Regional wage differentials - does distance matter?

Morten Marott Larsen.

University of Copenhagen, Institute of Economics & AKF (Institute of Local Government Studies - Denmark), Nyropsgade 37, DK-1602 Copenhagen V, Denmark (e-mail: ml@akf.dk)

September 2002

Abstract

This paper uses econometric methods to analyse causes of regional wage differentials in Denmark for the period1996-1999 and quantifies the importance of spatial proximity. Grouped data sets are used. Two concepts of distance are investigated. The first assumes that there is a positive production externality present in a centre that declines as the distance to the centre increases. Different definitions of a centre are tested. The second distance concept is commuting distance. Both distance measures have the expected sign and are significant, but the effects are small.

JEL classification: J31

Key words: Wage differentials, distance, grouped data.

1 Introduction

This paper examines regional wage differentials in Denmark during the years 1996-1999, and in particular, the wage effect of two kinds of distance: 1) from place of work to an urban centre and 2) from place of residence to the place of work. The importance of these effects is tested using econometric methods.

When evaluating the effect of distance it is essential to define distance is. As noted above two different approaches are dealt with. *The first hypothesis is that there is a positive externality associated with a centre*. Here the distances are from place of work to a centre. The externality is on the production side of the economy and producers are able to pay higher wages because the workers are more productive in (and near) the centre. Firms located near the centre also benefit from the externality. It is assumed that the positive externality decreases as distance to the centre increases. Another important question within this first hypothesis is to define a centre. This paper tests three different measures: Distance to commuting centres, distance to the capital Copenhagen, and distance to the five large university towns: Copenhagen, Århus, Odense, Aalborg, and Esbjerg. The theoretical point of departure of the first hypothesis is production theory. From the producers' point of view workers are hired up to the point where the marginal product of the labour equals the wage. The observed wage can thus be interpreted as marginal productivity.

The second hypothesis is that commuting distance affects wages. Commuting distances are from place of residence to place of work. They can be interpreted as a workplace disamenity and therefore workers should require a higher wage if they have a longer commuting distance. The theoretical point of departure of the second hypothesis is compensated wage differentials[1]. In The Wealth of Nations Adam Smith argued that wages paid to workers should compensate for or equalize differences in workplace amenities and disamenities. Both monetary differences and also non-monetary advantages are reflected in the wage.

The wage effects of commuting distances may be underestimated in this paper. One reason is that workers in Denmark cab obtain a tax deduction for commuting if the worker commutes more than 24 kilometres per day, a fact which is not included in the estimation here.

Another matter when describing regional wage differentials is price differences across regions. Housing costs are, for instance, often higher in the centres which may equalize real wage across regions. For example, Dumond et all (1999) describe the importance of cost of living indexes in the USA. The subject is not explicitly addressed in this paper, but it is worth remembering when the results of the estimation are interpreted.

It is also important to describe what other factors determine the wage, besides distance. Factors such as gender, age, education, and sector, are often significant. A lot of research confirms positive relationships between wage and age, education and the male sex, see for example Albæk et al. (1999), Trigg and Madden (1995), and Berndt (1991). If one sector is preferred to another it could be interpreted as workplace disamenities within the theoretical setup of this paper. Normally, wages in the private sector are higher, see for example Berndt (1991), indicating some kinds of workplace disamenities in the private sector.

[1]Brown [1980].

The paper is organized as follows. Section 2 describes the data and section 3 presents summary statistics of the data. Section 4 deals with econometric questions such as estimation with grouped data. In section 5 the results regarding wage differentials and distance as a positive externality are presented. Section 6 concerns wage differentials and commuting distance. Finally, the conclusion is in section 7.

2 The data

The Social Accounting Matrix for Danish Municipalities (SAM-K) is the main data source in this paper, and a complete description of the data is presented in Madsen et al. (2001). The SAM is based upon the entire Danish population, but the data used in this paper are grouped.

Two main raw data sets are used. The SAM-K data sets are constructed from data on individuals, but are grouped together on the basis of variables that are named axes. *Data set 1* focuses on the production side of the regional economy. There are different types of employment and wage measures, but for estimation purposes “a full year's work for one person” and “wage arising from the employment” are chosen to represent “employment” and “wage” in this paper. It could be possible to include some profits of private owned firms and pensions, but these are presumed to be unimportant. Data set 1 has the following axes: Year (1996-99), place of work (275 municipalities), sector (132 groups), education (5 groups), gender, and age (0-14, 15-29, 30-44...).

The following observations are eliminated from the sample: If place of work is abroad, not available, or on the small island of Christiansø, and if age is below 15 or above 59. Furthermore, some extreme observations are removed, if the average wage is above 2 mill. DKK, and if total employment is under 5 working days the year concerned. This all reduces the number of observations to 908,947 in the 4-year-period with an average of 9.0 full year's workers per observation and a standard deviation of about 35 full year's workers. About 1 per cent of the observations contains more than 100 workers. The highest number of workers in an observation is around 2,200.

Data set 2 contains the same wage and employment measures as in data set 1. The differences are that no sector variable is present, but on the other hand place of residence is available at municipal level. Additionally, information in data set 2 about ongoing education is not used. The same sample selection criteria are applied as in data set 1. The total number of observations is 822,721 with an average of 9.9 full year's workers per observation and a larger standard deviation of about 89 full year's workers.

A variable capturing unemployment is also present in data set 2. A member of the labour force (15-59 years old) is unemployed if he is unemployed at the end of November when the statistics are collected.

Different measures of distances are applied in the regressions, but the source is the same. Distances are at municipal level. The distance from one municipality to another is the number of kilometres from the main post office in the first municipality to the main post office in the second municipality. An internal distance inside the municipality is calculated including factors such as size and shape of the municipality. When crossing water, kilometres are not an appropriate measure because there would usally be more cost connected with crossing water.

Therefore the kilometres are transformed into DKK with the assumption that one kilometre on land equals one DKK. When crossing water the price of a ferry ticket is applied instead of kilometres. The distances are from 1996.

The distances could be defined in other ways taking factors into account such as congestion, speed limits, time values, etc., but more accurate distances will not be pursued further in this paper. Further research could apply the actual commuting distances of the single worker and not just an average distance at municipal level.

In 1997 for rail and in 1998 for cars and lorries, The Great Belt link between Zealand and Funen opened and motorways in Northern Jutland were also finished. However, it has not been possible to include these substantial changes in the distance variable used in this paper.

The distance measure in this paper is not accurate because the distances are defined at municipal level. A coefficient on a badly measured variable is biased toward zero[2]. Therefore, it should be expected that a more accurate distance measure would lead to more significant results than in this paper.

3 Summary statistics

To give an overview of wage differentials in Denmark some preliminary statistics are presented.

Typically, the wage effects of gender and education have different interpretations. The wage

[2]See for example Hausman (2001).

differentials of education are normally related to productivity, but wage differentials of gender are normally described to discrimination.

Tabel 3.1. Mean yearly wage 1999 in DKK (number of full year's workers in brackets)

	Male	Female	Total
Basic education	246,147 (386,628.62)	209,736 (327,467.55)	229,450 (714,096.17)
Skilled worker	283,904 (492,501.47)	226,151 (357,866.84)	259,599 (850,368.32)
Short further and higher education.	318,280 (54,396.94)	255,663 (48,717.68)	288,696 (103,114.62)
Middle-range further and higher education.	374,491 (107,542.18)	262,664 (173,176.57)	305,505 (280,718.75)
Long further and higher education.	441,082 (85,897.68)	339,669 (52,876.49)	402,441 (138,774.18)
Total	293,234 (1,126,966.90)	234,888 (960,105.13)	266,393 (2,087,072.03)

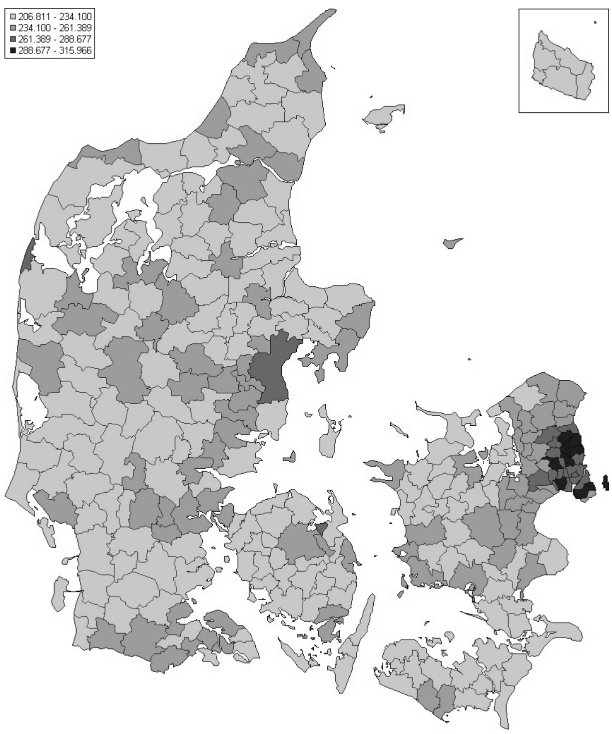
Source: AKF’s regional accounting system.

Table 3.1 indicates that the average wage for a male increases faster with education level, but on average males must also forgo a higher wage to attend education. That is, the opportunity cost of education is higher for males.

Wages per employed worker increases with age. The youngest age group, 15-29 years old, earns nearly 200,000 DKK per year, whereas a person in the older age groups, 30-44 years old and 45-59 years old, are earning around 265,000 DKK and 280,000 DKK respectively.

Wages per employed worker have a regional pattern. In figure 3.1 average wages are plotted for the municipalities in Denmark. The metropolitan area of Copenhagen stands out, but the larger cities of Århus, Aalborg, Odense, and perhaps also Esbjerg have an average wage above the average municipality. This also applies to the triangle of Vejle, Kolding, and Fredericia. The highest average wages are found in selected municipalities in the metropolitan areas and the lowest average wages are to be found on the island of Bornholm.

Figure 3.1. Average wages per employed by place of production



4 The regression model and estimation with grouped data

The following regression equation is used:

$$\ln\left(\overline{w}_{r,t}\right) = f\left(X_{r,t}\right) + u_{r,t}$$

where $\ln\left(\overline{w}_{r,t}\right)$ is the natural log of average wage, $X_{r,t}$ are the explanatory variables, and $u_{r,t}$ is the random disturbance which is assumed to be normally distributed with mean zero and constant variance. The subscript *r* is region (municipality) and t is time (year). The explanatory variables can include constant terms, individual characteristics such as education and gender, and regional characteristics such as regional unemployment.

The functional form of the wage is often assumed to be logarithmic because data are skewed with the median wage usually being less than the mean, this being also the case with the present data. Furthermore, regressions with average wages and logarithmic average wages are compared, the logarithmic average wage regressions provides a better fit to the data.

There are differences when applying the logarithmic form with grouped data as compared to individual data. The logarithm of the expected value of average wage does not equal the expected value of the logarithm of average wage. This paper does not deal with this subject, but the issue is recognized.

As mentioned, the data used in the regression are grouped. A common practice is to weight the grouped data by the square root of the group size. But as pointed out by Dickens (1990) it may be inappropriate to use the square root of group size because of within group correlation of the error term.

The question is, how much wages within a group vary. If there is no within group correlation, then the square root of the group size would be the appropriate method, but the groups in the data have several common characteristics. Beforehand, it could not be rejected that within group correlation of the error term is present.

Dickens (1990) describes a test whether or not weighting with the square root of the group size is appropriate or not. A regression where the data are weighted with the square root of the group size can be run, and the residuals can be tested for heteroscedasticity. For instance, one can regress the squared residuals from the weighted regression on a constant and the size of the group. If the coefficient of group size is significantly different from zero, weighting with the square root is not appropriate.

Three types of estimators are tested in the regressions. Besides ordinary least square (OLS) and weighted least square (WLS, weight the data by the square root of the group size) also feasible generalised least square (FGLS) is carried out. The weights in FGLS are formed in the following manner. OLS is run and the log of the squared residuals is regressed on the log of the group size. The exponential of the predicted values from this regression are used as weights in the FGLS regression.

The sizes of the standard errors and the above mentioned test from Dickens (1990) are used to choose between the three estimators.

5 Wage differentials by place of production and distance as a positive externality

This section focuses on the first hypothesis of this paper: *that there is a positive externality of a centre*. Data set 1 is used.

First, a model with gender, education, sector, and unemployment as explanatory variables is set up and compared with a model which also includes municipalities as fixed effects. An F-test[3] cannot reject that the constant terms are all equal and therefore fixed effects are left out. In the fixed effects model with time dummies, unemployment is insignificant which corresponds to the results in Albæk et al. (1999).

Three measures of the centre are applied now in a model with gender, time, education, sector, and unemployment as explanatory variables. Fundamentally, more than one measure could enter into the model, but here it is assumed that there is only one type of centre in which one positive externality is present. The logarithmic transformation of all three measures describes the data better. Distance to the commuting centre is the worst measure comparing with the two others because the adjusted r^2 is smallest comparing the three models which only differ with respect to the distance measure. It might be because a definition of 35 commuting centres is chosen where some of the centres are small islands where no positive externality in production would be expected. The second best measure is the distance to Copenhagen. Given the geography of

[3]See appendix B, I).

Denmark one would expect that it could be difficult to identify some distances because of the many belts, seas, sounds, and straits.

Table A.1 in appendix A contains the variable “distance5" this being the best of the three proposed measures. “Distance5" is the distance to five university towns: Copenhagen, Århus, Odense, Aalborg, and Esbjerg. Whether or not to interpret it as distance to an economic centre or distance to a university city is a matter of choice.

The test described regarding heteroscedasticity[4] rejects the hypothesis that there is no heteroscedasticity in all three models with the explanatory variables of table A.1 in appendix A. The results of the FGLS regression are preferred to WLS and OLS because the t-value of β is numerically smaller in the heteroscedasticity test.

Because both the average wage and distances are in logarithmic form the estimated parameters are elasticities. However, the elasticity is small: -0.04. If an improvement in infrastructure could be interpreted as a shorter distance then a10% improvement in infrastructure would result in a 0.4% higher wage. In the context of the theoretical setup of this paper it is also a welfare gain because the higher wages are due to a positive externality.

Even though the elasticity is small the total welfare gain is worth calculating. In 1999 1.5 million workers had a place of work outside the 5 centres and their average wage was 260,000 DKK. If all distances outside the centres were improved by 10% there would on average per worker be

[4]See appendix C, I).

a welfare gain of 1040 DKK. The total welfare gain would be 1.5 million workers times 1040 DKK; a total of 1.560 billion DKK.

When dealing with infrastructure investment the positive externality associated with a centre is not the only benefit. Therefore the total welfare gain of the positive externality could be worth calculating as one of the benefits.

When comparing the size of the estimates in the regression the most important contributions to the average wages are gender, education, some sectors, and age. Distance contributes a little less than the difference between 1996 and 1998 and more than the natural log of unemployment. As mentioned the fixed effects model is rejected, which means that regional unemployment has a small but significant estimated parameter. An interpretation could be that higher unemployment lowers wages because of competition for vacancies.

6 Wage differentials and commuting distance

The second hypothesis of this paper is *that commuting distance affects wages*. To test this, a regression analysis on data set 2 is carried out.

We recall that data set 2 has information about both place of residence and place of work. Both place of residence and place of work could be used as a fixed effect, but when comparing adjusted r^2, place of work is chosen. However, an F-test[5] does not support the municipalities as fixed effects and because of that it is abandoned.

The same problems concerning grouped data are present in this regression as well. The test described rejects the hypothesis that there is no heteroscedasticity[6] in all three models with the explanatory variables of table A.2 in appendix A. Again, the results of the FGLS regression are preferred to WLS and OLS because the t-value of β is numerically smaller in the heteroscedasticity test.

Gender, education, age, year, unemployment by place of residence and commuting distances are all significant in the model using FGLS, WLS, and OLS. Estimates and standard deviations are presented in appendix A, table A.2. Comparing the estimated parameter with the regression using data set 1, the estimated parameters of gender have increased by 17% (using FGLS) and other changes have also occurred. The estimated parameter of unemployment is still small, though it has increased. An explanation could be that unemployment by place of residence is chosen and in the regression using data set 1 unemployment by place of work is used. Another reason could be simply that it is another data set.

Commuting distance has a positive effect on wages. The estimated parameter is around 0.03 (using FGLS). With an elasticity of 0.03 one would get 3% more in average wage if commuting distance doubles.

---

[5] See appendix B, II).

[6] See appendix C, II).

7 Conclusion

There is a positive externality present in a centre and the best measure of a centre is the distance to five larger university towns. The paper does not test whether or not it is because of the existence of a university.

Even though the elasticity of distance is small it could be worth measuring if one is to evaluate infrastructure investments, because the positive externality of the centre is only one of several benefits.

If your commuting distance is long you would on average get a higher wage. This is the result of the wage regression including commuting distances.

Of course one could argue about the direction of causality assumed in the regression analysis. It is implicitly assumed that commuting distances affect wages and not the other way around. It could be argued that if one has a higher wage one is able to locate further away and this is why workers with higher wages have higher commuting distances.

Distance does matter - perhaps not a lot - but regression analyses of wage differentials presented in this paper, do suggest that distance does make a positive contribution to explaining the variation in the data.

References

Albæk, Karsten, Rita Asplund, Stig Blomskog, Erling Barth, Björn Runar Gudmundsson, Vifill Karlsson, and Erik Strøjer Madsen (1999). Dimensions of the Wage-Unemployment Relationship in the Nordic Countries: Wage Flexibility without Wage Curves. Discussion Paper 99-24, University of Copenhagen, Institute of Economics.

Berndt, Ernst R. (1991). The Practice of Econometrics: Classic and Contemporary. Addison-Wesley Publishing Company, Inc.

Brown, Charles (1980). Equalizing Differences in the Labor Market. Quarterly Journal of Economics; 94:1, February, pages 113-34.

Dickens, William T. (1990). Error Component in Grouped Data: Is it Ever Worth Weighting? The Review of Economics and Statistics, Volume 72, Issue 2, May 1990, 328-333.

Dumond, Michael J., Barry T. Hirsch, and David A. Macpherson (1999). Wage Differentials Across Labour Markets and Workers: Does Cost of Living Matter?. Economic Inquiry, vol. 37, No. 4, October 1999, 577-598.

Hausman, Jerry (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. Journal of Economic Perspectives, vol. 15, No. 4.

Madsen, Bjarne, Chris Jensen-Butler and Poul Uffe Dam (2001). A Social Accounting Matrix for Danish Municipalities (SAM-K). AKF Forlaget, Copenhagen.

Trigg, Andrew B. and Moss Madden (1995). A CGE solution to the household rigidity problem in extended input-output models. In Social & Demographic Accounting, edited by Geoffrey J. D. Hewings & Moss Madden. Cambridge University Press.

Appendix A

Table A.1. Wage differentials by place of production. Estimated parameters and standard errors.

Dependent variable: ln(average wage)	FGLS	WLS	OLS
Intercept	12.04955 (0.00262)	12.07617 (0.00221)	12.0701 (0.00354)
Gender	0.17744 (0.00056923)	0.19182 (0.00042372)	0.15730 (0.00080864)
Year=1999	0.07181 (0.00086704)	0.06812 (0.00063209)	0.06648 (0.00124)
Year=1998	0.05261 (0.00083882)	0.05011 (0.00059852)	0.04885 (0.00121)
Year=1997	0.02129 (0.00078997)	0.01956 (0.00055930)	0.02116 (0.00114)
Long further and higher education	0.44529 (0.00116)	0.48476 (0.00089247)	0.38434 (0.00153)
Middle-range further and higher education.	0.32206 (0.00094904)	0.31477 (0.00069157)	0.29656 (0.00132)
Short further and hi. edu.	0.19079 (0.00102)	0.19089 (0.00084233)	0.16910 (0.00135)
Skilled worker	0.10002 (0.00065551)	0.10167 (0.00045153)	0.08367 (0.00098931)
Public and personal services	0.03522 (0.00158)	-0.00264 (0.00148)	0.08941 (0.00202)
Financial intermediation	0.19359 (0.00167)	0.19702 (0.00155)	0.19622 (0.00212)
Transport and communic.	0.09228 (0.00182)	0.11263 (0.00161)	0.08528 (0.00239)
Wholesale and retail trade	0.05271 (0.00164)	0.07839 (0.00150)	0.04573 (0.00213)
Construction	0.01555 (0.00207)	-0.00886 (0.00162)	0.03066 (0.00321)
Electricity, gas, & water.	0.15069 (0.00296)	0.10845 (0.00271)	0.17680 (0.00381)
Manufacturing	0.10046 (0.00158)	0.08810 (0.00148)	0.10430 (0.00201)
Age 45-59	0.26275 (0.00071966)	0.28820 (0.00053279)	0.22788 (0.00101)
Age 30-44	0.20693 (0.00070084)	0.22936 (0.00051082)	0.17211 (0.00099141)
ln(unemployment)	-0.02047(0.00081802)	-0.02993(0.00061965)	-0.02309 (0.00117)
ln(distance5)	-0.03893(0.00027692)	-0.04252(0.00017878)	-0.03165(0.00045022)
* Adjusted R^2	-	-	0.215

Note: All variables are significant at 10% level. If all dummy variables are zero the representation is: Year=1996, basic education, sector= Agriculture, and age 15-29 years old.

Table A.2. Wage differentials and commuting distances. Estimated parameters and standard errors.

Dependent variable: ln(average wage)	FGLS	WLS	OLS
Intercept	11.99731 (0.00198)	11.96282 (0.00120)	11.95378 (0.00310)
Gender	0.20739 (0.00058077)	0.20313 (0.00032394)	0.18550 (0.00089959)
Year=1999	0.04847 (0.00088965)	0.6484 (0.00050635)	0.04807 (0.00137)
Year=1998	0.03423 (0.00086104)	0.04703 (0.00048466)	0.03470 (0.00133)
Year=1997	0.01399 (0.00081145)	0.01898 (0.00045713)	0.01323 (0.00125)
Long further and higher education.	0.46169 (0.00109)	0.49915 (0.00071247)	0.45697 (0.00158)
Middle-range further and higher education.	0.29248 (0.00090276)	0.29033 (0.00055063)	0.30186 (0.00138)
Short further and higher education.	0.17716 (0.00101)	0.17132 (0.00068469)	0.19049 (0.00150)
Skilled worker	0.10374 (0.00070002)	0.09785 (0.00036837)	0.10002 (0.00111)
Age 45-59	0.30252 (0.00074048)	0.27874 (0.00042973)	0.31246 (0.00112)
Age 30-44	0.24254 (0.00071004)	0.22502 (0.00041411)	0.24134 (0.00106)
ln(unemployment)	-0.07263(0.00082992)	-0.03924(0.00047434)	-0.06847 (0.00130)
ln(distance)	0.03071 (0.00022535)	0.03067 (0.00016048)	0.04145 (0.00036696)
* Adjusted R^2	-	-	0.2626

Note: All variables are significant at 10% level. If all dummy variables are zero the representation is: Year=1996, basic education, and age 15-29 years old.

Appendix B. F-test for fixed effects

F-test: $$\text{F(n-1,nT-n-K)}=\frac{\left(\mathrm{R}_u^2-\mathrm{R}_p^2\right)/\left(n-1\right)}{\left(1-\mathrm{R}_u^2\right)/\left(nT-n-K\right)}$$

I) F-test in a model with wage differentials by place of production (data set 1) and explanatory variables gender, education, sector, age, and unemployment:

$$\text{F(274,810)}=\frac{\left(0.2214-0.2083\right)/\left(274\right)}{\left(1-0.2214\right)/\left(810\right)}\cong 0.05$$

As F(4,4)=1 the hypothesis that all constant terms are equal could not be rejected.

II) F-test in a model with wage differentials (data set 2) and explanatory variables gender, education, age, unemployment, and commuting distance:

$$\text{F(274,810)}=\frac{\left(0.2854-0.2613\right)/\left(274\right)}{\left(1-0.2854\right)/\left(816\right)}\cong 0.10$$

As F(4,4)=1 the hypothesis that all constant terms are equal could not be rejected.

Appendix C

I) A model with wage differentials by place of production (data set 1) and explanatory variables gender, time, education, sector, age, unemployment, and distance is examined. The squared residuals of the weighted regression (u^2) are estimated on the size of the group (N):

$$u^2 = \beta N + \varepsilon$$

The results are:

	FGLS	WLS	OLS
β	0.01863	0.0060891	-0.00079502
t-value of β	26.89	36.53	-40.9

All of the t-values are significant and therefore a group error component could be present.

II) A model with wage differentials and explanatory variables gender, time, education, age, unemployment, and distance is examined.

The test described above gives the following results:

	FGLS	WLS	OLS
β	0.00461	0.00788	-0.00016861
t-value of β	9.84	329.21	-14.96

All of the t-values are significant and therefore a group error component could be present.