Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n3p491

New iterative AM estimation procedure for fitting the simple linear measurement error models

By Al-Nasser, Al-Sliti, Edous

Published: November 2016

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 03, November 2016, 491-501
DOI: 10.1285/i20705948v9n3p491

New iterative AM estimation procedure for fitting the simple linear measurement error models

Amjad D. Al-Nasser*[a], Ayat Al-Sliti[b], and Midhet Edous[b]

[a] *Quality Assurance and Institutional Effectiveness Center, Al Falah University, Dubai, UAE*
[b] *Department of Statistics, Science Faculty, Yarmouk University, Irbid, Jordan*

Published: November 2016

This article proposed a modified AM estimation procedure. The procedure uses the grouping estimators iteratively after dividing the sample into clusters. Then, the grouping AM procedure used to fit the structural relationship with measurement error considering there is no equation error model. The performance of the iterative grouping estimator is compared with the traditional two group estimators. Simulation study showed that in terms of mean square error the proposed estimator is robustify the traditional two group estimator. A real data analysis for studying the relationships between happiness rate and the corruption perceptions index in the Arabs states is discussed.

keywords: Measurement Error Model, Grouping Method, Happiness Rate, Corruption Perceptions Index.

1 Introduction

In studying the relationships between two variables; the researchers usually use the simple linear regression to fit the relationship between two variables. However, simple linear regression is not reliable in terms it consider the error exist in one variable only, to overcome this problem we consider of using the measurement error model (MEM) or the so called errors-in-variables model (Fuller, 2009; Kendall and Stuart, 1961), which fit the simple linear relations when both variable are subject to errors. In MEM, the two latent variables ξ and η are assumed to be functionally related in the form:

*Corresponding author: amjadyu@yahoo.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

$$\eta = \alpha + \beta\xi$$

The two latent variables ξ and η are observed subject to mutually independent errors δ_i and ϵ_i) which have known means and unknown variances σ^2 and τ^2; respectively. Then for a random sample of size n, the observed pairs (x_i, y_i) are used to measure the actual values that satisfies the linear relationship in the following form:

$$y_i = \eta_i + \epsilon_i$$
$$x_i = \xi_i + \delta_i$$

where

$$\eta_i = \alpha + \beta \times \xi_i; i = 1, 2, ..., n$$

The MEM in (1) contains (n+4) parameters to be estimated,$\alpha, \beta, \sigma^2, \tau^2$ and the latent variables or the so-called incidental parameters $\xi_1, \xi_2, ..., \xi_n$ (or equivalently $\eta_1, \eta_2, ..., \eta_n$). The presence of the incidental parameter leads to inconsistencies of the estimators (Cheng and Van Ness, 1999). To overcome the inconsistencies of the estimators, either one piece of information on the errors variances or the intercept is required; or a non-parametric estimation procedure is used. Most of the estimation methods used in the literature to estimate the model parameters based on normality assumption and uses extra information about the precision error. However, in some situations it is unreliable to use the normality assumption to identify the model, therefore the second option of using a non-parametric estimation approach will be a wise alternative. (Hussin, 2004; Al-Nasser, 2010; Al-Nasser, 2014; Carpita and Ciavolino, 2016; Ciavolino and Carpita, 2015; Ciavolino et al., 2015; Carpita and Ciavolino, 2014). In this article, we are more interested in the non-parametric procedures.

There are several non-parametric estimation methods used to fit the MEM. The iterative reweighted least squares approach is used by Huwang and Huang (2000), to estimate the unknown parameters in the polynomial MEM Berkson model. Fan and Truong (1993); construct a new class of kernel estimators in nonparametric regression and account the errors in covariates. Al-Nasser and Ebrahem (2005) suggested a robust Wald-type estimator for MEM model based on the idea of L-statistic by considering the trimmed and Winsorized means. Also, Ebrahem and Al-Nasser (2005) suggested the AM estimation procedure to fit MEM; this procedure also used by Al-Nasser and Ebrahem (2005) to fit the simple linear regression. Both articles found that the AM estimator to be a good alternative to traditional methods including OLS estimation method, Wald type methods, repeated median, Geometric mean and Housner Bernnan (HB) estimation methods, and it has the ability to achieve satisfactory results even in the presence of a large number of outliers. Later, Ghapor et al. (2015) and Al Mamun et al. (2015) suggested a modified AM procedure to estimate the slope parameter in MEM.

In this article the wald type estimators will be merge with the AM estimation procedure to come up with a new iterative estimation procedure. The remainder of this paper is divided into three sections. Section 2 Proposed the new iterative AM grouping estimation procedure. Section 3 presents Monte Carlo evidence on the numerical performance of the proposed estimators in fitting the MEM. Section 4 include a real data analysis to study the relationships between happiness and transparency in the Arabs world; and Section 5 presents concluding comments.

2 Iterative AM Estimation Procedure

The iterative AM estimation procedure merge between AM procedure (Al-Nasser and Ebrahem, 2005) and the two group mean based estimation procedure (Wald, 1940). The procedure consists of the following two main steps:

- **Step 1**: AM stage
  a. Arrange the observations in ascending order on the basis of the values of x_i ; i.e.,$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$, and the associated $y_{[1]}, y_{[2]}, ..., y_{[n]}$ of the original data are taken; then the new pairs will be $(x_{(i)}, y_{[i]})$
  b. Divide the data into m subgroup each of size r such that $m \times r = n$; then the sample can be rewritten in the form:

$$(x_{(1)}, y_{[1]}); (x_{(2)}, y_{[2]}); ....; (x_{(r)}, y_{[r]})$$
$$(x_{(r+1)}, y_{[r+1]}); (x_{(r+2)}, y_{[r+2]}); ....; (x_{(2r)}, y_{[2r]})$$
$$\vdots$$
$$(x_{((m-1)r+1)}, y_{[(m-1)r+1]}) ; (x_{((m-1)r+2)}, y_{[(m-1)r+2]}) ;....;(x_{(mr)}, y_{[mr]})$$

  It is suggested by Al-Nasser and Ebrahem (2005) m can be chosen to be the maximum divisor of n such that $m \leq r$; in this article we suggest that m can be chosen to be any divisor, and based on intensive simulation results we found the vis versa selection gave better estimates in many cases.

- **Step 2**: Compute the grouping estimator iteratively:
  c. Find all possible pair grouping slopes:
  In such case we have to compute all possible two groups estimators between all sub-samples:

$$b(j) = \{\frac{\overline{y_{[g2]j}} - \overline{y_{[g1]j}}}{\overline{x_{(g2)j}} - \overline{x_{(g1)j}}}; j = 1, 2, ..., \frac{m(m+1)}{2}\}$$

  d. The estimated value of the slope can be defined as follows:

$$\widehat{\beta} = \{b(1), b(2), ..., b(\frac{m(m+1)}{2})\}$$

  and

$$\widehat{\alpha} = \overline{y} - \beta\overline{x}$$

3 Simulation study

In this section, a simulation study is conducted to investigate the performance of the MEM estimators using the proposed procedure and the classical grouping estimators. In the simulation study, 10.000 samples of sizes; 50, 75, 100 and 200 were generated from the MEM under the following conditions:

1. The parameter values are ($\alpha = 1, \beta = 1$);
2. The error terms were generated from standard normal distributions;
3. Sample size is selected as given in the following table for iterative AM grouping procedure.

Table 1: Sample size selections

n	m	r
50	10	5
75	15	5
100	10	10
200	20	10

4. Then the data was contaminated; at each step a certain percentages 10 $\perp$ and 20 $\perp$ of the observations were deleted and replaced with outliers observations. The contaminated data point was generated according to the given relationship where:

- In y only outliers, $\epsilon_i : N(0, 16); \delta_i : N(0, 1)$
- In x only outliers, $\epsilon_i : N(0, 1); \delta_i : N(0, 16)$.
- In both x and y outliers, $\epsilon_i : N(0, 16); \delta_i : N(0, 16)$

The properties of these estimators were investigated in terms of the simulated bias, mean squared error (MSE) and the efficiency (Eff):

$$Bias(\widehat{\theta}) = \frac{\sum(\widehat{\theta}_i - \theta)}{10000},\ MSE(\widehat{\theta}) = \frac{\sum(\widehat{\theta}_i - \theta)^2}{10000} \text{ and } Eff = \frac{MSE(\widehat{\theta}_{AM})}{MSE(\widehat{\theta}_{twogroup})}$$

where $\widehat{\theta} = \widehat{\alpha}$ or $\widehat{\beta}$ and $\widehat{\theta}_i$ is the estimate of the unknown parameter based on the i^{th} sample; given that $i = 1, 2, ..., 10000$.

The simulations results of the Bias, MSE and efficiency are tabulated in Table 2 in case no outliers in the data and in Tables 3, 4 and 5 for the case contamination in X only, Y only and in Both, respectively. In all cases and for all sample sizes the iterative

AM estimation procedure robustify the two grouping methods as all the AM estimates are more efficient with efficiency value more than 1. These results indicate that the proposed estimation procedure is better to be use in fitting the MEM.

Table 2: Comparisons results when the data contains no outliers

Sample Size	Method	$\widehat{\alpha}$			$\widehat{\beta}$		
		Bias	MSE	Eff	Bias	MSE	Eff
50	Two Groups	-.002	.061	1.23	-.995	1.035	1.02
	Iterative AM	.001	.049		1.006	1.013	
75	Two Groups	-.002	.041	1.03	-1.002	1.031	1.02
	Iterative AM	-.004	.039		1.004	1.008	
100	Two Groups	-.0003	.0031	1.07	-1.009	1.023	1.03
	Iterative AM	-.003	.029		.993	.987	
200	Two Groups	.001	.015	1.09	-.999	1.008	1.02
	Iterative AM	-.0007	.013		.991	.985	

Table 3: Comparisons results when the data contains outliers in x Only

outliers %	Sample size	Estimation Method	$\widehat{\alpha}$			$\widehat{\beta}$		
			Bias	MSE	Eff	Bias	MSE	Eff
10%	50	Two Groups	-.002	.399	1.58	1.564	2.449	1.17
		Iterative AM	.002	.253		-1.432	2.082	
	75	Two Groups	.0018	.252	1.59	1.524	2.325	1.12
		Iterative AM	-.0046	.158		-1.432	2.072	
	100	Two Groups	-.006	.202	1.22	1.479	2.294	1.33
		Iterative AM	-.002	.164		-1.304	1.732	
	200	Two Groups	.006	.101	1.42	1.551	2.205	1.27
		Iterative AM	.005	.077		-1.313	1.717	
20%	50	Two Groups	.009	.956	3.24	2.225	4.459	1.97
		Iterative AM	-.013	.295		-1.579	2.515	
	75	Two Groups	-.501	.617	3.37	2.193	4.812	1.90
		Iterative AM	-.001	.184		-1.587	2.531	
	100	Two Groups	.004	.461	2.12	2.181	4.761	2.38
		Iterative AM	.008	.217		-1.408	1.995	
	200	Two Groups	.003	.228	2.15	2.187	4.681	2.37
		Iterative AM	-.006	.106		-1.403	1.974	

Table 4: Comparisons results when the data contains outliers in y Only

outliers %	Sample size	Estimation Method	$\widehat{\alpha}$			$\widehat{\beta}$		
			Bias	MSE	Eff	Bias	MSE	Eff
10%	50	Two Groups	-.001	.692	1.19	1.239	1.538	1.08
		Iterative AM	.005	.557		-1.001	1.417	
	75	Two Groups	.003	.429	1.17	1.208	1.526	1.24
		Iterative AM	-.008	.367		-.999	1.239	
	100	Two Groups	-.006	.194	1.11	1.137	1.508	1.02
		Iterative AM	-.002	.173		-.998	1.484	
	200	Two Groups	.002	.163	1.12	1.123	1.217	1.12
		Iterative AM	.001	.145		-.997	1.091	
20%	20%	Two Groups	-.003	1.321	1.41	1.488	2.216	1.23
		Iterative AM	-.004	.934		-.996	1.801	
	75	Two Groups	-.007	.829	1.50	1.485	2.206	1.25
		Iterative AM	-.012	.551		-1.002	1.763	
	100	Two Groups	.001	.396	1.37	1.452	2.182	1.57
		Iterative AM	.005	.289		-1.005	1.388	
	200	Two Groups	.006	.370	1.33	1.451	2.167	1.83
		Iterative AM	-.001	.278		-1.002	1.180	

Table 5: Comparisons results when the data contains outliers in both x and y

outliers %	Sample size	Estimation Method	$\widehat{\alpha}$			$\widehat{\beta}$		
			Bias	MSE	Eff	Bias	MSE	Eff
10%	50	Two Groups	.003	.867	1.08	1.653	2.742	1.24
		Iterative AM	.003	.797		-1.433	2.188	
	75	Two Groups	.001	.495	1.04	1.600	2.563	1.19
		Iterative AM	-.001	.474		-1.432	2.138	
	100	Two Groups	-.007	.440	1.05	1.541	2.516	1.42
		Iterative AM	-.034	.422		-1.307	1.765	
	200	Two Groups	.0002	.431	2.08	1.456	2.423	1.38
		Iterative AM	.002	.207		-1.314	1.752	
20%	20%	Two Groups	-.016	1.586	1.13	2.467	6.092	2.33
		Iterative AM	-.015	1.403		-1.575	2.607	
	75	Two Groups	-.004	.933	1.03	2.489	5.982	2.91
		Iterative AM	-.002	.904		-1.581	2.577	
	100	Two Groups	.008	.659	1.28	2.487	5.977	2.91
		Iterative AM	.009	.514		-1.407	2.048	
	200	Two Groups	-.005	.369	1.01	2.479	5.914	2.96
		Iterative AM	-.003	.365		-1.401	1.995	

4 Relationships between CPI and Happiness Rate in the Arabs world

The Corruption Perceptions Index (CPI) scores is consider a composite statistical barometer of transparency that measure which indicates the perceived level of public sector corruption on a scale of 0-100, where 0 means highly corrupted country and 100 means that a country is perceived as very clean (see Transparency International, 2016. The Corruption perception index 2015, *https://issuu.com/transparencyinternational/docs/2015_corruptionperceptionsindex_rep?e=2496456/33011041.* Retrieved April, 2016). On the other hand, Happiness rate (HR) is an interesting proper social measure and start to be a goal of public policy. In the Arab region, consider one of the sadness region on the earth, still has some promising countries and the people who lives in these countries (such as United Arab Emirates and Qatar) are more happy than the people who lived in a developed countries. It worth to say that; His Highness Sheikh Mohammed bin Rashid Al Maktoum, Vice President and Prime Minister of the United Arab Emirates (UAE), and Ruler of Dubai usually talking about the importance of well-being as a guide for their nations and as new development of the happiness factor in his country they create a local ministry of happiness in 2015. Happiness rate depends on many factors and one of the most interesting factors is the CPI, and we believe that CPI effect positively in increasing the citizens happiness attitude in any country. In this article, we used a publish data of 18 Arabs countries that have both CPI and HR measures (see Transparency International - above - and Sustainable Development Solutions Network, SDSN, 2016. World Happiness Rate Report. *http://worldhappiness.report/.* Retrieved April, 2016.).

Table 6 represents a descriptive statistics of CPI and HR in the Arabs world. the results indicated that the minimum HR was in Syria (3.069) however, the maximum HR was United Arab Emirates (6.54). Also, the Arab country that is highly perceived corrupted is Somalia (CPI = 8); while Qatar is the least perceived corrupted Arab country with CPI equal to 71. Also, it worth to say that there is a strong positive and significant linear correlation (r = 0.69, p=0.001) between CPI and HR.

Table 6: Descriptive Statistics

Variable	Min	Max	Mean	Stdev	Correlation Coeffecient
PCI	8	71	36.28	19.213	0.693 ; p =0.001
HR	3.069	6.545	5.32	1.030	

In order to have better insight on the Arabs region, the two variables; CPI and HR, where classified into four coordinates based on the sample quartiles (See Table 7).

to have in-depth vision on this region a scatter plot is constructed in Figure 1; and different references lines were used based on the quartiles criterion that given in Table 7.

Table 7: Arabs Countries classification based on HR & CPI

Status	HR	Level	CPI
Very Sad	HR $\leq$ 4.52	Very Low	CPI $\leq$ 17.5
Sad	4.52<HR $\leq$ 5.37	Low	17.5<CPI $\leq$ 36
Happy	5.37<HR $\leq$ 6.27	High	36<CPI $\leq$ 51.25
Very Happy	HR >6.27	Very High	CPI >51.25

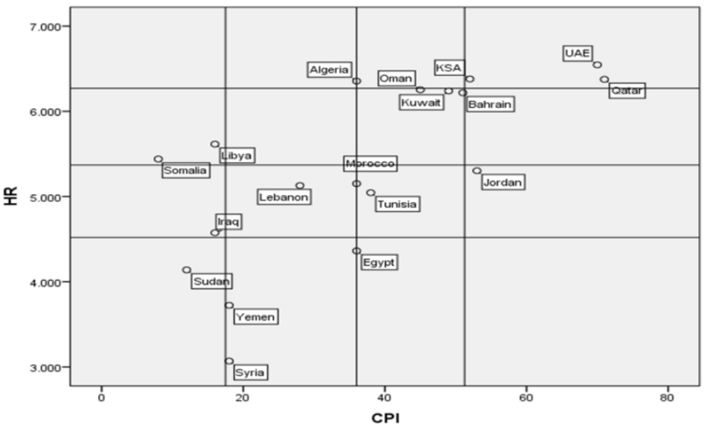

Figure 1: Arab countries classification

The results in Figure 1 indicated that the Arabs region can be classified into three set of countries; the first one is the happiest and least corrupted countries which include UAE and Qatar, and we believe that these two countries have luxury living standard in the region, the second one in between which has slightly happy citizens and medium corruptions or vis versa, which includes KSA, Kuwait, Oman, Bahrain and Algeria and the last set of countries are the sad countries with highly perceived public sector corruption; which includes Jordan, Morocco, Tunisia, Egypt, Iraq, Syria, Libya, Yemen Lebanon, Somalia and Sudan. It is very clear, the first two sets includes only the Arab gulf countries and Algeria. However, the third set of countries includes Arab countries that have a very serious problem that caused by political changed or problems.

5 Fitting Data to Equation

CPI and HR can be modeled in a linear relationship based on Figure 2.

However, and we believe that linear MEM is better to be used rather than using the simple linear regression, because both variables were measured in different criterion and based on several factors, the over all computation leads to the given results. Therefore, two types of error could occurs in computing both measures, one the sampling error and

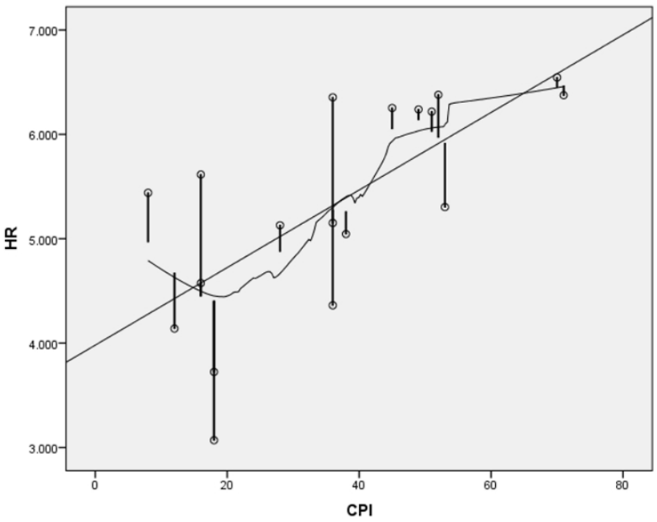

Figure 2: Structural Relationships Between HR and CPI

the other in reformatting and approximation the results. Therefore, in equation 1, η will be the exact value of HR, and ξ will be the exact value of CPI.

$HR = \alpha + Beta \times (CPI - \delta) + \epsilon$

Table 8: Parameter estimation

Method	Intercept	Slope	MSR
Two Groups	3.868	0.041	0.527
Iterative Grouping Method m =6, r =3	4.457	0.024	0.581
Iterative Grouping Method m =3, r =6	4.0885	0.0342	0.524

Estimating the unknown parameters in this model based on the observed values of the 18 Arabs countries the results are given in Table 8. However; Figure 3 give a residual comparisons between the three suggested models.

6 Concluding Remarks

This study proposed a new non-parametric estimation procedure to fit the MEM. The new procedure used iterative steps of the AM procedure and in each single step a grouping estimator were used. The Monte Carlo simulations provide a good evidence for the superiority of the proposed estimation procedure on the classical two groups methods in all cases of the data (with and without outliers) and for any sample size. Moreover, the estimation procedure were applied on a real data obtained about the Arabs regions to study the effect of the CPI on the HR. The data analysis suggested that there is a

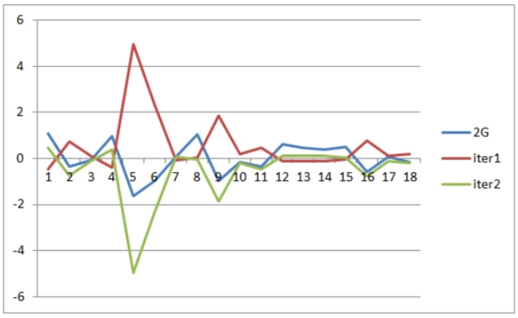

Figure 3: Residual comparisons

strong positive relationship between both variables under study. Also, the data analysis suggested that the Arab's region could be classified into three countries sets based on the CPI and HR.

References

Al Mamun, A. S. M., Zubairi, Y. Z., Hussin, A. G., and Imon, A. R. (2015). A new technique to solve the unidentifiablity problem of linear structural relationship model. In *THE 2ND ISM INTERNATIONAL STATISTICAL CONFERENCE 2014 (ISM-II): Empowering the Applications of Statistical and Mathematical Sciences*, volume 1643, pages 242–247. AIP Publishing.

Al-Nasser, A. D. (2010). An information-theoretic approach to ultra structural measurement error model. *Statistics in Transition New Series*, 11(1):9–24.

Al-Nasser, A. D. (2014). Two steps generalized maximum entropy estimation procedure for fitting linear regression when both covariates are subject to error. *Journal of Applied Statistics*, 41(8):1708–1720.

Al-Nasser, A. D. and Ebrahem, M. (2005). A new nonparametric method for estimating the slope of simple linear measurement error model in the presence of outliers. *PAKISTAN JOURNAL OF STATISTICS-ALL SERIES-*, 21(3):265.

Carpita, M. and Ciavolino, E. (2014). Mem and sem in the gme framework: Statistical modelling of perception and satisfaction. *Procedia Economics and Finance*, 17:20–29.

Carpita, M. and Ciavolino, E. (2016). A generalized maximum entropy estimator to simple linear measurement error model with a composite indicator. *Advances in Data Analysis and Classification*, pages 1–20.

Cheng, C. and Van Ness, J. (1999). Statistical regression with measurement errorarnold.

Ciavolino, E. and Carpita, M. (2015). The gme estimator for the regression model with a composite indicator as explanatory variable. *Quality & Quantity*, 49(3):955–965.

Ciavolino, E., Carpita, M., and Al-Nasser, A. (2015). Modelling the quality of work in the italian social co-operatives combining npca-rsm and sem-gme approaches. *Journal of Applied Statistics*, 42(1):161–179.

Ebrahem, M. A.-H. and Al-Nasser, A. D. (2005). Estimating the slope of simple linear regression in the presence of outliers. *Journal of Modern Applied Statistical Methods*, 4(2):509–513.

Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, pages 1900–1925.

Fuller, W. A. (2009). *Measurement error models*, volume 305. John Wiley & Sons.

Ghapor, A., Zubairi, Y., Mamun, A., and Imon, A. (2015). Pak. j. statist. 2015 vol. 31 (3), 339-350 a robust nonparametric slope estimation in linear functional relationship model. *Pak. J. Statist*, 31(3):339–350.

Hussin, A. G. (2004). Numerical comparisons for various estimators of slope parameters. *Matematika*, 20:19–30.

Huwang, L. and Huang, Y. S. (2000). On errors-in-variables in polynomial regression-berkson case. *Statistica Sinica*, pages 923–936.

Kendall, M. and Stuart, A. (1961). Vol. 2: Inference and relationship.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11(3):284–300.