Journal of Forensic Economics 19(3), 2006, pp. 317-323

Out-of-Step with Cohort Analysis: The Problems with "Plain and Fancy Regression"

A.E. Rodriguez*

I. Introduction

In a recent paper in this journal, Peterson (2003) argues that "cohort analysis" is an empirical technique that is no different from basic plain-vanilla multiple regression analysis. I argue in this paper that Peterson's argument is incorrect, or at the very least, inaccurate.

There are two fundamental interrelated problems with the Peterson paper. First, he confounds readers by labeling his exposition "cohort analysis." Cohort analysis (Glenn 2005) is a well known technique in econometrics but unfortunately, not the technique Peterson actually uses in his paper.[1] To be sure, cohort analysis is a phrase often heard in courts and in the forensic literature distinct from the econometric technique, and even then the definition of a cohort can vary widely depending on the circumstances of the analysis. (Piette and Thornton, 1995) But Peterson's "cohort-analysis" refers to a *sui generis* series of rank sum, non-parametric tests on idealized data which are neither the former nor an example of the latter. Second, Peterson presumes to show that his cohort-analysis allegedly conveys the same information as a regression model with a complete set of interactive effects among the independent variables when used to empirically examine the presence of gender discrimination in a hypothetical firm. However, this similarity is the case only in the stylized example offered by Peterson. Put differently, for an expert witness mulling the appropriate tool to appraise the likelihood of gender or race discrimination (or both), Peterson's-cohort-analysis does not represent a valid alternative to multiple regression.

Peterson's simulated data carefully and ingeniously avoid problems that may limit the use of the rank sum test in the first place; if the rank sum test is limited then there can be no generality. To demonstrate the lack of generality, I offer at least two common conditions in which regression analysis is robust but where rank sums fear to tread; for one condition I use the same model of pay discrimination that Peterson does; for the other I rely on Berndt's (1991) well-known publicly available labor market discrimination data. In so doing I rebut the claim that Peterson's "cohort analysis" is "just a regression plain and fancy."

The two elements of Peterson's argument that I criticize are interrelated because—paradoxically—*classical* cohort analysis could be of use in the analysis of pay discrimination if there is any indication that the *aging* of the work force had an asymmetric impact in the determination of pay across gender (the only

*Associate Professor, Department of Economics & Finance, University of New Haven, CT 06516.
[1]See Rodgers, et al., (1996) for an example of classical cohort analysis published in this journal.

protected grouping in Peterson's simulated data). However, the simulated data generated and used by Peterson does not contain this covariance, by design; if it did, rank sum analysis may not be able to provide the answers and conclusion Peterson desires. In fact, Peterson's rank sum tests work *because* they contain no covariance whatsoever with continuous variables, for the most part.

In the section immediately following we first describe Peterson's simulation methodology and then explain it. Then we point to the confusion arising from Peterson's use of the name "cohort analysis." Last, we show how rank sum tests do not hold up under scrutiny; that it is *the same as a multiple regression analysis* only under special conditions that are unlikely to appear in a litigation context.

There is some appeal to Peterson's approach: its inherent simplicity, a feature that is especially appealing to specialists tasked with presenting information in a way that is both accurate and understandable to jurors or a judge who may not be statistically literate. A technique that is simpler than multivariate regression might be useful as long as it does not misrepresent the conclusion of a more thorough analysis of the data. But no forensic economist should use Peterson's "cohort analysis" in place of regression analysis without also having conducted a thorough regression analysis first and ensuring that the results of the method used are robust under other more sophisticated methods.

II. Peterson Simulates Perfect—and Idiosyncratic—Data

To explain and motivate his argument, Peterson creates an artificial data set for a hypothetical engineering consulting firm called *Air Quality Consultants*. Air Quality Consultants employs engineers and accountants. The data are parsed by specialty, level (based on "prior experience") and by gender. Employee pay in this hypothetical firm is set by experience level and by specialty (i.e., whether an employee is an accountant or an engineer). Thus, the firm's employees are classified as level 1 through level 4 depending on prior experience. The pay data for each employee within each cohort, i.e. accountant-level-1, are set by a random drawing from a uniform distribution; the endpoints of the uniform distribution (and which completely characterize the distribution) are set by Peterson but they are increasing in experience levels. Engineers are paid more than accountants; the differential between engineers and accountants is established by Peterson. No difference is attributed to gender. This setup ensures that the data reflect no evidence of gender disparity in pay.

Air Quality Consultants comes under Peterson's scrutiny for pay discrimination across gender. Peterson statistically tests a null hypothesis of no gender discrimination; alas, at the first run of a simple model in levels, a test of the null-hypothesis of no gender discrimination cannot be rejected! This is of course, at first blush, a surprise because we were privy to how the data were generated to specifically avoid gender discrimination. This false positive is a result of deliberately embedding heteroskedasticity in the data obtained by increasing the variability in pay at the higher salary levels for engineers and at the same time reducing the number of women present among the highly-paid engineers.

The heteroskedasticity becomes transparent when a full-fledged multivariate regression model is estimated; a model that relies on a full set of levels and interactions between variables for gender levels and for both engineers and for accountants. Of course, the coefficients having been correctly estimated, the null is not rejected.

Peterson proceeds to show that it is possible to arrive at the same inference of no gender discrimination by conducting a series of rank sum tests, testing for discrimination at each cohort and then aggregating the rank sums across all cohorts:

> ...we have completed the transition from the simple homoscedastic regression model with no interaction to the rank sum cohort analysis...showing that the latter is in effect a regression analysis... (p. 169)

It is—but only in the particular and unlikely special case designed by Peterson; we show this below.

III. Peterson's Cohort Analysis Analyzes Cohorts But It is Not Classical Cohort Analysis

Peterson refers to those employees classified in a particular level and a specialty as a "cohort." For example, engineers-in-level 3 constitute a "cohort" separate and distinct from engineers-in-level 1 or accountants-in-level 2. He then proceeds to analyze each cohort for the presence of gender discrimination within each cohort: he does this via a multivariate analysis with interaction and via a rank sum analysis. But even though Peterson's analysis involves cohorts—as the term is understood by laymen: "a group or company"—it would ordinarily not be considered "cohort analysis" by an econometrician.

Classical "cohort analysis" refers to a specific and different data-examining strategy not present in either of the two techniques Peterson uses to scrutinize discrimination. A popular tract on cohort analysis explains as follows:

> The term cohort analysis is usually reserved for studies in which two or more cohorts are compared with regard to at least one dependent variable measured at two or more points in time. (Glenn, 2005, **p.)**

The most common goal of cohort analysis is the assessment on a particular variable of the consequences of growing older, or, in other words, estimation of *age effects*.

Ironically, had Peterson incorporated age effects into his simulation specification he may have relied on classical cohort analysis to examine the possible impact of aging on any observed gender discrimination. Of course, classical cohort analysis comes with its own intellectual baggage, often demanding a sufficiently restrictive set of conditions that easily persuade any analyst to choose an alternative method.

The presence of age effects—a continuous variable—on gender differentials in the firm, with their possible interactions with other explanatory variables would have likely vitiated Peterson's conclusion: that you get the same answers with rank sum tests of the "cohorts." This is the general point of the next section.

An additional complication to the Peterson model, from the presence of a continuous variable such as age, would have occurred from the necessary loss of information resulting from the formation of age intervals required for the rank sum tests. Peterson ingeniously avoids this loss of information problem in the first instance with the otherwise continuous "experience" variable by neatly assuming that employees are categorized based on experience, thereby converting the continuous variable to an interval variable prior to the multivariate regression rank sum equivalence argument. The loss of information would have increased variability and may have disallowed Peterson's claim that rank sum tests and multivariate regression elicit the same information from the data. These additional complications to the Peterson argument are not illustrated in this paper.

IV. Peterson's Rank Sum Method Does Not (Necessarily) Work when the Pay Generating Mechanisms are Characterized by Different Distributions

Peterson's results arise only from the presence of "tidy" data representing Air Quality Consultants' pay structure. Tidy data follow from assuming that the same pay-generating distribution underscores the reported pay data for the cohorts: a uniform distribution. But what happens to the rank sum equivalence if the distributions are dissimilar? It may disappear. Indeed, in a litigation context there is no reason to assume that the generating distributions are similar.[2]

To demonstrate, I will alter one of the distributions generating salary for women in cohort engineer level-2. I chose to use a Pareto distribution to generate the pay data for women and continue to use the same data for men, generated by Peterson using a uniform distribution with $70,000 and $74,000 as the relevant endpoints. Consistent with Peterson's original specification, the average pay for women and men remain approximately the same.

The Pareto distribution is often used in the analysis of pay inequality to simulate earnings data (Bernstein and Mishel, 1997). In generating the data used here I allow women's pay to be characterized as a random variable distributed Pareto.[3]

Table 1

	Females	Males
Mean	72,412	73,424
Standard Deviation	1,156	93,306
Trials	20	90

[2]Paradoxically, Peterson acknowledges this, but proceeds undisturbed: "Of course, in any practical situation, we would most probably not know that the underlying data came from uniform populations, or from any other particular family." (p. 168)

[3]The mathematical characterization of the parameters of the Pareto distribution is left out to conserve space.

V. Results

We generated 20 pay data points from the Pareto distribution characterized by a minimum of 70,000 and a shape parameter of 2.5. To maintain the 9:2 ratio of men to women I sample 90 times from the uniform distribution characterized by a minimum of 70,000 and 74,000. Table 1 contains the generated data. Consistent with Peterson, the difference of means is structured to establish no statistically significant difference across gender; where statistical significance is the Hazelwood standard and the test is a conventional t-test.

We find that when the data are generated from different distributions, the two sample t-tests with unequal variances are robust to the inference of discrimination: that is, it finds, correctly, that the means are not different.

This is well known: "Most standard tests are remarkably robust to many kinds of non-normality..." (Kennedy, 1995, p. 91). However, the rank sum tests find a difference, which is not surprising since the ranks tests are testing for different populations.

VI. Peterson's Rank Sum Method Does Not (Necessarily) Work in the Presence of Continuous Variables

Peterson's argument turns on the presence of neat categorical groupings of employees: e.g., Engineer-Female-Level 3; put differently, if confronted with incorporating a continuous variable into his framework, such as age, experience or education, Peterson's approach would necessitate collapsing the continuous variable into dichotomous or polychotomous categorical variables.[4,5] Such a collapsing of variables necessarily reduces the information and imputes bias in the variable.

In addition, there is considerable evidence in the discrimination literature of interaction effects between gender and race, and many continuous variables including age, education and experience. However, rank sum tests are not robust to the presence of interaction effects.

Peterson claims that his results reflect possible interactions. Note how cleverly Peterson's model enables him to assert as much: the interactions are all between categorical variables (i.e., does there exist pay asymmetry for female engineers in level 3? Where engineer, female and level 3 are all categorical variables.) Interaction effects are generally considered omitted variable problems. Again, the rank sum tests on categories would not be robust to this unless you know, ahead of time, that you were missing variables. For example: consider the well known Berndt data on discrimination (Berndt, 1991). Berndt's 1991 econometrics textbook contained extracts (drawn randomly) from the May 1978 and May 1985 Current Population Survey; it can be found online at: http://shazam.econ.ubc.ca/student/berndt/data.htm. The May 1978

---

[4]Peterson recognizes the problem without acknowledging the loss of information: "One way to deal with this difficulty is to base each intra-cohort comparison on pay rate rank, rather than the exact numerical values of the pay rates themselves." (p. 168)

[5]Again, Peterson fails to heed his own caveat: "If pay rates are possibly non-normal and possibly interdependent, the reliability of even the non-constant variance full-interaction regression model is called into question." (p. 168)

CPS extract contains observations on 20 variables for 550 individuals; the May 1985 CPS extract contains 534 observations on 19 variables.

Suppose we are concerned with gender discrimination in unions. A simplistic linear model regressing the logarithm of wages on variables controlling for union membership, gender and a cross-dummy for gender discrimination in unions results in a t-statistic of -1.72 (p = 0.087) for the presence of gender discrimination in unions using the CPS extract for May 1978. Although the model is necessarily simplified by omitting relevant variables for expository purposes, the result would not support an inference of gender discrimination in a litigation context under the Hazlewood standard.

Now let's use Peterson's approach on the same 1978 CPS extract: this would imply a series of rank sum tests of pay data for unions by gender. Thus, we have, two Peterson cohorts: union and non-union. The rank sum test finds significant differences between pay for women and men in both union and nonunion cohorts contrary to what we know. The table below is a reproduction of Peterson's Table 12 containing the aggregate statistic supporting a finding of no discrimination.

The table of results supports a finding of gender discrimination; however, we know that neither the simplistic specification discussed above (nor the full specification, for that matter) suggests statistically significant evidence of gender pay disparity in unions. Clearly, the rank tests cannot account for omitted variables in this instance.

VII. Concluding Comments

Rank sum tests do not hold up under scrutiny to the generality claimed by Peterson; they can be equivalent to a multiple regression analysis but only under very special conditions. The special conditions are unlikely to arise in a context of gender discrimination litigation; in fact, the opposite is most likely true. Incomplete data sets are likely to be the norm in litigation. Peterson's method may not be robust to models with covariance between regressors or the possibility of interactions with different distributions representing the pay generation mechanisms of men and women, as the case may be.

There may be a place in the forensic economist's tool-kit for Peterson's approach: its simplicity is its strongest recommendation. But the technique's simplicity is beguiling because its use demands strong assumptions. Thus, before turning to Peterson's approach—when applicable—an expert should continue to rely on sound multivariate regression analysis.

Table 2

		NORMALIZED			
Cohorts	**Rank Sum**	**Rank Sum**	**Mean**	**Variance**	**Standard Deviations**
union	1082	11.15	25973.58	159500.971	-65.01
non-union	41969	95.60	25973.58	94516222.70	-2.66
summary		106.76	51947.15	94675723.7	-5.33

References

Berndt, Ernst, *The Practice of Econometrics*, New York: Addison-Wesley, 1991.

Bernstein, Jared, and Lawrence Mishel, "Has Wage Inequality Stopped Growing?," *Monthly Labor Review*, December 1997, 3-16.

Glenn, Norval D., *Cohort Analysis*, Thousand Oaks, CA: Sage Publications, 2005.

Kennedy, Peter, "Randomization Tests in Econometrics," *Journal of Business & Economic Statistics*, January 1995, 13(1), 85-94.

Peterson, David W., "Cohort Analysis: A Regression Plain and Fancy," *Journal of Forensic Economics*, 2003, 16(2), 153-176.

Piette, Michael J., and Janet Thornton, "Cohort Analysis and the Determination of Economic Damages Resulting from Employment Discrimination," *Journal of Forensic Economics*, 1995, 8(1), 69-80.

Rodgers, James D., Michael L. Brookshire, and Robert J. Thornton, "Forecasting Earnings Using Age-Earnings Profiles and Longitudinal Data," *Journal of Forensic Economics*, 1996, 9(2), 169-210.