R. Liska (2007) "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association* 102, 603-617.

[41] Ho, Teck-Hua and Su, Xuanming (2009) "Peer-induced fairness in games." *American Economic Review* 99 (5), 2022-2049.

[42] Levitt, S. (2004) "Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not," *Journal of Economic Perspectives* 18: 163-190.

[43] MacGillivray, H.L. (1981) "The mean, median, mode inequality and skewness for a class of densities." *Australian Journal of Statistics*, 23 (2), 247-250.

[44] Mann, H.B. and Whitney, D.R. (1947) "On a test of whether one of two random variables is stochastically larger." *Annals of Mathematical Statistics*, 18 (1), 50-60.

[45] Marsh, Patrick (2010) "A two-sample nonparametric likelihood ratio test." *Journal of Nonparametric Statistics*, 22 (8) Nov. 2010, 1053-1065.

[46] Moody, C. and T. Marvell (2010) "On the Choice of Control Variables in the Crime Equation," *Oxford Bulletin of Economics and Statistics* 72: 696-715.

[47] Nadarajah, Saralees and Kotz, Samuel (2005) "Some bivariate beta distributions." *Statistics*, 39 (5), 457-466

[48] Nadarajah, Saralees (2005) "Reliability for some bivariate beta distributions." *Mathematical Problems in Engineering*, 1, 101-111.

[49] Office of National Drug Control Policy, U.S. (2013) "Texas Drug Control Update," White House. 08 Oct. 2013.

[50] Parker, J. and D. Sul (2013) "Identification of Common Factors: Leaders and Followers," mimeo University of Texas at Dallas.

[51] Pesaran, M. (2006) "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967-1012.

[52] Stock, J. and M. Watson (2005) "Implication of Dynamic Factor Models for VAR Analysis," NBER Working Paper.

[53] Wilcoxon, Frank (1945) "Individual comparisons by ranking methods." *Biometrics Bulletin*, 1 (6), 80-83.

VITA

Jason Parker earned his undergraduate degree in mathematics with a specialization in economics in three years. Having graduated from The University of Chicago in 2009, he went on to earn a master's of science in economics at The University of Texas at Dallas in 2010. His research interests include panel data econometrics, panel information forecasting, and the economics of crime. Upon completion of his Ph.D., Jason Parker will begin a post-doctoral fellowship at Michigan State University and the North Central Regional Center for Rural Development (NCRCRD).