Statistica Sinica **7**(1997), 1021-1038

# WEIGHTED DEGENERATE U- AND V-STATISTICS WITH ESTIMATED PARAMETERS

Grace S. Shieh

*Academia Sinica*

*Abstract:* A new test statistic based on Pearson correlation coefficient ($r_n$) for cross-sectional correlation in panel data is studied. The limiting distribution of $r_n$ is presented. A simulation study shows that $r_n$ is useful in general situations. Computing $r_n$ on the residuals from models for panel data motivates the study of a new class of statistics, namely weighted degenerate U-statistics with estimated parameters ($WU_n(\hat{\lambda})$). The limiting distributions of $WU_n(\hat{\lambda})$ and weighted degenerate V-statistics with estimated parameters are established. Whether their limiting distributions are affected by using estimates of the parameters depends on whether or not a certain mean function has zero derivative. Applications to testing correlation in panel data and to testing for goodness-of-fit are presented.

*Key words and phrases:* Asymptotics, correlation, residuals, U-statistics.

## 1. Introduction

Panel data (also called longitudinal data) are measurements taken from experimental units (typically individuals or economic entities) over time. For the $i$th experimental unit, a set of panel data over time periods may be denoted by $Y_{it}$ where $t = 1, \ldots, T$. For further details of panel data, we refer to Hsiao (1986) and Greene (1993).

Cross-sectional correlation occurs often in the social sciences due to a common environment. For instance, two automobile manufacturing companies operating in a common economic environment such as the same market and the same work rules will show correlation in their trading behaviors.

According to Frees (1995), experimenters in social sciences are reluctant to consider the data over long period of time because basic conditions appear unstable. In this paper, we consider situations in which the time period for the panel data ($T$) is short whereas the number of experimental units ($n$) is large. For the importance of this case, see Hsiao (1986).

Classical regression analysis assumes that experimental units are unrelated; however, with panel data this is often not the case. Frees (1992, 1993) showed that ignoring cross-sectional correlation results in overstated prediction bands in interstate migration rates forecasts. Hence, assessing cross-sectional correlation in panel data is important.

In this paper, we first study a new test statistic $r_n$ for cross-sectional correlation with panel data. The asymptotic and finite sample properties of $r_n$ are provided. Then we investigate the asymptotics of a class of new statistics, namely weighted degenerate U-statistics with parameters, which was obtained by computing $r_n$ on residuals from models for panel data.

To illustrate a new test statistic $r_n$ proposed in this paper, we first consider a simple model $Y_{it} = \mu_i + \sigma_i e_{it},\ i = 1,\ldots,n,\ t = 1,\ldots,T$, where $\{Y_{it}\}, t = 1,\ldots,T$ are observations from the $i$th experimental unit. Assume that $\{e_{it}\}$ are mean zero unobservable random variables (r.v.'s). We note that $\mu_i$ could be random or fixed effects because the correlation of panel data is not affected by location and scale parameters. We consider more general models in Section 3.

The cross-sectional correlation between the $i$th and the $j$th experimental units is defined as $\rho_{ij} = \mathrm{Corr}(Y_{it}, Y_{jt})$. To determine whether there is cross-sectional correlation, we shall test the null hypothesis that $H_o : \rho_{ij} = 0$, for each $i$ and $j$.

We propose a new test statistic for $H_0$ which is based on the Pearson correlation coefficient, and call it $r_n$.

$$r_n = \frac{\sum_{i\neq j}\ (S_{ij}^2 - S_i^2\ S_j^2)}{\sum_{i\neq j}\ S_i^2\ S_j^2}, \tag{1}$$

where $S_{ij}^2 = \sum_{t=1}^T Y_{it}^2 Y_{jt}^2/T$ and ${S_i}^2 = \sum_{t=1}^T Y_{it}^2/T$. For a statistic of the Pearson correlation coefficient type and other rank test statistics for cross-sectional correlation with panel data, we refer to Frees (1995).

In Section 2, the null limiting distribution (limiting distribution henceforth) and a simulation of the properties of $r_n$ are presented. The simulation study shows that under various situations, $r_n$ is more powerful than its rank version studied in Frees (1995).

Since, in econometrics, the statistic is primarily used for model diagnostics, the statistic based on residuals is likely to be useful. Furthermore, Randles (1984) shows that under certain general conditions, test statistics computed on additive residuals have the same limiting distributions as those based on independent observations. This led the author to compute $r_n$ on the residuals, which motivates a class of new statistics, namely weighted degenerate U-statistics with estimated parameters ($WU_n(\hat{\lambda})$ henceforth). (For details, see Section 3.)

In Section 3, we first establish the limiting distribution of weighted degenerate V-statistics with estimated parameters ($WV_n(\hat{\lambda})$), which is a class of statistics closely related to $WU_n(\hat{\lambda})$. Then the limiting distribution of $WU_n(\hat{\lambda})$ is obtained via $WV_n(\hat{\lambda})$. $WU_n(\hat{\lambda})$ extends the class of weighted degenerate U-statistics in Shieh, Johnson and Frees (1994). Conditions under which the use of parameter estimates affects the limiting distributions are specified. Two applications are given. We conclude with some remarks in Section 4.

## 2. The Test Statistic $r_n$

Recall that the new statistic proposed in this paper, $r_n$, assumes the form $r_n = \sum_{i\neq j}(S_{ij}^2 - S_i^2\ S_j^2)/[\sum_{i\neq j}\ S_i^2\ S_j^2]$, where $S_{ij}^2 = \sum_{t=1}^T Y_{it}^2 Y_{jt}^2/T$ and $S_i{}^2 = \sum_{t=1}^T Y_{it}^2/T$. $r_n$ is based on the Pearson correlation coefficient. If for $t = 1, \ldots, T$ the $\{Y_{it} :\ i = 1, \ldots, n\}$ are independent mean zero variables with variance $\sigma_i^2$, then $E\sum_{i\neq j}(S_{ij}^2 - S_i^2 S_j^2) = 0$, and it can be shown that $|r_n| \leq 1$. Thus the statistic $r_n$ can be used to detect deviations from the independence assumption. Frees (1995) studied a rank version of $r_n$, $R_{AVE}^2$, which will be compared to $r_n$ in a simulation study in the next section.

The statistic $r_n$ is the ratio of a degenerate U-statistic to a non-degenerate U-statistic and thus can be analyzed by results for degenerate U-statistics (Gregory (1977) or Serfling (1980)). We establish its limiting distribution in the next section.

### 2.1. The limiting distribution of $r_n$

Let $U_n = [n(n-1)]^{-1}\sum_{i\neq j}\ (S_{ij}^2 - S_i^2 S_j^2)$ and $H_n = [n(n-1)]^{-1}\sum_{i\neq j}\ S_i^2 S_j^2$, where $\{S_{ij}^2\}$ and $\{S_i^2\}$ are defined under the equation (1). It can be checked that $n[r_n - U_n/E(H_n)] \rightarrow_D 0$, provided that

$$H_n - E(H_n) \rightarrow_P 0 \text{ and } E(H_n) \rightarrow\ C_H, \tag{2}$$

where $C_H$ is a positive constant. The implications in (2) follow from a result for non-degenerate U-statistics (Theorem 5.4.A in Serfling (1980)). After straightforward algebra, it can be shown that

$$nr_n \rightarrow_D\ [(T-1)(\mu_4 - \mu_2^2)/(T^2 C_H)](\chi_T^2 - T),$$

where $\mu_4 = E(Y_{11}^4)$, $\mu_2 = E(Y_{11}^2)$ and $\chi_T^2$ is a chi square random variable with $T$ degree of freedom. Frees (1995) showed that

$$n(R_{AVE}^2 - (T-1)^{-1}) \rightarrow_D a(T)[\chi_{T-1}^2 - (T-1)] + b(T)[\chi_{T(T-3)/2}^2 - T(T-3)/2],$$

where $\chi_{T-1}^2$ and $\chi_{T(T-3)/2}^2$ are independent $\chi^2$ random variables with $T-1$ and $T(T-3)/2$ degrees of freedom, respectively, and $a(T) = 4(T+2)/[5(T-1)^2(T+1)]$ and $b(T) = 2(5T+6)/[5T(T-1)(T+1)]$. As anticipated, the limiting distribution of $R_{AVE}^2$ is distribution-free because it is based on ranks.

### 2.2. Power study of $r_n$

In this section, the results of a simulation study are presented to examine the power of $r_n$ and then the power of $r_n$ is compared to that of a rank statistic, $R_{AVE}^2$, studied in Frees (1995). $R_{AVE}^2$ is based on the Spearman correlation coefficient and is given by $R_{AVE}^2 = 2/[n(n-1)]\sum_{i<j} r_{ij}^2/[r_{ii}\ r_{jj}\ ]$, where $r_{ij} =$

$(T-1)^{-1}\sum_{t=1}^{T}(R_{it}-(T+1)/2)(R_{jt}-(T+1)/2)$ and $\{R_{i1},\ldots,R_{iT}\}$ are the ranks of $\{Y_{i1},\ldots,Y_{iT}\}$. Note that $r_{ij}/(r_{ii}\, r_{jj})^{1/2}$ is the Spearman rank correlation coefficient between the $i$th and $j$th experimental unit.

The family of alternative models used is

$$Y_{it} = \alpha_i\beta_t + e_{it}, \tag{3}$$

where the $\{\beta_t\}$ and the $\{e_{it}\}$ are i.i.d. r.v.'s with variances $\sigma_\beta^2$ and $\sigma_e^2$, $E(\beta)=1$ and $E(e)=0$. The $\{\alpha_i\}$ are not random.

Under model (3), the cross-sectional correlation equals

$$\rho_{ij} = \mathrm{Corr}\,(Y_{it},Y_{jt}) = \frac{\alpha_i\alpha_j\sigma_\beta^2}{[(\alpha_i^2\sigma_\beta^2+\sigma_e^2)(\alpha_j^2\sigma_\beta^2+\sigma_e^2)]^{1/2}}.$$

The case $\sigma_\beta^2 = 0$ gives the zero cross-sectional correlation model, namely the model under the null hypothesis. Taking $\sigma_e^2 = 1$ and the fixed parameters $\alpha_i$ to be either $-1$ or 1, gives cross-sectional correlations

$$\rho_{ij} = \frac{\alpha_i\alpha_j\sigma_\beta^2}{\sigma_\beta^2+1}.$$

Let $p$ be the proportion of ones among the values of $\alpha_i$. Three choices of $p$ were studied, namely (1) $p=1$, corresponding to all positive $\rho_{ij}$'s, (2) $p = 0.5+1/(2\sqrt{n})$, corresponding to the case that $\sum_{i<j}\rho_{ij}=0$, that is, there are as many positive correlations as negative ones and (3) $p=0.75+1/(4\sqrt{n})$, most of the correlations are positive but there are many negative ones as well.

The powers of the test statistics under the alternative family in (3) with $\beta_t \sim N(1,\sigma_\beta^2)$, $\beta_t \sim \mathrm{Exp}(1-\sigma_\beta,\sigma_\beta)$, $e_{it}\sim N(0,\sigma_e^2)$, and with the above three choices of $p$ are summarized in Tables 1-3, respectively. In each simulation, the number of replications used was 5,000 which yields a standard error about 0.0071. The sample sizes studied are 10, 20, 50 and 250; $T$ varies from 5 to 10; and the choices of $\sigma_\beta = 1.0, 0.5, 0.1$ and 0.0 correspond to $|\rho_{ij}| = 0.5, 0.2, 0.0909$ and 0.00.

In Tables 1-3, both statistics detected cross-sectional correlations whose absolute values are equal to or greater than 0.2. Power is increasing in $n$, $T$ and $|\rho_{ij}|$. When $\beta_t$ has a $N(1,\sigma_\beta^2)$ distribution, in general, $r_n$ is more powerful than $R^2_{AVE}$ which is because $r_n$ is based on the Pearson correlation coefficient whereas $R^2_{AVE}$ is based on the Spearman correlation coefficient.

When $\beta_t$ has the exponential or uniform distribution with mean one and variance $\sigma_\beta^2$, $r_n$ dominates $R^2_{AVE}$ in most cases, and in the rest of cases they are equivalent. When $\beta_t$ is exponentially distributed, both $r_n$ and $R^2_{AVE}$ exhibit slightly lower power than when $\beta_t$ is normally distributed. When $\beta_t$ is uniformly distributed, the powers of both statistics are similar to those when $\beta_t$ is normally distributed. The details of the powers of both statistics when $\beta_t$ is uniformly distributed is available from the author.

Table 1. Power of $r_n$ and $R^2_{AVE}$ with $p$, the proportion of intercepts equal to 1, equal to 1.

| $n$ | $\|\rho_{ij}\|$ | $\beta_t \sim N(1,\sigma^2_\beta)$ | | | | $\beta_t \sim$ Exp$(1-\sigma_\beta,\sigma_\beta)$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $T=5$ | | $T=10$ | | $T=5$ | | $T=10$ | |
| | | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ |
| 10 | 0.5 | .848 | .671 | .980 | .955 | .812 | .518 | .963 | .845 |
| | 0.2 | .571 | .171 | .797 | .423 | .465 | .166 | .682 | .340 |
| | .0909 | .080 | .046 | .081 | .049 | .099 | .048 | .101 | .051 |
| | 0.00 | .054 | .049 | .047 | .050 | .054 | .049 | .047 | .050 |
| 25 | 0.5 | .960 | .849 | .999 | .995 | .911 | .705 | .992 | .951 |
| | 0.2 | .861 | .366 | .980 | .737 | .660 | .300 | .849 | .570 |
| | .0909 | .126 | .051 | .145 | .052 | .105 | .049 | .111 | .055 |
| | 0.00 | .051 | .050 | .041 | .049 | .051 | .050 | .041 | .049 |
| 50 | 0.5 | .998 | .984 | 1.000 | 1.000 | .990 | .952 | 1.000 | 1.000 |
| | 0.2 | .988 | .881 | 1.000 | .996 | .927 | .805 | .991 | .974 |
| | .0909 | .346 | .436 | .545 | .517 | .516 | .437 | .601 | .488 |
| | 0.00 | .054 | .054 | .055 | .055 | .053 | .049 | .050 | .051 |
| 250 | 0.5 | .999 | .998 | 1.000 | 1.000 | .998 | .986 | 1.000 | 1.000 |
| | 0.2 | .999 | .973 | 1.000 | 1.000 | .983 | .928 | 1.000 | .998 |
| | .0909 | .849 | .538 | .983 | .711 | .650 | .526 | .762 | .646 |
| | 0.00 | .050 | .053 | .057 | .055 | .052 | .054 | .056 | .051 |

Table 2. Power of $r_n$ and $R^2_{AVE}$ with $p$, the proportion of intercepts equal to 1, equal to $0.5 + 1/(2\sqrt{n})$.

| $n$ | $p$ | $\|\rho_{ij}\|$ | $\beta_t \sim N(1,\sigma^2_\beta)$ | | | | $\beta_t \sim$ Exp$(1-\sigma_\beta,\sigma_\beta)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $T=5$ | | $T=10$ | | $T=5$ | | $T=10$ | |
| | | | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ |
| 10 | 0.658 | 0.5 | .842 | .669 | .979 | .962 | .805 | .513 | .966 | .852 |
| | | 0.2 | .574 | .180 | .802 | .417 | .473 | .169 | .690 | .354 |
| | | .0909 | .080 | .052 | .078 | .058 | .102 | .050 | .101 | .056 |
| | | 0.00 | .050 | .045 | .056 | .047 | .050 | .045 | .056 | .047 |
| 25 | 0.600 | 0.5 | .962 | .845 | .998 | .996 | .919 | .702 | .994 | .956 |
| | | 0.2 | .848 | .354 | .984 | .746 | .639 | .294 | .856 | .593 |
| | | .0909 | .114 | .050 | .138 | .054 | .108 | .050 | .110 | .056 |
| | | 0.00 | .049 | .048 | .048 | .052 | .049 | .048 | .048 | .052 |
| 50 | 0.571 | 0.5 | .996 | .984 | 1.000 | 1.000 | .992 | .951 | 1.000 | 1.000 |
| | | 0.2 | .990 | .876 | 1.000 | .996 | .938 | .803 | .992 | .970 |
| | | .0909 | .315 | .431 | .541 | .512 | .503 | .447 | .577 | .484 |
| | | 0.00 | .048 | .051 | .052 | .051 | .050 | .051 | .053 | .049 |
| 250 | 0.532 | 0.5 | 1.000 | .998 | 1.000 | 1.000 | .998 | .988 | 1.000 | 1.000 |
| | | 0.2 | 1.000 | .974 | 1.000 | .999 | .979 | .920 | 1.000 | .998 |
| | | .0909 | .849 | .544 | .981 | .714 | .640 | .522 | .764 | .644 |
| | | 0.00 | .050 | .052 | .057 | .054 | .053 | .048 | .055 | .052 |

Table 3. Power of $r_n$ and $R^2_{AVE}$ with $p$, the proportion of intercepts equal to 1, equal to $0.75+1/(4\sqrt{n})$.

| $n$ | $p$ | $\|\rho_{ij}\|$ | $\beta_t \sim N(1,\sigma^2_\beta)$ | | | | $\beta_t \sim$ Exp$(1-\sigma_\beta,\sigma_\beta)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $T=5$ | | $T=10$ | | $T=5$ | | $T=10$ | |
| | | | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ | $r_n$ | $R^2_{AVE}$ |
| 10 | 0.829 | 0.5 | .852 | .671 | .982 | .961 | .816 | .519 | .961 | .843 |
| | | 0.2 | .578 | .183 | .807 | .419 | .474 | .173 | .680 | .332 |
| | | .0909 | .083 | .052 | .086 | .051 | .103 | .051 | .098 | .047 |
| | | 0.00 | .050 | .057 | .056 | .046 | .050 | .057 | .056 | .046 |
| 25 | 0.800 | 0.5 | .961 | .860 | .999 | .994 | .914 | .698 | .994 | .953 |
| | | 0.2 | .843 | .367 | .982 | .748 | .633 | .305 | .854 | .579 |
| | | .0909 | .108 | .053 | .143 | .054 | .100 | .056 | .116 | .054 |
| | | 0.00 | .044 | .060 | .049 | .049 | .044 | .060 | .049 | .049 |
| 50 | 0.785 | 0.5 | .996 | .985 | 1.000 | 1.000 | .989 | .948 | 1.000 | .999 |
| | | 0.2 | .987 | .892 | 1.000 | .993 | .929 | .800 | .992 | .971 |
| | | .0909 | .318 | .439 | .526 | .504 | .512 | .430 | .584 | .486 |
| | | 0.00 | .043 | .051 | .050 | .051 | .051 | .046 | .053 | .052 |
| 250 | 0.766 | 0.5 | 1.000 | .997 | 1.000 | 1.000 | .998 | .989 | 1.000 | 1.000 |
| | | 0.2 | .999 | .974 | 1.000 | 1.000 | .984 | .925 | .999 | .996 |
| | | .0909 | .858 | .539 | .984 | .695 | .659 | .534 | .776 | .646 |
| | | 0.00 | .058 | .051 | .056 | .050 | .049 | .051 | .054 | .048 |

In Frees (1995), $R^2_{AVE}$ is compared to $R_{AVE}$, where $R_{AVE} = \sum_{i<j} r_{ij}/[(r_{ii}r_{jj})^{1/2}]$. Under the alternative model in (3), with $\beta_t \sim N(1,\sigma^2_\beta)$ and $e_{it} \sim N(0,\sigma^2_e)$, for the choices of $p=1$ and $p=0.75+1/(4\sqrt{n})$, $r_n$ has power similar to that of $R_{AVE}$ in most cases, but is slightly less powerful than $R_{AVE}$ in the small sample size cases ($n=10$ and 25). In situations with $p=0.5+1/(2\sqrt{n})$, $r_n$ is much more powerful than $R_{AVE}$.

The simulation studies show us when to apply each statistic. When either positive or negative correlations will prevail, one should use $R_{AVE}$ or $r_n$. When it is unknown whether $\rho_{ij}$ is positive or negative or if there is a mixture of positive and negative cross-sectional correlations, then one should use $r_n$ or $R^2_{AVE}$. A mixture of positive and negative cross-sectional correlations occurs in the study of migration between states in Frees (1992).

## 3. $r_n(\hat{\lambda})$ and Weighted Degenerate U-statistics with Estimated Parameters

The simulation study shows that the statistic $r_n$ is powerful and thus is useful in general situations. In econometrics, often $r_n$ is used for model diagnosis. Hence, in practice, $r_n$ will be calculated using residuals from a complex model. The following example illustrates this point. Assume that the panel data follow a regression model. In this model, slope coefficients are constants and the intercept varies over individuals. Let

$$Y_{it} = \mu_i + C'_{it}\beta + \sigma_i e_{it}, \text{ for } i = 1, \ldots, n \text{ and } t = 1, \ldots, T, \tag{4}$$

where $\beta$ is an unknown parameter, the $\{\sigma_i\}$ are known unequal constants, and the $\{e_{it}\}$ are i.i.d. r.v.'s with mean zero.

Let $\hat{\beta}$ be any consistent estimator of $\beta$. Define $\hat{\lambda} = \hat{\beta} - \beta$, $\lambda = 0$, the residuals $\hat{e}_{it} = Y_{it} - \hat{Y}_{it} = \sigma_i e_{it} + \mu_i - \hat{\mu}_i - C_{it}\hat{\lambda}$, and

$$\tilde{e}_{it} = \sigma_i^{-1}(\hat{e}_{it} - \bar{\hat{e}}_i) = e_{it} - \bar{e}_i - \sigma_i^{-1}(C_{it} - \bar{C}_i)\hat{\lambda} \ .$$

Substituting $\{Y_{it}\}$ of $r_n$ in (1) with "residuals" $\{\sigma_i \tilde{e}_{it}\}$, we obtain

$$r_n(\hat{\lambda}) = \frac{\sum_{i \neq j} d_{ijn}[T^{-1}\sum_{t=1}^T \tilde{e}_{it}^2 \tilde{e}_{jt}^2 - (T^{-1}\sum_{t=1}^T \tilde{e}_{it}^2)(T^{-1}\sum_{t=1}^T \tilde{e}_{jt}^2)]}{\sum_{i \neq j} d_{ijn}(T^{-1}\sum_{t=1}^T \tilde{e}_{it}^2)(T^{-1}\sum_{t=1}^T \tilde{e}_{jt}^2)} \equiv \frac{A_n(\hat{\lambda})}{B_n(\hat{\lambda})}, \tag{5}$$

where $d_{ijn} = \sigma_i^2 \sigma_j^2/(n^2)$.

It turns out that the numerator of $r_n(\hat{\lambda})$, $A_n(\hat{\lambda})$, is a weighted degenerate U-statistic with estimated parameters ($WU_n(\hat{\lambda})$ defined in (6) below). (For details, see Section 3.2.1.) The limiting distribution of $r_n(\hat{\lambda})$ depends on $WU_n(\hat{\lambda})$. This led the author to investigate the limiting distribution of $WU_n(\hat{\lambda})$.

In Section 3.1, the limiting distribution of weighted degenerate V-statistics with estimated parameters ($WV_n(\hat{\lambda})$) is established. The limiting distribution of $WU_n(\hat{\lambda})$ is obtained via $WV_n(\hat{\lambda})$ and turns out to be a weighted sum of chi-square variates. Conditions under which both $WU_n(\hat{\lambda})$ and $WU_n(\lambda)$ assume the same limiting distribution (referred to as Case I) and different limiting distributions (referred to as Case II) are specified.

### 3.1. The limiting distributions of $WU_n(\hat{\lambda})$ and $WV_n(\hat{\lambda})$

Let $X_1, X_2, \ldots$ denote i.i.d. r.v.'s. Assume that $E(X_1) = 0, \text{Var}\,(X_1) = 1$, and $E(X_1^4) < \infty$. Let the kernel $h(\cdot)$ be a symmetric real valued function with finite second moment. Weighted degenerate U-statistics with estimated parameters are defined as follows:

$$WU_n(\hat{\lambda}) = \sum_{i \neq j} d_{ijn} h(X_i, X_j; \hat{\lambda}). \tag{6}$$

Here $\{d_{ijn}\}$ are symmetric but non-stochastic weights, $\hat{\lambda} = \hat{\lambda}(X_1, ..., X_n)$ is a consistent estimator of the $p$-vector parameter $\lambda$, and $h$ is degenerate in the sense that $\text{Var}\,[h_1(X_1)] = 0$, where $h_1(x_1) = E[h(x_1, X_2; \lambda)]$. If $\lambda$ is known, $WU_n(\lambda)$ is a weighted degenerate U-statistic. Whether and how the variability in $\hat{\lambda}$ affects the limiting distribution of $WU_n(\hat{\lambda})$ is an important question.

For investigations on the asymptotics of statistics with estimated parameters, see De Wet and Randles (1987) for a thorough review. This paper investigates how the estimated parameter $\hat{\lambda}$ affects the limiting distributions of $WU_n(\hat{\lambda})$ and $WV_n(\hat{\lambda})$. The result is based on the asymptotic results for degenerate U- and V-statistics with estimated parameters in DeWet and Randles (1987) and weighted degenerate U-statistics ($WU_n(\lambda)$ henceforth) in Shieh, Johnson and Frees (1994).

A class of statistics closely related to $WU_n(\lambda)$ is the class of weighted degenerate V-statistics which assumes the following form.

$$WV_n(\lambda) = \sum_{i,j} d_{ijn}\ h(X_i, X_j; \lambda),$$

where $\sum_{i,j}$ denotes the summation in which both indices $i$ and $j$ run from 1 to $n$. To avoid the strong assumption that the kernel $h$ is differentiable in $\hat{\lambda}$, we assume that

$$h(x, y; \lambda) = \int_{-\infty}^{\infty} g(x, t; \lambda) g(y, t; \lambda) dM(t), \tag{7}$$

where $g$ is some real-valued function and $M(t)$ is a finite positive measure.

Conditions 1 and 3 below are basic regularity conditions on the function $g$ and hence are conditions indirectly on the kernel $h$. Condition 2 is the usual asymptotic linearity applied to $\hat{\lambda}$. Conditions 4, 5 and W are necessary for the limiting distributions of $WV_n(\hat{\lambda})$ and $WU_n(\hat{\lambda})$. We assume that $\mu(t;\lambda)$, where $\mu(t;\gamma) = E_\lambda[g(X_1, t; \gamma)]$, has an $L_2(R, M)$ differential at $\lambda$ as follows. Suppose that for any $\epsilon > 0$, there is a bounded sphere $\mathcal{C}$ in $R^p$ centered at $\lambda$ such that $\gamma \in \mathcal{C}$ implies

$$\|\gamma - \lambda\|^{-2} \int_{-\infty}^{\infty} [\mu(t;\gamma) - \mathbf{d}_1\mu(t;\lambda)'(\gamma - \lambda)]^2\ dM(t)\ <\ \epsilon.$$

**Condition 1.** suppose that $\mu(t;\gamma)$ exists and $\mu(t;\gamma) \equiv 0$ for every $t$ and $\gamma$ in a neighborhood of $\gamma = \lambda$. Furthermore, assume that $\mu(t, \gamma)$ has an $L_2(R, M)$ differential at $\gamma = \lambda$ with partial derivative vector $\mathbf{d}_1\mu(t;\lambda)$ and $\int_{-\infty}^{\infty}[\mathbf{d}_1\mu(t;\lambda)_r]^2 dM(t) < \infty$, for $r = 1, \ldots, p$, where $\mathbf{d}_1\mu(t;\lambda)_r$ is the $r$th component of the vector $\mathbf{d}_1\mu(t;\lambda)$.

**Condition 2.** $\hat{\lambda} = \lambda + n^{-1}\sum_{i=1}^n \alpha(X_i) + o_p(n^{-1/2})$, where $E[\alpha(X_i)_r] = 0$ and $E[\alpha(X_i)_r\ \alpha(X_i)_{r'}] < \infty$, for all $1 \le r \le r' \le p$.

**Condition 3.** Suppose that there exists a number $K_1 > 0$ and a neighborhood $K(\lambda)$ of $\lambda$ such that (a) if $\gamma \in K(\lambda)$ and $D(\gamma, d)$ is a sphere centered at $\gamma$ with radius $d$ such that $D(\gamma, d) \subset K(\lambda)$, then

$$\int_{-\infty}^{\infty} \Big\{E\Big[\sup_{r' \in D(\gamma,d)} |g(X_1, t; \gamma') - g(X_1, t; \gamma)|\Big]\Big\}^2 dM(t) \le K_1\ d^2$$

and (b) for any $\epsilon > 0$, there is a $d^* > 0$ such that $0 < d < d^*, \gamma \in K(\lambda)$ and $D(\gamma, d) \subset K(\lambda)$ imply

$$\int_{-\infty}^{\infty} E\Big[\sup_{\gamma' \in D(\gamma,d)} |g(X_1, t; \gamma') - g(X_1, t; \gamma)|^4\Big] dM(t) < \epsilon.$$

**Condition 4.** Suppose that $g(X_1, t; \lambda) \in D[-\infty, \infty]$ and for fixed $t_0$, $P\{g(X_1, t_0) \neq g(X_1-, t_0)\} = 0$. Further, Let $K_1(s,t) = E[g(X_1, s; \lambda) g(X_1, t; \lambda)]$. $K_1(s,s) + K_1(t,t) - 2K_1(s,t) \leq |t-s|^{\alpha}$, where $\alpha > 1$.

**Remark.** Since $K_1$ is a covariance matrix, in general it is smooth and thus Condition 4 is not restrictive. When $K_1$ has finite discontinuity points, we can choose $s$ and $t$ from a small interval such that Condition 4 holds.

**Remark.** When $K_1$ is smooth, we can take $\alpha = 2$.

Let $\{\delta_k\}$ be the eigenvalues of $h(x, y; \lambda)$ corresponding to the eigenfunctions $\{\phi_k(\cdot)\}$, and

$$\lim_{K \to \infty} E\left\{\left[h(x,y;\lambda) - \sum_{k=1}^{K} \delta_k \phi_k(x)\phi_k(y)\right]^2\right\} = 0. \tag{8}$$

**Condition 5.** $E[h^2(X_1, X_i)] < \infty$, for $i = 1$ or $2$. $\sum_{k=1}^{\infty} |\delta_k| < \infty$ and $h(X, X, \lambda) = \sum_{k=1}^{\infty} \delta_k \phi_k^2(X)$.

The following notation is needed for Condition W. Let $\{b_{imn} : i = 1, \ldots, n; m = 1, 2, \ldots\}$ and $\{\eta_m : m = 1, 2, \ldots\}$ be real numbers, and $\delta_{km} = 1$, if $k = m$, and 0 otherwise. Let $\eta_m$ be limiting eigenvalues of the weight matrix, $\mathbf{D_n} = (nd_{ijn})$, in the following sense. Since $\mathbf{D_n}$ is symmetric, there exists an orthogonal matrix $\mathbf{B_n} = (b_{imn})$ such that $\mathbf{B_n}'\mathbf{D_n}\mathbf{B_n} = \mathbf{\Lambda_n}$. Let $\eta_{mn}$ be the $m$th diagonal element of $\mathbf{\Lambda_n}$ and $\lim_{n\to\infty} \eta_{mn} = \eta_m$. (For details, see Shieh, Johnson and Frees (1994).)

**Condition W.** (i) $\max_{1 \leq i \leq n} |b_{imn}| \to 0$ as $n \to \infty$ for each $m$,
(ii) $\sum_{i=1}^{n} \sum_{j=1}^{n} (nd_{ijn})^2 \to \sum_{m=1}^{\infty} \eta_m^2 < \infty$,
(iii) $\sum_{m=1}^{\infty} |\eta_m| < \infty$ and $\sum_{m=1}^{\infty} |\eta_{mn} - \eta_m| \to 0$, as $n \to \infty$,
(iv) $|b_{imn}| \leq b/\sqrt{n} < \infty$.
(v) $(1/\sqrt{n}) \sum_{i=1}^{n} b_{imn} \to c_m$, where $\{c_m\}$ are constants.

By (7), we can express $WV_n(\hat{\lambda})$ as $WV_n(\hat{\lambda}) = \sum_{i,j} d_{ijn} \int_{-\infty}^{\infty} g(X_i, t; \hat{\lambda}) g(X_j, t; \hat{\lambda}) dM(t)$. By spectrum analysis in linear algebra, $(nd_{ijn})$ is symmetric, thus we can expand $nd_{ijn}$ into $\sum_{m=1}^{n} \eta_{mn} b_{imn} b_{jmn}$. Thus, $WV_n(\hat{\lambda}) = n^{-1} \sum_{m=1}^{n} \eta_{mn} \int_{-\infty}^{\infty} [\sum_{i=1}^{n} b_{imn} g(X_i, t; \hat{\lambda})]^2 \, dM(t)$. If $b_{imn} \equiv 1/n$, then the term in square brackets (for each fixed $t$) is a degree-1 V-statistic with estimated parameters, and will be denoted by $V_{1n}(\hat{\lambda})$. Let $\hat{\lambda} = \lambda(F_n)$, a functional of the empirical d.f. $F_n$. Similar to the approach in De Wet and Randles (1987) (a standard technique applied to V-statistics), with $b_{imn} \equiv 1/n$ and $t$ fixed, we write $V_{1n}(\hat{\lambda})$ as a functional $T(F_n) = \int_{-\infty}^{\infty} g(x; \lambda(F_n)) dF_n(x)$. The first *Gâteaux* differential of $T(\cdot)$ at $F$ in the direction of $F_n - F$ is

$$d_1T(F; F_n - F) = \int_{-\infty}^{\infty} g(x; \lambda) d(F_n - F) + \mathbf{d_1}\theta(\lambda)' d_1\lambda(F; F_n - F), \quad (9)$$

where $\mathbf{d_1}\theta(\lambda)$ denotes the partial derivative of $\theta(\gamma) = E_\lambda[g(X_1; \gamma)]$ with respect to $\gamma$, and the expectation assumes $\lambda$ the actual parameter value. Equation (9) motivates us to approximate $g(X_i, t; \hat{\lambda})$ by

$$g(X_i, t; \lambda) + \mathbf{d_1}\theta(\lambda)'(\hat{\lambda} - \lambda). \quad (10)$$

Note that through this approximation, we do not need to assume that $g(\cdot)$ is differentiable in $\gamma$ at $\gamma = \lambda$, which is often not true, while $\theta(\cdot)$ is differentiable in $\gamma$. Furthermore, if $\theta(\cdot)$ has a zero derivative at $\gamma = \lambda$, then the limiting distribution of $V_{1n}(\lambda)$ is not affected if $\lambda$ needs to be estimated.

By (10) and Condition 2 we define $WV_n$, an approximation of the weighted degenerate V-statistic with estimated parameters, as follows:

$$WV_n = \sum_{m=1}^{n} \eta_{mn} \int_{-\infty}^{\infty} \Big\{ \sum_{i=1}^{n} b_{imn} [g(X_i, t; \lambda) + \mathbf{d_1}\mu(t, \lambda)' \overline{\alpha(X)}] \Big\}^2 dM(t),$$

where $\mu(t; \gamma) = E_\lambda[g(X_1, t; \gamma)]$ and $\overline{\alpha(X)} = n^{-1} \sum_{k=1}^{n} \alpha(X_k)$. We note that the effect of $\hat{\lambda}$ is captured in $\mathbf{d_1}\mu(t; \lambda)$. The following lemmas are needed in Theorem 1, which establishes the limiting distribution of $WV_n(\hat{\lambda})$.

**Lemma 1.** (De Wet and Randles (1987), Lemma 4.1.) *Let $X_1, \ldots, X_n$ be i.i.d. and suppose $k_n(x, y) = k_n(y, x)$ for every $x$, $y$ and $n$. In addition, assume for every $x$ and $n$ that $E[k_n(x, X_2)] = 0$ and $E[k_n^2(X_1, X_i)] = o(n^2)$ for $i = 1$ and 2, as $n \to \infty$, then $W_n = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} k_n(X_i, X_j) \to_P 0$.*

Define

$$Q_n(\mathbf{s}) = \int_{-\infty}^{\infty} \Big\{ \sum_{i=1}^{n} b_{imn} \Big[ g(X_i, t; \lambda + n^{-1/2}\mathbf{s}) - g(X_i, t; \lambda) - \mu(t; \lambda + n^{-1/2}\mathbf{s}) \Big] \Big\}^2 dM(t).$$

Note that for simplicity we have supressed the index $m$ in $Q_n$.

**Lemma 2.** *Suppose that Conditions* 1, 2 *and W hold. Then* $Q_n(\sqrt{n}(\hat{\lambda} - \lambda)) \to_P 0$.

**Proof.** See Appendix 1 of Shieh (1994).

The following theorem shows that when Conditions 1–4 and W hold, both $WV_n(\hat{\lambda})$ and $WV_n$ have the same limiting distribution, which is a weighted sum of chi-square variates. When $\mathbf{d}_1\mu(t; \lambda) \equiv \mathbf{0}$ (case I situation) the estimator $\hat{\lambda}$ does not affect the limiting distribution of $WV_n(\hat{\lambda})$, and the limiting distribution is a weighted sum of independent chi-square one variates. For the case II situation, $\mathbf{d}_1\mu(t;\lambda) \neq \mathbf{0}$, so the effect of $\hat{\lambda}$ is captured in $\mathbf{d}_1\mu(t;\lambda)$.

**Theorem 1.** *Suppose that Conditions* 1-4, *and W hold. Then* $n[WV_n(\hat{\lambda}) - WV_n] \to_P 0$ *and*

$$nWV_n(\hat{\lambda}) \to_D \sum_{m=1}^{\infty} \eta_m \int_{-\infty}^{\infty} [Z_m(t) + c_m \mathbf{d}_1\mu(t;\lambda)'\mathbf{Z})]^2 dM(t),$$

*where the* $\{\eta_m\}$ *are the limiting eigenvalues of the weight matrix* $(nd_{ijn})$ *defined above Condition W. Further, for any finite constant* $M$*, the* $\{Z_m(t), m = 1, \ldots, M\}$ *are i.i.d. Gaussian processes with covariance matrix function* $K_1(u,v) = E[g(X_1,u,\lambda)\ g(X_1,v,\lambda)]$*, and* $\mathbf{Z} \sim N_p(\mathbf{0},\mathbf{\Sigma})$ *with* $\mathbf{\Sigma} = E[\alpha(X_1)\alpha(X_1)']$*. The covariance matrix function of* $\{Z_m(t), m = 1, \ldots, M\}$ *and* $\mathbf{Z}$ *equals* $K_2(u) = \{c_1 E[g(X_1,\ u,\lambda)\alpha(X_1)], \ldots, c_M\ E[g(X_1,u,\lambda)\alpha(X_1)]\}$.

**Proof.** An intermediate term between $WV_n$ and $WV_n(\hat{\lambda})$ is

$$Y_n(\hat{\lambda}) = n^{-1} \sum_{m=1}^{n} \eta_{mn} \int_{-\infty}^{\infty} \Big\{ \sum_{i=1}^{n} b_{imn} \Big[ g(X_i,t;\lambda) + \mathbf{d_1}\mu(t;\lambda)(\hat{\lambda} - \lambda) \Big] \Big\}^2 dM(t),$$

where the $(b_{imn})$ are above Condition W. We first show that

$$n[WV_n(\hat{\lambda}) - Y_n(\hat{\lambda})] \to_P 0. \tag{11}$$

$$n[WV_n(\hat{\lambda}) - Y_n(\hat{\lambda})] = \sum_{m=1}^{n} \eta_{mn} [WV_{mn}(\hat{\lambda}) - Y_{mn}(\hat{\lambda})],$$

where $WV_{mn}(\hat{\lambda}) = \int_{-\infty}^{\infty} [\sum_{i=1}^{n} b_{imn} g(X_i,t;\hat{\lambda})]^2 dM(t)$ and

$$Y_{mn}(\hat{\lambda}) = \int_{-\infty}^{\infty} \Big\{ \sum_{i=1}^{n} b_{imn} [g(X_i,t,\lambda) + \mathbf{d}_1\mu(t;\lambda)'(\hat{\lambda} - \lambda)] \Big\}^2 dM(t).$$

To prove (11), it suffices to show that $\sum_{m=1}^{n} \eta_{mn}^{(s)} [WV_{mn}(\hat{\lambda}) - Y_{mn}(\hat{\lambda})] \to_P 0$. Let $s$ be "+" or "−" and let $\eta_{mn}^{(+)}$ and $\eta_{mn}^{(-)}$ be the positive part and the negative part of $\eta_{mn}$, respectively. Treating $[\sum_{m=1}^{n} \eta_{mn}^{(s)} \int_{-\infty}^{\infty} (\cdot)^2 dM(t)]^{1/2}$ as a norm, it suffices to show that

$$\Big( \sum_{m=1}^{n} \eta_{mn}^{(s)} WV_{mn}(\hat{\lambda}) \Big)^{1/2} - \Big( \sum_{m=1}^{n} \eta_{mn}^{(s)}\ Y_{mn}(\hat{\lambda}) \Big)^{1/2} \to_P 0, \tag{12}$$

since $nY_n(\hat{\lambda})$ will be shown to be bounded later. To prove (12), we need to show that

$$\Big( \int_{-\infty}^{\infty} \sum_{m=1}^{n} \eta_{mn}^{(s)} \Big\{ \sum_{i=1}^{n} b_{imn} [g(X_i,t;\hat{\lambda}) - g(X_i,t,\lambda) - \mathbf{d_1}\mu(t;\lambda)'(\hat{\lambda} - \lambda)] \Big\}^2 dM(t) \Big)^{1/2} \to 0. \tag{13}$$

Now

$$\Big(\int_{-\infty}^{\infty}\Big\{\sum_{i=1}^{n} b_{imn}[g(X_i,t;\hat{\lambda})-g(X_i,t,\lambda)-\mathbf{d}_1\mu(t;\lambda)'(\hat{\lambda}-\lambda)]\Big\}^2 dM(t)\Big)^{1/2}$$
$$\leq \Big(\int_{-\infty}^{\infty}\Big\{\sum_{i=1}^{n} b_{imn}[g(X_i,t;\hat{\lambda})-g(X_i,t;\lambda)-\mu(t;\hat{\lambda})]\Big\}^2 dM(t)\Big)^{1/2}$$
$$+\Big(\int_{-\infty}^{\infty}\Big\{\sum_{i=1}^{n} b_{imn}[\mu(t;\hat{\lambda})-\mathbf{d}_1\mu(t;\lambda)'(\hat{\lambda}-\lambda)]\Big\}^2 dM(t)\Big)^{1/2} \equiv S_{1mn}+S_{2mn},$$

the inequality holds by triangle inequality. In Appendix 1 of Shieh (1994), we show that $Q_n(\sqrt{n}(\hat{\lambda}-\lambda)) \to_P 0$ by Lemma 1. Thus $0 \leq S_{1mn} = [Q_n(\sqrt{n}(\hat{\lambda}-\lambda))]^{1/2} \to_P 0$. Let $M_0 = \int_{-\infty}^{\infty} dM(T) < \infty$. $S_{2mn}^2 \leq M_0(\sum_{i=1}^{n} b_{imn})^2[\mu(t,\hat{\lambda})$ $\mathbf{d}_1\mu(t;\lambda)'(\hat{\lambda}–\lambda)]^2 \to_P 0$, provided that Conditions 2 and W hold. Thus $S_{2mn} \to_P 0$, since $0 \leq S_{2mn}$. Now since $\eta_{mn}^{(s)} \geq 0$, $\eta_{mn} \to \eta_m$ and Condition W(iii) holds, this is sufficient for (13). In Appendix 2 of Shieh (1994), we show that $n[Y_n(\hat{\lambda}) - WV_n] \to_P 0$. Thus $n[WV_n(\hat{\lambda}) - WV_n] \to_P 0$.

In the following, we proceed to derive the limiting distribution of $nWV_n$.

$$nWV_n = \sum_{m=1}^{n}\eta_{mn}\int_{-\infty}^{\infty}\Big\{\sum_{i=1}^{n} b_{imn} g(X_i,t;\lambda) + n^{-\frac{1}{2}}\sum_{i=1}^{n} b_{imn}\mathbf{d}_1\mu(t;\lambda)'\sqrt{n}\ \overline{\alpha(X)}\Big\}^2 dM(t)$$
$$\equiv \sum_{m=1}^{n}\eta_{mn}\int_{-\infty}^{\infty}\Big\{Z_{mn}(t,\lambda)+c_{mn}\mathbf{d_1}\mu(t;\lambda)'\mathbf{Z_n}\Big\}^2 dM(t).$$

First, for any finite $M$ we have

$$\{Z_{mn}(t), m=1,\ldots,M\} \to_D \{Z_m(t), m=1,\ldots,M\} \quad \text{and} \quad \mathbf{Z_n} \to_D \mathbf{Z},$$

where $\{Z_m(t), m = 1,\ldots,M\}$ are i.i.d. Gaussian processes with mean 0 and covariance matrix function $K_1(u,v) = E[g(X_1,u;\lambda)g(X_1,v;\lambda)]$, and the covariance matrix function of $\{Z_m(t), m = 1,\ldots,M\}$ and $\mathbf{Z}$ equals $K_2(u) = \{c_1 E[g(X_1,u,\lambda)\alpha(X_1)],\ldots,c_M E[g(X_1,u,\lambda)\alpha(X_1)]\}$. Further, $\mathbf{Z} \sim \mathbf{N_p}(\mathbf{0},\mathbf{\Sigma})$, where $\mathbf{\Sigma} = E[\alpha(X_1)\alpha(X_1)']$.

In Appendix 3 of Shieh (1994), we show that $\{Z_{mn}(t), m = 1,\ldots,M\}$ and $\mathbf{Z_n}$ satisfy conditions in Theorem 15.6 of Billingsley (1968); thus $\{Z_{mn}, m = 1,\ldots,M\}$ and $\mathbf{Z_n}$ jointly converge to a sequence of Gaussian processes and a random vector $\mathbf{Z}$. From the strong representation theorem in Skorohod (1956), we can find an appropriate probability space in which to define a probability such that $\{Z_{mn}(t), m = 1,\ldots,M\}$ almost surely by converge to $\{Z_m(t), m = 1,\ldots,M\}$ in the Skorohod topology. Since the limiting process is continuous, the convergence is also uniform with probability 1. Hence we can apply the Dominated convergence theorem to $nWV_n$ to obtain

$$\begin{aligned} nWV_n &= \sum_{m=1}^{n} \eta_{mn} \int_{-\infty}^{\infty} \Big\{ Z_{mn}(t) + c_{mn}\mathbf{d_1}\mu(t;\lambda)'\mathbf{Z_n} \Big\}^2 dM(t) \\ &\to_D \sum_{m=1}^{\infty} \eta_m \int_{-\infty}^{\infty} \Big\{ Z_m(t) + c_m\mathbf{d_1}\mu(t;\lambda)'\mathbf{Z} \Big\}^2 dM(t), \end{aligned}$$

where $c_m = \lim_{n\to\infty} c_{mn}$.

We note that the limiting distribution is a weighted sum of chi-square variates. If the weight matrix $(d_{ijn})$ has a special structure, the limiting distribution of $nWV_n(\hat{\lambda})$ in Theorem 1 can be simplified.

**Corollary 1.** *Suppose that $\sum_{i=1}^{n} d_{ijn} = C$, where $C$ is a constant, then $c_m = 0$, for $m = 2, \ldots$, and $b_{i1n} \equiv (1/\sqrt{n})$, $i = 1, \ldots, n$. Thus $nWV_n(\hat{\lambda})$ reduces to*

$$\eta_1 \sum_{k=1}^{\infty} \delta_k^* \chi_{1k}^2 + \sum_{m=2}^{\infty} \eta_m \sum_{k=1}^{\infty} \delta_k^{*1} \chi_{mk}^2,$$

*where $\{\delta_k^*\}$ and $\{\delta_k^{*1}\}$ are eigenvalues of $h_*$ and $h_{*1}$ respectively defined below, and $\{\chi_{mk}^2, m = 1, \ldots\}$ are independent chi-square one variates.*

Now $nWV_n = \eta_{1n} \sum_{i,j} h_*(X_i, X_j) + \sum_{m=2}^{n} \eta_{mn} \sum_{i,j} h_{*1}(X_i, X_j) dM(t)$, where

$$\begin{aligned} h_*(X_i, X_j) = (1/n) \int_{-\infty}^{\infty} \{ \, [\, g(X_i, t; \lambda) + (1/n)\mathbf{d}_1\mu(t;\lambda)'\alpha(X_i)] \times \\ [\, g(X_j, t; \lambda) + (1/n)\mathbf{d}_1\mu(t;\lambda)'\alpha(X_j)]\} dM(t) \end{aligned}$$

and $h_{*1}(X_i, X_j) = \int_{-\infty}^{\infty} \{ b_{imn} b_{jmn} [g(X_i, t; \lambda)][g(X_j, t; \lambda)]\} dM(t)$. Following the methods in De Wet and Randles (1987), we can obtain $\{\delta_k^*\}$ and $\{\delta_k^{*1}\}$.

When $\mathbf{d}_1\mu(t;\lambda) \equiv 0$, the limiting distribution of $nWV_n$ is not affected by $\hat{\lambda}$, and

$$nWV_n \to_D \sum_{m=1}^{\infty} \eta_m \sum_{k=1}^{\infty} \delta_k \ \chi_{1k}^2,$$

where the $\{\chi_{1k}^2\}$ are independent chi-square one variates.

The following theorem establishes the limiting distribution of weighted degenerate U-statistics with estimated parameters.

**Theorem 2.** *Suppose that Conditions 1−5, W and $E[\int_{-\infty}^{\infty} g^2(X_1, t; \lambda)\, dM(t)] < \infty$ hold. Then*

$$nWU_n(\hat{\lambda}) \to_D \sum_{m=1}^{\infty} \eta_m \int_{-\infty}^{\infty} [Z_m(t) + c_m \mathbf{d}_1\mu(t;\lambda)'\mathbf{Z})]^2 dM(t) - \sum_{m=1}^{\infty} \eta_m \sum_{k=1}^{\infty} \delta_k,$$

*where the* $\{\eta_m\}$, *the* $\{Z_m(t), m = 1, \ldots, M\}$, $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, *and the covariance matrix function of* $\{Z_m(t), m = 1, \ldots, M\}$ *and* $\mathbf{Z}$ *are defined below Theorem* 1.

**Proof.** Note that $WU_n(\hat{\lambda}) = \sum_{i \neq j} d_{ijn} \int_{-\infty}^{\infty} g(X_i, t; \hat{\lambda}) g(X_j, t; \hat{\lambda}) dM(t)$ and thus $nWU_n(\hat{\lambda}) = n[WV_n(\hat{\lambda}) - WU_{2n}(\hat{\lambda})]$, where

$$WU_{2n}(\hat{\lambda}) = \sum_{i=1}^{n} d_{iin} \int_{-\infty}^{\infty} g^2(X_i, t; \hat{\lambda}) dM(t). \tag{14}$$

Recall that $n[\ WV_n(\hat{\lambda}) - WV_n\ ] \to_P 0$ from Theorem 1. In Appendix 4 of Shieh (1994), we show that $nWU_{2n}(\hat{\lambda}) \to_P \sum_m^{\infty} \eta_m \sum_{k=1}^{\infty} \delta_k$, where $\{\delta_k\}$ are eigenvalues of $h(x, y; \lambda)$ defined in (8). This completes the proof of the theorem.

Theorem 2 states that the limiting distribution of $nWU_n(\hat{\lambda})$ is a weighted sum of chi-square variates. For case I situations, $\mathbf{d_1}\mu(t; \lambda) = \mathbf{0}$, and hence $WU_n(\hat{\lambda})$ and $WU_n$ have the same limiting distribution, which is a weighted sum of independent chi-square one variates. For case II situations $\mathbf{d_1}\mu(t; \lambda) \neq \mathbf{0}$, and the effect of $\hat{\lambda}$ is captured in $\mathbf{d_1}\mu(t; \lambda)$.

### 3.2. Applications

In this section, applications of the asymptotic result of $WU_n(\hat{\lambda})$ and $WV_n(\hat{\lambda})$ to two statistics are presented. The first example considers the test statistic, $r_n(\hat{\lambda})$, for cross-sectional correlation in panel data. The test statistic falls into Case I in which the use of estimates of parameters does not affect the limiting distribution of $WU_n(\hat{\lambda})$. The second example is a $\chi^2$ goodness-of-fit test statistic which falls in Case II, i.e., the limiting distribution of $WV_n(\hat{\lambda})$ is affected by the estimated parameters.

#### 3.2.1. Testing for cross-sectional correlation in panel data under a linear regression model

We assume that the panel data follow a regression model in (4) and define the cross-sectional correlation between the $i$th and the $j$th experimental units as $\rho_{ij} = \mathrm{Corr}(Y_{it}, Y_{jt})$. To determine whether there is cross-sectional correlation, we test the null hypothesis, $H_o : \rho_{ij} = 0$ for each $i, j$. Consider the case when the number of experimental units, $n$, is large compared to the number of observations per experimental unit $T$.

The model used here is one of the two most widely used when analyzing panel data (Hsiao (1986) and Greene (1993)). Recall that in the panel data example introduced in Section 3, we have derived $r_n(\hat{\lambda})$, the test statistic based on residuals, for testing cross-sectional correlation in panel data. In (5) we have that

$$r_n(\hat{\lambda}) = \frac{\sum_{i \neq j} d_{ijn} [T^{-1} \sum_{t=1}^{T} \tilde{e}_{it}^2 \tilde{e}_{jt}^2 - (T^{-1} \sum_{t=1}^{T} \tilde{e}_{it}^2)(T^{-1} \sum_{t=1}^{T} \tilde{e}_{jt}^2)]}{\sum_{i \neq j} d_{ijn} (T^{-1} \sum_{t=1}^{T} \tilde{e}_{it}^2)(T^{-1} \sum_{t=1}^{T} \tilde{e}_{jt}^2)} \equiv \frac{A_n(\hat{\lambda})}{B_n(\hat{\lambda})},$$

where $d_{ijn} = \sigma_i^2\sigma_j^2/(n^2)$.

Note that $n[r_n(\hat\lambda) - A_n(\hat\lambda)/E(B_n(\lambda))] \to_D 0$, provided that

$$B_n(\hat\lambda) - E(B_n(\lambda)) \to_P 0 \text{ and } E(B_n(\lambda)) \to \ C_B, \tag{15}$$

where $C_B$ is a positive constant. Note that (15) can be shown by techniques similar to Iverson and Randles (1989). Hence the limiting distribution of $r_n(\hat\lambda)$ depends on $A_n(\hat\lambda)$ which is a $WU_n(\hat\lambda)$.

In the following, we apply Theorem 2 to obtain the limiting distribution of $nr_n(\hat\lambda)$. Let $\tilde e_{i\cdot}^2 = T^{-1}\sum_{t=1}^T \tilde e_{it}^2$, for $1 \le i \le n$. Then we can write $A_n(\hat\lambda)$ as

$$A_n(\hat\lambda) = \sum_{i\neq j} d_{ijn} h(Z_i, Z_j; \hat\lambda),$$

where $h(z_1, z_2; \gamma) = \int_{-\infty}^{\infty} g(z_1, t; \gamma)g(z_2, t; \gamma)\ dM(t)$, $g(z_1, t; \gamma) = \tilde e_{1t}^2 - \tilde e_{1\cdot}^2$, $\tilde e_{1t} = e_{1t} - \bar e_1 + \sigma_1^{-1}(C_{1t} - \bar C_1)\gamma$, and $M(t)$ is the finite measure that places mass $T^{-1}$ on the integers $1, \ldots, T$. It can be checked that the kernel of $A_n(\hat\lambda)$ is degenerate and hence $A_n(\hat\lambda)$ is a $WU_n(\hat\lambda)$. It is easily seen that

$$\begin{aligned}\mu(t;\gamma) &= E[g(Z_1, t; \gamma)] \\ &= E[(e_{1t} - \bar e_1)^2 - E(e_{1t} - \bar e_1)^2] + (\gamma/\sigma_1)^2[(C_{1t}^2 - 2C_{1t}\bar C_1 - \overline{C_{1\cdot}^2} + 2\bar C_1^2] \\ &= (\gamma/\sigma_1)^2[(C_{1t}^2 - 2C_{1t}\bar C_1 - \overline{C_{1\cdot}^2} + 2\bar C_1^2],\end{aligned}$$

since the $\{e_{it}\}$ are i.i.d. with mean zero. This clearly shows that $\mu(t; \mathbf{0}) = \mathbf{0}$, for all $t$ and $\mathbf{d_1}\mu(t;\gamma)|_{\gamma=\mathbf{0}} \equiv \mathbf{0}$. Hence the limiting distribution of $nA_n(\hat\lambda)$ is not affected by the estimated parameter $\hat\lambda$, and is a weighted sum of independent chi-square one variates.

If $\sigma_i^2 \equiv \sigma^2$, then

$$\mathbf{D_n} = (nd_{ijn}) = \frac{\sigma^4}{n}\begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots \\ \vdots & & \ddots & \\ 1 \cdots & 1 & 0 & \end{pmatrix}.$$

Thus, the limiting eigenvalues of the weight matrix $(\mathbf{D_n})$ are $\eta_1 = \sigma^4$ and $\eta_i = 0$, for $i = 2, 3, \ldots$ By Theorem 2 and straightforward algebra, it can be shown that the limiting distribution of the test statistic $nr_n(\hat\lambda)$ is

$$[\sigma^4(T-1)(\mu_4 - \mu_2^2)/(T^2C_B)]\ (\chi_T^2 - T),$$

where $C_B$ is defined in (15), $\mu_4 = E(Y_{11}^4)$, $\mu_2 = E(Y_{11}^2)$ and $\chi_T^2$ is a chi square random variate with $T$ degree of freedom.

#### 3.2.2. $\chi^2$ goodness-of-fit test statistic (DeWet and Randles (1987), Example 3.3)

Let $X_1, \ldots, X_n$ be i.i.d. $F((x-\mu)/\sigma)$, and suppose that we wish to test $H_o: F = F_o$, where $F_o$ is completely specified. Let $\hat{\lambda} = (\hat{\mu}, \hat{\sigma})$ be a consistent estimator for $\lambda = (\mu, \sigma)$ satisfying Condition 2, and let $F_n(t)$ be the empirical distribution function of $X_1, \ldots, X_n$. Define $p_j = F_o(b_j) - F_o(b_{j-1})$ and $\hat{p}_j = F_n(\hat{\mu} + b_j\hat{\sigma}) - F_n(\hat{\mu} + b_{j-1}\hat{\sigma})$ for $j = 1, \ldots, k$, where $-\infty = b_0 < b_1 < \cdots < b_k = \infty$. Form the $\chi^2$ goodness-of-fit test statistic

$$\begin{aligned} WV_n(\hat{\lambda}) &= n \sum_{j=1}^{n} p_j^{-1} \, (\hat{p}_j - p_j)^2 = \sum_{i,j=1}^{n} d_{ijn} \, h(X_i, X_j; \hat{\lambda}) \\ &= n^{-1} \int_{-\infty}^{\infty} \Big[ \sum_{i=1}^{n} g(X_i, t; \hat{\lambda}) \Big]^2 dM(t), \end{aligned}$$

where $d_{ijn} \equiv 1/n$, $g(x, t; \hat{\lambda}) = I[\hat{\mu} + \hat{\sigma} b_{t-1} < x \leq \hat{\mu} + \hat{\sigma} b_t] - p_t$, and $M$ is the discrete measure placing mass $p_j^{-1}$ on the point $t = j$, where $j = 1, \ldots, k$. Thus

$$\mu(t, \gamma) = E_\lambda[g(X_i, t; \gamma)] = F_0(\gamma_1 + \gamma_2 b_t) - F_0(\gamma_1 + \gamma_2 b_{t-1}) - p_t.$$

Note that $WV_n(\hat{\lambda})$ has been scaled up by $n$, so $d_{ijn} \equiv 1/n$. Without loss of generality, take $\mu = 0$ and $\sigma = 1$. The conditions of Theorem 1 can be shown to hold with

$$\mathbf{d}_1 \mu(t_j; \lambda) = \begin{bmatrix} f_0(b_j) - f_0(b_{j-1}) \\ b_j f_0(b_j) - b_{j-1} f_0(b_{j-1}) \end{bmatrix}, \; j = 1, \ldots, k,$$

where $f_0(b_0) = f_0(b_k) = 0$. Since $d_{ijn} \equiv 1/n$, we have $\eta_1 = 1$ and $\eta_i = 0$, for $i = 2, \ldots$ By Theorem 1, $WV_n(\hat{\lambda})$ converges to

$$\int_{-\infty}^{\infty} \Big\{ Z_1(t) + c_1 \mathbf{d}_1 \mu(t; \lambda)' \mathbf{Z} \Big\}^2 dM(t).$$

For given $F_o$ and $\alpha(\cdot)$, the limiting distribution of $WV_n(\hat{\lambda})$ can be obtained.

## 4. Concluding Remarks

The limiting distribution of a test statistic based on the Pearson correlation coefficient is presented and a simulation study shows that the statistic is useful in general situations. Note that the test statistics studied in this paper can also be applied to estimate association of state-to-state migration rates.

The limiting distributions of weighted degenerate U- and V-statistics with estimated parameters have been established. These results extend the literature concerning weighted degenerate U-statistics.

Classes of U-, V- and L-statistics having kernels of order 1 and 2 with estimated parameters have been studied since Sukhatme (1958). The study grew intense after the appearance of Pierce (1982) and Randles (1982). The limiting distribution of symmetric statistics of arbitrary order was established in Dynkin and Mandelbaum (1983). Exploring the limiting distribution of symmetric statistics of arbitrary order with estimated parameters will extend the study of statistics with estimated parameters in the kernels.

## Acknowledgements

I wish to express my deep gratitude to Jed Frees for sharing ideas of this problem, to Tertius De Wet for his careful reading and suggestions which helped improve an early draft of this work greatly, and to Z. D. Bai for his valuable consultation during many discussions. The author also thanks Paul Speckman, M. T. Tsai and H. C. Ho for helpful discussions, and an associate editor and a referee for constructive comments. Research was supported in part by NSF Grant DMS-9308756, Summer Research Fellowship and Research Council Grant from University of Missouri-Columbia, and NSC Grant 85-2121-M001-003.

## References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

De Wet, T. and Randles, R. H. (1984). On limiting chi square U- and V-statistics forms with auxiliary estimators. Technical Report No. 227, Dept. of Statist. Univ. of Florida.

De Wet, T. and Randles, R. H. (1987). On the effect of substituting parameter estimators in limiting $\chi^2$, U and V statistics. *Ann. Statist.* **15**, 398-412.

Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* **11**, 739-745.

Frees, E. W. (1992). Forecasting state-to-state migration rates. *J. Business and Economic Statist.* **10**, 153-167.

Frees, E. W. (1993). Short-term forcasting of internal migration. *Envior. Planning Ser. A* **25**, 1593-1606.

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *J. Econom.* **69**, 393-414.

Greene, W. H. (1993). *Econometric Analysis.* MacMillan, New York.

Gregory, G. G. (1977). Large sample theory for U-statistics and test of fit. *Ann. Statist.* **5**, 110-123.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge University Press, New York.

Iverson, H. K. and Randles, R. H. (1989). The effects on convergence of substituting parameter estimates into U-statistics and other families of statistics. *Probab. Theory Related Fields* **81**, 453-471.

Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10**, 475-478.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462-474.

Randles, R. H. (1984). On tests applied to residuals. *J. Amer. Statist. Assoc.* **79**, 349-354.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Shieh, G. S. (1994). Weighted degenerate U- and V-statistics with estimated parameters. Technical Report No. C94-11, Instit. of Statistical Sci., Academia Sinica, Taiwan.

Shieh, G. S., Johnson, R. A. and Frees, E. W. (1994). Testing independence of bivariate circular data and weighted degenerate U-statistics. *Statist. Sinica.* **4**, 729-747.

Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1**, 261-290.

Sukhatme, B. V. (1958). Testing the hypothesis that two populations differ only in location. *Ann. Math. Statist.* **29**, 60-78.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

E-mail: gshieh@stat.sinica.edu.tw

(Received June 1994; accepted December 1996)