

## **Detecting Neglected Parameter Heterogeneity with Chow Tests<sup>†</sup>**

Joachim Zietz \*

*Middle Tennessee State University, Murfreesboro, TN*

### *Abstract*

The paper demonstrates through a number of Monte-Carlo experiments that, for the type of cross-section data sets typically encountered in applied economics, Chow tests on sorted variations of the data matrix can detect neglected parameter heterogeneity. The paper focuses on heterogeneity in the behavioral responses of economic actors that belong to different economically meaningful groups, such as the young, middle-aged, and old. Since the suggested methodology is easy to implement yet powerful, its routine use by applied economists would be desirable given the very significant estimation bias that can result from neglecting parameter heterogeneity.

*Key words:* parameter heterogeneity, Chow test, cross-section data, Monte-Carlo study

*JEL category:* C21, C52

<sup>†</sup> A revised version has been published in *Applied Economics Letters*, Vol. 13, No. 6 (May 2006): 369-374..

\*Joachim Zietz, Professor, Department of Economics and Finance, Middle Tennessee State University, Murfreesboro, TN 37132, phone: 615-898-5619, email: [jzietz@mtsu.edu](mailto:jzietz@mtsu.edu), url: [www.mtsu.edu/~jzietz](http://www.mtsu.edu/~jzietz).

# 1 Introduction

Tests for parameter heterogeneity are rarely used in econometric applications despite a considerable amount of work in the econometric literature following White's (1982) seminal study on the information matrix (IM) test and its reinterpretation by Chesher (1984) as a test for neglected parameter heterogeneity.<sup>1</sup> One reason for the lack of applications may lie in the typically poor power characteristics of the IM-derived test statistics. Another likely reason is that these tests are relatively cumbersome to use. The combination of these facts may make neglected parameter heterogeneity one of the more under-tested problems for cross-section data. This is unfortunate because, as shown by Zietz (2001), neglected parameter heterogeneity can be a serious problem for applied regression analysis in economics because it significantly biases the estimated coefficients, even to the extent of making them economically meaningless.

The purpose of this paper is to show why and how the well-known Chow test for structural break can be usefully applied to identify cases of neglected parameter heterogeneity in pure cross-section regressions.<sup>2</sup> By suggesting that a well-understood and easily implemented test can be helpful in detecting neglected parameter heterogeneity in cross-section data sets it is hoped that tests for parameter heterogeneity will stimulate research along a dimension that has become commonplace in panel data applications.

The paper is organized as follows. The next section discusses what is meant with parameter heterogeneity in this paper and explains why Chow tests are able to detect most practical forms of parameter heterogeneity that one may encounter in cross-section data sets. The following section provides some Monte-Carlo evidence on the power of Chow tests to detect parameter heterogeneity in the linear regression model. The paper concludes with a brief summary.

## 2 The Test Procedure

For pure cross-section data, one can differentiate between (a) random parameter variation across individual observations, as considered for example by Breusch and Pagan (1979) and Chesher (1984), and (b) systematic group-wise parameter variation, where the parameters vary across but not within a small number of subsets of observations.<sup>3</sup>

For a sample of  $n$  observations, random parameter variation for coefficient  $\beta$  can be described as

$$y_i = \alpha + \beta_i x_i + \varepsilon_i, \quad \text{where } \beta_i = \beta + \zeta_i \quad (1)$$

$$\zeta_i \sim N(0, \sigma_\zeta^2), \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2), \quad E(\zeta_i, \varepsilon_i) = 0, \quad \text{for } i = 1, \dots, n$$

By contrast, parameter heterogeneity by subsample implies that the sample consists of a small number of  $h$  mutually disjoint subsamples  $\{n_1, \dots, n_j, \dots, n_h\}$  with  $n = \sum_{j=1}^h n_j$ . For simplicity, it is assumed that each subsample  $j$  has associated with it a unique parameter  $\beta_j$ , that is fixed for all  $n_j$  observations of subsample  $j$ . In other words, there is parameter variation across subsamples but not within.

Two assumptions are distinguished for the distribution of  $x$ . These are critical for the ability of the Chow test to detect parameter heterogeneity. If  $x_j \sim iid(\bar{x}, \sigma_j^2)$ , that is, if the mean of  $x$  is identical across all  $h$  subsamples and the observations are not provided to the researcher already sorted by subsample, then Chow tests cannot identify underlying parameter heterogeneity. However, if  $x_j \sim iid(\bar{x}_j, \sigma_j^2)$ , that is, if the mean of  $x$  varies by subsample  $j$ , then parameter heterogeneity can in principle be detected by a Chow test after appropriately sorting the data set.

To make the Chow test a useful tool to detect neglected parameter heterogeneity one has to assume that the latter case, where the mean of  $x$  varies by subsample, is much more common in practical applications in economics than the case where  $x$  has the same mean across all subsamples. Recent work by Brock and Durlauf (2001) on social interactions as determinants of economic behavior provides some theoretical foundation for the assumption that parameter variation in economics is systematic rather than random. The additional assumption that economically relevant behavioral differences are typically associated with differences in the moments of at least some of the observed variables is rather trivial.<sup>4</sup>

The suggested procedure for detecting neglected parameter heterogeneity consists of a sequence of independent Chow tests.<sup>5</sup> For each test, the data matrix is sorted on a different regressor variable, with the exception of the dummy variables. The Chow test is conducted by splitting the sample into two.<sup>6</sup> For the standard assumptions of least squares and cross-section data sets, sorting by any of the right-hand variables is immaterial for residual-based test statistics. If any of the Chow tests is statistically significant, neglected parameter heterogeneity may be the reason, especially if other statistical indicators of neglected parameter heterogeneity are also present, such as heteroskedasticity and non-normal residuals. Since neglected parameter heterogeneity is invariably associated with heteroskedasticity,<sup>7</sup> Chow tests that are insensitive to heteroskedasticity, as discussed for example in Thursby (1992), should typically be employed, although standard unadjusted Chow tests may also work in practice.

Since every Chow test can be conducted in dummy variable format (e.g., Ashley 1984), there is an alternative to sorting the data matrix on each of the non-dummy regressors and then conducting a Chow test. To illustrate, consider the example where a linear regression equation with dependent variable  $y$  consists of only two regressors outside of the constant term,  $x$  and  $z$ . Then, the test methodology suggested so far would consist of

(a) sorting the data by  $x$  and (b) conducting a Chow test. The procedure would be repeated for variable  $z$ . Rather than conducting these two independent Chow tests, one could estimate two auxiliary test equations,

$$y = \alpha_0 + a_1x + a_2z + \beta_0D_x + \beta_1D_x x + \beta_2D_x z + \varepsilon \quad (2)$$

$$y = \delta_0 + \delta_1x + \delta_2z + \gamma_0D_z + \gamma_1D_z x + \gamma_2D_z z + \varepsilon \quad (3)$$

where  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ , and where dummy variable  $D_x$  is unity for all values of  $x$  that are equal to or larger than the median value of  $x$  and where dummy variable  $D_z$  is unity for all values of  $z$  that are equal to or larger than the median of  $z$ . An F-test of the joint significance of  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  would be equivalent to the suggested Chow test for a regression of  $y$  on a constant,  $x$ , and  $z$  if the data are sorted by variable  $x$ . An F-test of the joint significance of  $\gamma_0$ ,  $\gamma_1$ , and  $\gamma_2$  would be equivalent to the suggested Chow test for a regression of  $y$  on a constant,  $x$ , and  $z$  if the data are sorted by variable  $z$ . Employing equations (2) and (3) to conduct the Chow tests has the advantage that the data do not need to be sorted. On the negative side, the dummy variable version of the Chow test doubles the number of regressors, which may cause a degrees of freedom problem for smaller data sets.<sup>8</sup> Ultimately, it is the choice of the applied researcher which form of the Chow test will be employed.

### 3 The Power Characteristics of the Test Procedure

The purpose of the Monte-Carlo experiments summarized in Tables 1 through 3 is to determine how powerful Chow tests are in identifying neglected parameter heterogeneity of the type described above (a) in small, medium, and large samples, (b) for different degrees of variation of the mean of the regressor variables across subsamples, (c) for different degrees of parameter heterogeneity across subsamples, and (d) for different numbers of regressor variables. The experiments also reveal potential differences in standard and heteroskedasticity-resistant Chow tests.

All Monte-Carlo experiments assume that the sample consists of three subsamples of equal size, each with a unique set of parameters that is defined over two regressor variables,  $Z$  and  $W$ , and the constant term. The mean of variable  $Z$  is assumed to vary across the three subsamples, while variable  $W$  is assumed to have the same mean throughout. To analyze the sensitivity of the results to sample size, three different sample sizes are examined. The small sample consists of 60 observations, with twenty observations per subsample. The medium sample has 150 observations, with 50 observations in each subsample, and the large sample consists of 900 observations, which are equally divided among three subsamples. All experiments are conducted with 10,000 replications and rejection rates for several statistical tests are recorded for size 0.05.

Table 1 provides Monte-Carlo results for the base case, where the coefficients of the two variables and the constant term vary across subsample from five to twenty percent and the mean of variable  $Z$  varies by about 67 percent across subsamples. Rejection rates are

recorded for White's (1980) test for heteroskedasticity, the normality test of Jarque and Bera (1987), and two Chow tests. The first Chow test is the one originally suggested by Chow (1960), the other is a heteroskedasticity resistant variation of the Chow test that is easy to calculate yet received good marks in a Monte-Carlo study by Thursby (1992).<sup>9</sup> Table 1 reveals that Chow tests are unable to detect neglected parameter heterogeneity when the sample is sorted by a random number or by a variable ( $W$ ) with constant mean across all subsamples. A different result emerges when the data are sorted by the variable ( $Z$ ) with varying mean across subsamples: parameter heterogeneity is easily detected for the base case. Heteroskedasticity and non-normality tests have little power in identifying neglected parameter heterogeneity in small samples. Their power increases as the sample size goes up. Since the information matrix test has been shown by Hall (1987) to be equivalent to the sum of tests for heteroskedasticity and non-normality for the linear least squares regression case, one can conclude from Table 1 that it will easily detect neglected parameter heterogeneity in large samples but that its power decreases with the sample size.

Table 2 examines the sensitivity of the base case (Table 1) when (a) the mean value of  $Z$  and (b) the true coefficient of  $Z$  vary across the three subsamples. The sample is sorted only by variable  $Z$ . It is apparent that the variation of the mean value of  $Z$  across subsamples is far more important than the variation in the coefficient associated with  $Z$ . In fact, even in the case where the coefficients of  $Z$  do not vary at all across subsamples, do the Chow tests detect the parameter variation that results from the intercept term and variable  $W$ .

Table 3 documents to what extent the results are affected by the addition of more variables to the data generating model. Again, the data are sorted only by variable  $Z$ . If ten more variables like  $W$  are added, with means that do not vary across the sample, the sensitivity of the Chow tests to detect neglected parameter heterogeneity deteriorates somewhat compared to the results depicted in the first part of Table 2. In particular, the degree to which the mean of variable  $Z$  has to vary over the subsamples has to be somewhat more pronounced in Table 3 to achieve the same rejection rates than those provided in the first part of Table 2. If ten more variables like  $Z$  are added, for which the means vary across subsamples, the results depicted in the first part of Table 2 are virtually unaffected.

For all experiments reported in Tables 1 through 3, the heteroskedasticity-resistant Chow test performs somewhat better than the regular Chow test. But the difference in performance is clearly not large enough to dismiss the results produced by the regular Chow test.

## 4 Conclusion

Based on a set of Monte-Carlo experiments, the paper has demonstrated that Chow tests can be helpful in detecting neglected parameter heterogeneity for least squares regressions on pure cross-section data sets. Parameter heterogeneity is defined in this context as systematic as opposed to purely random variation in the regression coefficients

across but not within a small number of subsamples. It is suggested that this type of parameter heterogeneity is rather realistic for economic data sets. All that is needed to detect it with Chow tests is to sort the data on each of the regressors that are not dummy variables. Standard Chow tests have very similar power characteristics as heteroskedasticity-resistant Chow tests. This makes a strong case for systematically testing for neglected parameter heterogeneity since it involves little extra effort. For small and medium size samples, the power of the Chow tests to detect neglected parameter heterogeneity is superior to tests for heteroskedasticity and non-normality, which play a key role in tests for neglected parameter heterogeneity on the basis of the information matrix test.

## References

- Ashley, R. (1984) A simple test for regression parameter instability, *Economic Inquiry*, 22, 253-68
- Breusch, T.S. and Pagan, A.R. (1979) A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47, 1287-94.
- Brock, W.A. and Durlauf, S.N. (2001) Discrete choice with social interactions, *Review of Economic Studies*, 68, 235-60.
- Chesher, A. (1984) Testing for neglected heterogeneity, *Econometrica*, 52, 865-72.
- Chesher, A. and Spady, R. (1991) Asymptotic expansions of the information matrix test statistic, *Econometrica*, 59, 787-815.
- Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, 28, 591-605.
- Davidson, R. and MacKinnon, J.G. (1992) A new form of the information matrix test, *Econometrica*, 60, 145-57.
- Godfrey, L.G. and Orme, C.D. (1994) The sensitivity of some general checks to omitted variables in the linear model. *International Economic Review*, 35, 489-506.
- Hall, A. (1987) The information matrix test for the linear model, *Review of Economic Studies*, 54, 257-63.
- Jarque, C. M. and Bera, A. K. (1987) A test for normality of observations and regression residuals, *International Statistical Review*, 55, 163-72.
- Thursby, J.G. (1992) A comparison of several exact and approximate tests for structural shift under heteroscedasticity, *Journal of Econometrics*, 53, 363-86.

White, H. (1980) A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 48, 817-38.

White, H. (1982) Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-26.

Zietz, J. (2001) Heteroskedasticity and neglected parameter heterogeneity, *Oxford Bulletin of Economics and Statistics*, 63, 269-79.

## **Acknowledgements**

The author would like to thank Jonathan Temple for numerous helpful comments. This research was partially funded by a summer research grant of the faculty research and creative activity program at Middle Tennessee State University.

## **Endnotes**

<sup>1</sup>Chesher and Spady (1991), Davidson and MacKinnon (1992), and Godfrey and Orme (1994) are some of the more recent studies.

<sup>2</sup>A significant amount of work on parameter heterogeneity exists in the literature dealing with panel data. Tests for varying parameters are also routinely applied in time series applications.

<sup>3</sup>Group-wise parameter heterogeneity could arise, for example, across the saving behavior of the young, middle-aged, and old, or across firms belonging to the manufacturing as opposed to the service sector.

<sup>4</sup>For example, the coefficients describing firm's R&D expenditures may vary by size category or saving behavior may be different by age group.

<sup>5</sup>Since the Chow tests are not used for a constructive specification search, the tests are independent and require no adjustment of significance levels.

<sup>6</sup>A more elaborate test strategy could involve splitting the sample at different points, such as, the 33rd quantile, the 50th quantile, and the 67th quantile, as measured by the marginal distribution of the regressor variable that is used to sort the data matrix.

<sup>7</sup>Zietz (2001) has recently re-emphasized the link between heteroskedasticity and parameter heterogeneity.

<sup>8</sup>If one decided to divide the data set into three rather than two segments for the Chow test, the number of variables would triple; and so forth.

<sup>9</sup>The calculations use the *gmac2* version as made available on the TSP web site: <http://www.stanford.edu/~clint/tspex/gmac2.tsp>.



**TABLE 1. MONTE CARLO RESULTS ON POWER OF CHOW TEST TO DETECT PARAMETER HETEROGENEITY**

**Data Generating Process**

$$Y_j = \alpha_j + \beta_j Z_j + \gamma_j W + \varepsilon \quad \text{for } j = 1, 2, 3$$

where

$$Z_j \sim N(\mu_j, 4), W \sim N(30, 4), \varepsilon \sim N(0, 1)$$

$$\alpha_j = \{100, 97, 95\}, \beta_j = \{1.2, 1.05, 1.0\}, \gamma_j = \{2.0, 2.05, 2.2\}, \mu_j = \{40, 32, 24\}$$

**Results for 10,000 Replications, Estimated Model:  $Y = a + bZ + cW$**

<i>H0 Tested</i>	<i>Type of Test</i>	Rejection Rates of H0 for Size 0.05 if Sample is Sorted		
		randomly	by <i>Z</i>	by <i>W</i>
Small Sample, $n_j = 20$ , for $j = 1,2,3$ ; $n = 60$				
Homoskedasticity	White	0.4576	0.4473	0.4519
Normality of Residuals	Jarque-Bera	0.4223	0.4252	0.4228
No Structural Break	Chow	0.0364	1.0000	0.0411
	Mac2	0.0359	1.0000	0.0434
Medium Sample, $n_j = 50$ , for $j = 1,2,3$ ; $n = 150$				
Homoskedasticity	White	0.9538	0.9530	0.9561
Normality of Residuals	Jarque-Bera	0.9986	0.9971	0.9977
No Structural Break	Chow	0.0362	1.0000	0.0563
	Mac2	0.0358	1.0000	0.0586
Large Sample, $n_j = 300$ , for $j = 1,2,3$ ; $n = 900$				
Homoskedasticity	White	1.0000	1.0000	1.0000
Normality of Residuals	Jarque-Bera	1.0000	1.0000	1.0000
No Structural Break	Chow	0.0394	1.0000	0.1851
	Mac2	0.0393	1.0000	0.1853

*Notes:*  $n_j$  identifies the number of observations per subset  $j$ . Homoskedasticity is tested with White's (1980) test, normality by the procedure suggested by Jarque and Bera (1987). The null of no structural break is tested using a regular Chow test and the *Mac2* test suggested by Thursby (1992), which is insensitive to heteroskedasticity.

**TABLE 2. SENSITIVITY ANALYSIS WITH RESPECT TO Z AND ITS COEFFICIENT VECTOR**

**Data Generating Process**

$$Y_j = \alpha_j + \beta_j Z_j + \gamma_j W + \varepsilon \quad \text{for } j = 1, 2, 3$$

Where

$$Z_j \sim N(\mu_j, 4), W \sim N(30, 4), \varepsilon \sim N(0, 1)$$

$$\alpha_j = \{100, 97, 95\}, \gamma_j = \{2.0, 2.05, 2.2\}$$

**Results for 10,000 Replications, Estimated Model:  $Y = a + b Z + c W$**

<i>Experiment</i>	Rejection Rates of Chow / Mac2 Tests for Size 0.05 if Sample is Sorted by Z		
	$n_j = 20, \text{ for } j = 1,2,3;$ $n = 60$	$n_j = 50, \text{ for } j = 1,2,3;$ $n = 150$	$n_j = 300, \text{ for } j = 1,2,3;$ $n = 900$
<i>Mean values of <math>Z_j</math> vary, <math>\beta_j = \{1.2, 1.05, 1.0\}</math></i>			
$\mu_j = (36, 30, 24)$	1.0000 / 1.0000	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (35, 30, 25)$	0.9991 / 0.9997	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (34, 30, 26)$	0.9543 / 0.9824	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (33, 30, 27)$	0.6678 / 0.8011	0.9895 / 0.9958	1.0000 / 1.0000
$\mu_j = (32, 30, 28)$	0.2326 / 0.3363	0.5751 / 0.6908	1.0000 / 1.0000
$\mu_j = (31, 30, 29)$	0.0672 / 0.0799	0.1068 / 0.1224	0.4459 / 0.4926
<i>Coefficients of <math>Z_j</math> vary, <math>\mu_j = \{35, 30, 25\}</math></i>			
$\beta_j = \{1.35, 1.05, 0.95\}$	0.9988 / 0.9998	1.0000 / 1.0000	1.0000 / 1.0000
$\beta_j = \{1.29, 1.05, 0.97\}$	0.9987 / 0.9997	1.0000 / 1.0000	1.0000 / 1.0000
$\beta_j = \{1.23, 1.05, 0.99\}$	0.9988 / 0.9997	1.0000 / 1.0000	1.0000 / 1.0000
$\beta_j = \{1.20, 1.05, 1.00\}$	0.9989 / 0.9998	1.0000 / 1.0000	1.0000 / 1.0000
$\beta_j = \{1.14, 1.05, 1.02\}$	0.9985 / 0.9990	1.0000 / 1.0000	1.0000 / 1.0000
$\beta_j = \{1.08, 1.05, 1.04\}$	0.9869 / 0.9878	1.0000 / 1.0000	1.0000 / 1.0000
$\beta_j = \{1.05, 1.05, 1.05\}$	0.9310 / 0.9392	1.0000 / 1.0000	1.0000 / 1.0000

*Notes:*  $n_j$  identifies the number of observations per subset  $j$ . The null of no structural break is tested using a regular Chow test and the *Mac2* test suggested by Thursby (1992), which is insensitive to heteroskedasticity..

**TABLE 3. SENSITIVITY ANALYSIS WITH RESPECT TO ADDITIONAL VARIABLES**

**Data Generating Process**

$$Y_j = \alpha_j + \beta_j Z_j + \gamma_j W + \sum_{k=1}^{10} (-1)^{k+1} \delta_{jk} X_{jk} + \varepsilon \quad \text{for } j = 1, 2, 3$$

Where

$$Z_j \sim N(\mu_j, 4), W \sim N(30, 4), \varepsilon \sim N(0, 1)$$

$$\alpha_j = \{100, 97, 95\}, \beta_j = \{1.2, 1.05, 1.0\}, \gamma_j = \{2.0, 2.05, 2.2\}, \delta_{jk} = \{\beta_j + (10 + k)/10\}$$

**Results for 10,000 Replications, Estimated Model:  $Y = a + bZ + cW$**

Experiment	Rejection Rates of Chow / Mac2 Tests for Size 0.05 if Sample is Sorted by Z		
	$n_j = 20, \text{ for } j = 1,2,3;$ $n = 60$	$n_j = 50, \text{ for } j = 1,2,3;$ $n = 150$	$n_j = 300, \text{ for } j = 1,2,3;$ $n = 900$
<i>Addition of 10 Variables like W, <math>X_{jk} \sim N(30, [2+0.1(k-1)]^2)</math>,</i>			
$\mu_j = (42, 30, 18)$	1.0000 / 1.0000	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (40, 30, 20)$	0.9990 / 1.0000	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (38, 30, 22)$	0.9725 / 0.9999	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (36, 30, 24)$	0.7524 / 0.9834	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (34, 30, 26)$	0.3401 / 0.7064	0.9150 / 0.9870	1.0000 / 1.0000
$\mu_j = (32, 30, 28)$	0.0794 / 0.1142	0.1520 / 0.2080	0.9048 / 0.9496
$\mu_j = (31, 30, 29)$	0.0521 / 0.0572	0.0670 / 0.0740	0.1331 / 0.1422
<i>Addition of 10 Variables like Z<sub>j</sub>, <math>X_{jk} \sim N(\mu_j, [2+0.1(k-1)]^2)</math>,</i>			
$\mu_j = (36, 30, 24)$	1.0000 / 1.0000	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (35, 30, 25)$	0.9999 / 0.9999	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (34, 30, 26)$	0.9867 / 0.9902	1.0000 / 1.0000	1.0000 / 1.0000
$\mu_j = (33, 30, 27)$	0.7748 / 0.8024	0.9995 / 0.9996	1.0000 / 1.0000
$\mu_j = (32, 30, 28)$	0.2366 / 0.2702	0.7048 / 0.7213	1.0000 / 1.0000
$\mu_j = (31, 30, 29)$	0.0598 / 0.0671	0.0877 / 0.0900	0.4694 / 0.4754

Notes:  $n_j$  identifies the number of observations per subset  $j$ . The null of no structural break is tested using a regular Chow test and the *Mac2* test suggested by Thursby (1992), which is insensitive to heteroskedasticity.