



Communication

Is Industry Size a Carrier for Wage Inequality? A Panel Study Addressing Independent Variables of Inherently Different Sizes across Units

Jarle Aarstad * and Olav Andreas Kvitastein

Mohn Centre for Innovation and Regional Development, Western Norway University of Applied Sciences, P.O. Box 7030, NO-5020 Bergen, Norway; olav.andreas.kvitastein@hvl.no

* Correspondence: jarle.aarstad@hvl.no

Abstract: We address how independent variables of inherently different sizes across units, e.g., small vs. large industries, in panel regression is an advantage interpretively. Analyzing a Norwegian industry panel, we find that wage inequality is a function of industry size, particularly size increase, in an absolute number of firms. A possible reason is that specialized skilled employees negotiate higher wages when there are many legal entities. The findings can also imply that wage inequality is more sensitive to random change, particularly an increase, in large rather than small industries. We conclude that particularly large industries are positive carriers of wage inequality and discuss potential underlying causal mechanisms such as monopolistic competition.

Keywords: Gini coefficients; wage inequality; unit size; fixed effects regression



Citation: Aarstad, Jarle, and Olav Andreas Kvitastein. 2021. Is Industry Size a Carrier for Wage Inequality? A Panel Study Addressing Independent Variables of Inherently Different Sizes across Units. *Journal of Risk and Financial Management* 14: 436. <https://doi.org/10.3390/jrfm14090436>

Academic Editor: Shuddhasattwa Rafiq

Received: 16 August 2021

Accepted: 6 September 2021

Published: 9 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Let us assume that we do panel regressions where the independent variables are of inherently different sizes across units (for a general introduction to panel regression, see, e.g., [Cameron and Trivedi 2010](#); [Wooldridge 2010, 2019](#)). An example can be a panel study of industries where some include a few employees or firms, and others include numerous employees or firms. In this case, modeling industry size as an independent variable merits particular attention, and we explain why and how such data can be an advantage interpretatively. Each issue we discuss here is not necessarily novel per se, but we argue that bringing the different pieces together as a part of a larger puzzle can serve as a helpful tool for researchers when analyzing and interpreting panel data where the independent variables are of inherently different sizes across units.

To illustrate our idea empirically, we assess if industry size in the number of employees and firms as independent variables affects wage inequality as a dependent variable. Particularly, we examine industry size in an absolute number of employees and firms, in conjunction with the number of employees and firms relative to each industry's average for the years included. Shortly, we address why and how such distinctions can provide novel insight and be an advantage interpretively.

There is an extensive literature examining wage inequality (for a recent summary of this literature, see, e.g., [Aarstad and Kvitastein 2021](#)). In particular, scholars have studied how gender and firm characteristics can explain the concept ([Heinze and Wolf 2010](#); [Mitra 2003](#)), and research has also hinted that industry characteristics may play a role ([Faggio et al. 2010](#); [Song 2020](#)). However, to our knowledge, previous research has not explicitly studied if wage inequality is a function of industry size. Nor have we seen previous studies explicitly assessing how to deal with independent variables that are of inherently different sizes across units and illuminating how such data can be an advantage interpretatively. Hence, we argue that our contribution fills important research gaps that have not been addressed in previous economics and econometric studies. First,

it contributes to the literature by examining if wage inequality is a function of industry size. Second, it contributes to the literature by shedding light on nuances concerning independent variables that are of inherently different sizes across units.

2. A Discussion of Independent Variables of Inherently Different vs. Similar Sizes across Units

Assuming that we in panel regression find that wage inequality is a function of industry size in an absolute number of firms, how can we interpret this result? The most intuitive and obvious interpretation, in our opinion, is that industry size affects wage inequality, but can we learn more from the data? Yes, we argue, and the reason, as we have asserted above, is that industries are of inherently different sizes across units.

To exemplify, an industry of 10,000 firms at the outset that increases its size by one percent will impact the dependent variable roughly 100 times more in magnitude than a small industry of only 100 firms that also increases its size by one percent. (As noted, empirically, we study industry size both in the number of employees and firms, but in the following illustration, we only refer to industry size in the number of firms.) According to our above argument, a marginal change in industry size from one year to another, by one percent, can partly be explained as a random process most marked for large industries in absolute terms. Due to this randomness, the larger the industry size is at the outset, the more impact it will have on the dependent variable.

Returning to the illustration of two industries of 10,000 and 100 firms at the outset, assuming some equally distributed random variation in the size from one year to another, the large industry's average impact on the dependent variable is roughly 100 times stronger than of the small industry. It implies that the impact on the dependent variable is not only a change in industry size in the absolute number of firms, but that wage inequality is more sensitive to a random variation in industry size in large rather than small industries (We also acknowledge that industry size in terms of the absolute number of firms at the outset may reflect other underlying causal agents, which we address in the conclusion of this article).

However, unless we also control for a relative change in industry size, we cannot know whether wage inequality is particularly sensitive to industry size in terms of an absolute number. The reason is that, although industries vary in an absolute number, we cannot know whether it is a change in absolute size, which is most marked for large industries, or a change in relative size, which affects the dependent variable. Therefore, we suggest including an additional variable measuring industry size regarding the number of firms relative to each industry's average number of firms for the years included in the panel.

Assuming that we carry out fixed effects panel regression, the modeling of the *absolute* value of the independent variable x of industry i at year t is $x_{it} - \bar{x}_i$. (1), while the modeling of the *relative* value can be shown to be x_{it}/\bar{x}_i . (2); \bar{x}_i is the average of x for all years industry i is included in the data. As an alternative to Expression (2), we can instead log-transform the independent variable, resulting in that the value of industry i at year t is $\ln(x_{it}) - \ln(\bar{x}_i) = \ln(x_{it}/\bar{x}_i)$. As a second alternative to Expression (2), we can divide the independent variable's value at year t by the value at year t_{-1} . As a third alternative to Expression (2), we can standardize industry-year observations within each industry, e.g., zero mean value and a standard deviation, taking the value of one for industry-year observations within each industry. The alternative measures of Expression (2) are subject to their own interpretations, but they all tap into the same idea of measuring the relative vs. absolute value of the independent variable. For instance, assuming modest variation and drift in the independent variable over time, the effect on the dependent variable concerning the second alternative will not be much altered compared to the first and initial alternative reflected in Expression (2); however, it induces a loss in observations by the number of industries included in the panel data.

As such, we capture the eventual effect of an industry's change in (1) an absolute value, which is most marked for large industries according to our above arguments, along with (2) a relative change, which does not substantially discriminate between industry size

at the outset. Taken together, in the panel regression, we suggest the inclusion of both measures, (1) and (2), as independent variables.

3. An Empirical Illustration

To illustrate our approach empirically, we revisit a panel dataset, the details of which are reported by [Aarstad and Kvitastein \(2021, pp. 2–3\)](#). The unit of analysis is Norwegian industries (digit-two NACE-codes) between 2001 and 2014. The industries are of inherently different sizes, concerning both the number of employees and the number of firms. The dependent variable is wage inequality at year t among full-time employees in industry i , measured by using the Gini coefficient, G_{it} . Formally, $G_{it} = 1 - 2L_{it}$, where L_{it} is the area under the Lorenz curve. On a horizontal scale from 0 to 1, the Lorenz curve sorts those earning wages in increasing order and, on a vertical scale from 0 to 1, reflects the cumulative amount of wages earned. If we theoretically assume that only one person in industry i at year t earned all the wages and the rest earned zero, $G_{it} = 1$ (as $L_{it} = 0$), and if all earned exactly the same wages, $G_{it} = 0$ (as $L_{it} = 0.5$).

Table 1 reports descriptive statistics, and Table 2 reports correlations. Table 3 reports fixed-effects panel regressions with robust standard errors accounting for potential heteroscedasticity and serial correlation within industries (please see, for instance, [Cameron and Trivedi 2010, pp. 237, 257](#), for a detailed explanation of the fixed effect estimator and pp. 84–85, 239, 257–58, and 334–36 for a detailed explanation of robust standard errors; [Wooldridge 2019](#) also discusses fixed effects regression and the use of robust standard errors).

Table 1. Descriptive statistics.

	Min.	Max.	Mean	Median	SD	Skewness	Kurtosis
Wage inequality at t	0.118	0.388	0.222	0.216	0.042	0.892	4.21
Number of full-time employees in absolute terms at t	1007	106,204	16,150.8	10,456	19,598.1	2.78	11.0
Number of full-time employees in relative terms at t	0.507	1.57	1.00	1.01	0.153	−0.70	3.95
Number of firms in absolute terms at t	100	35,520	2179.7	859	4350.8	4.59	28.0
Number of firms in relative terms at t	0.479	1.88	1.00	0.997	0.177	0.569	5.64

Table 2. Correlation matrix.

	1	2	3	4
Wage inequality at t (1)				
Number of full-time employees in absolute terms at t (2)	−0.008			
Number of full-time employees in relative terms at t (3)	0.077 *	0.111 **		
Number of firms in absolute terms at t (4)	0.293 ***	0.554 ***	0.040	
Number of firms in relative terms at t	0.085 *	0.073 *	0.603 ***	0.079 *

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Conservative two-tailed tests.

Models 1–3 in Table 3 include four different measures of industry size at year t as independent variables: (1) number of full-time employees in absolute terms; (2) number of full-time employees in relative terms; (3) number of firms in absolute terms; and (4) number of firms in relative terms (please see the discussion in the previous section concerning the distinction between modeling the independent variables in absolute and relative terms).

Models 1 and 3 show that only the number of firms in absolute terms at t significantly affects wage inequality. Table 2 reports that the correlation coefficient between the number of full-time employees in absolute terms at t and the number of firms in absolute terms at t is 0.554 and 0.603 between the number of full-time employees in relative terms at t and the number of firms in relative terms at t . Therefore, we add Models 2 and 3 in Table 3 to avoid potential problems with multicollinearity. The other correlation coefficients between the independent variables are low, taking a maximum absolute value of 0.111.

Table 3. Fixed effects regressions with robust standard errors and wage inequality at t as the dependent variable.

	Model 1	Model 2	Model 3	Model 4
Number of full-time employees in absolute terms at t	-3.12×10^{-7} (3.97×10^{-7})	-4.36×10^{-8} (4.11×10^{-7})		
Lagged variable of the number of full-time employees in absolute terms at t_{-1}				9.04×10^{-9} (4.87×10^{-7})
Number of full-time employees in relative terms at t	-2.96×10^{-3} (0.012)	-5.87×10^{-3} (0.012)		
Lagged variable of the number of full-time employees in relative terms at t_{-1}				-2.25×10^{-3} (0.013)
Number of firms in absolute terms at t	3.69×10^{-6} *** (6.28×10^{-7})		3.60×10^{-6} *** (6.23×10^{-7})	
Lagged variable of the of firms in absolute terms at t_{-1}				2.91×10^{-6} *** (6.51×10^{-7})
Number of firms in relative terms at t	-1.73×10^{-3} (6.93×10^{-3})		-4.46×10^{-3} (7.93×10^{-3})	
Lagged variable of the number of firms in relative terms at t_{-1}				-1.40×10^{-3} (7.65×10^{-3})
Year dummies included (but not reported)	Yes	Yes	Yes	Yes
N industry-year obs./industries	853/67	853/67	853/67	783/67
Min./avg./max. obs. per industry	2/12.7/14	2/12.7/14	2/12.7/14	1/11.7/13
F-value	30.6 ***	12.6 ***	12.7 ***	28.5 ***
R-sq. within/between	0.355/0.130	0.328/0.002	0.350/0.090	0.342/0.085

*** $p < 0.001$. Conservative two-tailed tests for regressors.

Model 4 in Table 3 replicates Model 1, except that we include lagged independent variables at t_{-1} . Only the number of firms in absolute terms at t_{-1} significantly affects wage inequality in Model 4, but the regression estimate is lower, and the robust standard error is higher, than in Model 1.

As conventional Hausman (1978) specification tests cannot be executed with robust standard errors, we replicate the results from Table 3 in Table A1 in Appendix A with random effects (instead of fixed effects) estimators. The results show no substantial difference between the estimates in the two tables.

To assess if the number of firms in absolute terms may cause wage inequality or vice versa, we first balance the panel, resulting in a drop from 853 to 798 observations and from 67 to 57 industries/units (for balancing a panel, please see Yujun 2009). Next, we perform panel data Granger causality tests, applying the Bayesian information criterion on the number of lags included (Dumitrescu and Hurlin 2012; Lopez and Weber 2017). Table 4 informs that the number of firms in absolute terms robustly appears to cause wage inequality, but the data also indicate a reverse, albeit less robust, causality.

Table 4. Panel data Granger causality tests.

	Wage Inequality at t	
	Dependent Variable	Independent Variable
Number of firms in absolute terms at t (NFA)	(1) NFA as independent variable Z-bar = 12.3 ($p < 0.001$) Z-bar tilde = 7.19 ($p < 0.001$)	(2) NFA as dependent variable Z-bar = 3.86 ($p < 0.001$) Z-bar tilde = 1.65 ($p < 0.10$)

To test for unit roots or stationarity concerning the dependent variable, we apply the [Harris and Tzavalis \(1999\)](#) test on the balanced panel, which is appropriate when the number of units is relatively high compared to the number of time periods. Executing the test with panel-specific means and no time trend (default options), but with a small sample adjustment to the number of time periods (14), it generates a point estimate $p = 0.709$ and a z statistic of -2.65 ($p < 0.01$). When accounting for potential cross-sectional correlation, by removing cross-sectional averages ([Levin et al. 2002](#)), $p = 0.652$ and the z statistic is -4.60 ($p < 0.001$). The tests reject the assumption of unit roots, and we conclude that the dependent variable is stationary.

4. Conclusions

The focus of this study was to assess how independent variables of inherently different sizes across units in panel regression can be an advantage interpretively. Revisiting a panel dataset by [Aarstad and Kvitastein \(2021\)](#), taking an industry level of analysis, we found that wage inequality is a function of industry size in an absolute number of firms. A possible reason is that specialized skilled employees negotiate higher wages when there are many legal entities.

According to our previous discussion, the finding can also imply that wage inequality is not genuinely a function of change in industry size in an absolute number of firms, but that wage inequality is more sensitive to random size variation in large rather than small industries. A possible explanation is that industries with many firms may tend to face monopolistic competition, e.g., restaurants or retail firms offering relatively similar products or services (for further readings on monopolistic competition, please see, e.g., [Krugman and Obstfeld 2018](#)). These industries typically have low entry and exit barriers and relatively many low-skilled employees. In times of expansion, as indicated by an increase in the number of firms, these industries may increase bonuses and wages among managers and specialized high-skilled employees, while the many low-skilled employees' wages are relatively sticky (as they can be recruited from a large pool of potential employees outside of the industry). Similarly, establishing new firms in these industries will increase the demand for experienced managers and specialized high-skilled employees (while low-skilled employees can be recruited from a larger workforce pool outside of the industry). *Ceteris paribus*, these issues will increase wage inequality. In times of retraction, as indicated by a decrease in the number of firms, these industries may decrease bonuses and wages among managers and specialized high-skilled employees, while the many low-skilled employees' wages are relatively sticky (as they can be recruited to a large workforce outside of the industry). Similarly, the reduction in the number of firms will decrease the demand for managers and specialized high-skilled employees (while low-skilled employees can be recruited to a large workforce pool outside of the industry). *Ceteris paribus*, this will decrease wage inequality.

To gain additional knowledge about wage inequality seemingly being more sensitive to a change in large rather than small industries, we replicated Model 1 in [Table 3](#) but restricted the sample to only include industry-year observations of an industry *increase* and *decrease*, respectively, in the number of firms between t_{-1} and t . In other words, we carried out two separate analyses, one where the number of firms in industry i at $t > t_{-1}$ and one where the number of firms in industry i at $t < t_{-1}$. Interestingly, the data showed that when industries increase in terms of the absolute number of firms ($t > t_{-1}$), the effect on

the dependent variable is 4.32×10^{-6} (standard error 7.81×10^{-7} and $p < 0.001$, two-tailed test), while it is merely 2.26×10^{-6} (standard error 8.63×10^{-7} and $p < 0.05$, two-tailed test) when industries decrease in terms of the absolute number of firms ($t < t_{-1}$). In other words, wage inequality is not only more sensitive to a change in industry size in large rather than small industries (according to our previous discussion), but the effect is more marked and robust when industries increase rather than decrease in size. Additionally, the data showed more industry-year observations with an increase in the absolute number of firms between t_{-1} and t (574) than with a decrease (273). An implication is that large industries appear to be particularly strong positive carriers of wage inequality.

Table 4 indicates that industry size in terms of the absolute number of firms robustly appears to cause wage inequality, but, in line with our discussion above, we cannot rule out other underlying and genuine causal mechanisms. To further investigate this issue, we replicated Model 1 in Table 3 with a dynamic unconditional quasi-maximum likelihood fixed effects model with robust standard errors, developed by Kripfganz (2016) and recently extended by Williams et al. (2019). Interestingly, it showed a non-significant association between industry size in terms of the absolute number of firms and wage inequality, indicating that the independent variable is not strictly exogenous but reflects other underlying causal mechanisms at play. (Using the same technique of dynamic modeling, we also checked if wage inequality as an independent variable is associated with industry size in terms of the absolute number of firms as a dependent variable, but the results were non-significant.)

A possible explanation of the non-significant association between industry size in terms of the absolute number of firms and wage inequality can be that gender distribution is a genuine causal agent of the dependent variable (Strittmatter and Wunsch 2021), but it hinges on the assumption that a change in gender distribution is reflected by a change in the absolute number of firms. We cannot find strong arguments for the plausibility of this assumption but encourage future research to address the issue by including data on gender distribution.

Another possible explanation, as discussed above, is that industries with many firms may tend to face monopolistic competition, which reflects their sensitivity to wage inequality. If this is the case, a policy implication is that industry stakeholders (e.g., politicians, employers, employees, owners, and trade unions) should be made aware of such a tendency when negotiating wages, work conditions, and the formation of bonuses among different groups of employees. Nonetheless, we encourage future research to understand these potential mechanisms further before explicit policy recommendations are communicated to relevant stakeholders.

Author Contributions: Conceptualization, J.A.; methodology, J.A. and O.A.K.; formal analysis, J.A. and O.A.K.; writing—original draft preparation, J.A.; writing—review and editing, J.A. and O.A.K. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are not publicly available due to restrictions imposed by their provider.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Random effects regressions with robust standard errors and wage inequality at t as the dependent variable.

	Model 1	Model 2	Model 3	Model 4
Number of full-time employees in absolute terms at t	-3.98×10^{-7} (2.75×10^{-7})	-3.92×10^{-8} (2.59×10^{-7})		
Lagged variable of the number of full-time employees in absolute terms at t_{-1}				-1.97×10^{-7} (3.20×10^{-7})
Number of full-time employees in relative terms at t	-1.81×10^{-3} (0.012)	-5.90×10^{-3} (0.012)		
Lagged variable of the number of full-time employees in relative terms at t_{-1}				5.24×10^{-4} (0.013)
Number of firms in absolute terms at t	3.84×10^{-6} *** (5.97×10^{-7})		3.48×10^{-6} *** (5.41×10^{-7})	
Lagged variable of the of firms in absolute terms at t_{-1}				3.21×10^{-6} *** (5.94×10^{-7})
Number of firms in relative terms at t	-1.93×10^{-3} (6.86×10^{-3})		-4.33×10^{-3} (7.82×10^{-3})	
Lagged variable of the number of firms in relative terms at t_{-1}				-1.84×10^{-3} (7.51×10^{-3})
Year dummies included (but not reported)	Yes	Yes	Yes	Yes
N industry-year obs./industries	853/67	853/67	853/67	783/67
Min./avg./max. obs. per industry	2/12.7/14	2/12.7/14	2/12.7/14	1/11.7/13
Wald χ^2	532.9 ***	179.1 ***	346.4 ***	489.0 ***
R-sq. within/between	0.355/0.136	0.328/0.002	0.350/0.089	0.341/0.119

*** $p < 0.001$. Conservative two-tailed tests for regressors.

References

- Aarstad, Jarle, and Olav A. Kvitastein. 2021. Do Operating Profits Induce a Wage Premium Equally Shared among Employees Earning High or Low Incomes? *Economies* 9: 81. [CrossRef]
- Cameron, Adrian Colin, and Pravi Trivedi. 2010. *Microeconometrics Using Stata*, revised ed. College Station: Stata Press.
- Dumitrescu, Elena-Ivona, and Christophe Hurlin. 2012. Testing for Granger non-causality in heterogeneous panels. *Economic Modelling* 29: 1450–60. [CrossRef]
- Faggio, Giulia, Kjell G. Salvanes, and John Van Reenen. 2010. The evolution of inequality in productivity and wages: Panel data evidence. *Industrial and Corporate Change* 19: 1919–51. [CrossRef]
- Harris, Richard D. F., and Elias Tzavalis. 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91: 201–26. [CrossRef]
- Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–71. [CrossRef]
- Heinze, Anja, and Elke Wolf. 2010. The intra-firm gender wage gap: A new view on wage differentials based on linked employer–employee data. *Journal of population Economics* 23: 851–79. [CrossRef]
- Kripfganz, Sebastian. 2016. Quasi-Maximum likelihood estimation of linear dynamic short-T panel-data models. *The Stata Journal* 16: 1013–38. [CrossRef]
- Krugman, Paul R., and Maurice Obstfeld. 2018. *International Economics: Theory and Policy*, 11th ed. Westminster: Pearson Education.
- Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu. 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108: 1–24. [CrossRef]
- Lopez, Luciano, and Sylvain Weber. 2017. Testing for Granger Causality in Panel Data. *The Stata Journal* 17: 972–84. [CrossRef]
- Mitra, Aparna. 2003. Establishment size, employment, and the gender wage gap. *The Journal of Socio-Economics* 32: 317–30. [CrossRef]
- Song, Sang-Yoon. 2020. Employer Size and Wage Inequality: Rent-Sharing Role of Performance Pay. *Korean Economic Review* 36: 415–44.
- Strittmatter, Anthony, and Conny Wunsch. 2021. The Gender Pay Gap Revisited with Big Data: Do Methodological Choices Matter? *arXiv* arXiv:2102.09207.
- Williams, Richard, Paul Allison, and Enrique Moral-Benito. 2019. XTDPDML: Stata Module to Estimate Dynamic Panel Data Models Using Maximum Likelihood. Available online: <https://econpapers.repec.org/software/bocbocode/s458210.htm> (accessed on 1 July 2021).

Wooldridge, Jeffrey Marc. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge: MIT Press.

Wooldridge, Jeffrey Marc. 2019. *Introductory Econometrics: A Modern Approach*, 7th ed. Boston: Cengage.

Yujun, Lian. 2009. *XTBALANCE: Stata Module to Transform the Dataset into Balanced Panel Data*. Statistical Software Components S457094. Chestnut Hill: Boston College. Available online: <https://econpapers.repec.org/software/bocbocode/s457094.htm> (accessed on 1 July 2021).