American Economic Review: Papers & Proceedings 2017, 107(5): 398–403
https://doi.org/10.1257/aer.p20171094

Worker Betas: Five Facts about Systematic Earnings Risk†

By Fatih Guvenen, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo*

How are the labor earnings of a worker tied to the fortunes of the aggregate economy, his employer, and his industry? How do these risk exposures vary by gender, age, the worker's earnings level, industry, and the size of his employer? The answers to these questions have important implications for various theories in macro, labor, and financial economics. In this paper, we use a large and clean panel dataset on individual earnings from the US Social Security Administration to answer these questions. In particular, we estimate the risk exposure of a worker's earnings to three important risk factors: the aggregate economy, the performance of the worker's employer, and the performance of the worker's industry.

Our main empirical approach is a pooled ordinary least squares (OLS) regression of a worker's real annual earnings growth on either real gross domestic product (GDP) growth or real stock returns, the average earnings growth of the worker's employer, and the average earnings growth of the worker's industry. The use of big data allows us to estimate these risk exposures, which we call *worker betas,* more accurately and at a granular level, resulting in a clearer picture of how income risk from various sources is distributed across the population.

We document five main facts:

(i) Aggregate risk exposure (to either GDP growth or stock returns) is U-shaped with respect to the earnings level.

(ii) Males are more exposed to aggregate risk than females.

(iii) Younger workers are more exposed to aggregate risk than older workers, except at the top of the earnings distribution where the relation reverses.

(iv) In the middle of the earnings distribution, males, younger workers, and those in construction and durable manufacturing have the highest aggregate risk exposure. At the top of the earnings distribution, those in finance have the highest aggregate risk exposure. Workers in health and education have the lowest aggregate risk exposure throughout the earnings distribution.

(v) Workers in larger employers are less exposed to aggregate risk, but they are more exposed to a common factor in employer-level earnings, especially at the top of the earnings distribution. Within an employer, higher-paid workers have higher exposure to employer-level risk than lower-paid workers.

Our measurement exercise has important implications for several literatures in macro, labor, and financial economics.

For example, our findings relate to a large literature that estimates or calibrates individual income processes motivated by an incomplete markets model. This literature usually specifies a worker's log earnings as the sum of an aggregate shock (modeled as a time fixed effect), a Mincer earnings function of age and education, and the residual that is interpreted as an idiosyncratic

* Guvenen: Department of Economics, University of Minnesota, 4-101 Hanson Hall, 1925 Fourth Street South, Minneapolis, MN 55455, and NBER (e-mail: guvenen@umn.edu); Schulhofer-Wohl: Economic Research Department, Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604 (e-mail: sam@frbchi.org); Song: Social Security Administration, One Skyline Tower, 5107 Leesburg Pike, Falls Church, VA 22041 (e-mail: jae.song@ssa.gov); Yogo: Department of Economics, Princeton University, Julis Romo Rabinowitz Building, Princeton, NJ 08544, and NBER (e-mail: myogo@princeton.edu). We thank Gerald Ray at the Social Security Administration for assistance with the data. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago, the Social Security Administration, or the National Bureau of Economic Research.

† Go to https://doi.org/10.1257/aer.p20171094 to visit the article page for additional materials and author disclosure statement(s).

shock. According to our findings, the standard approach underestimates systematic risk by ignoring the differential exposure across workers to aggregate risk as well as employer- and industry-level risk. Thus, the standard approach misinterprets the residual from the wage regression as purely idiosyncratic when in fact it contains several sources of systematic risk. Properly decomposing earnings through a factor model, as we do in this paper, makes the residual closer to the theoretical concept of idiosyncratic risk that is unrelated to aggregate outcomes and pertains only to the circumstances of the individual worker.

The decomposition of earnings into systematic versus idiosyncratic components is a key input into the macro debate on the cost of business cycles and the benefits of stabilization policies (Lucas 2003). The standard specification of earnings based on time fixed effects that we described above would imply that business cycles and stabilization policies have homogeneous effects on income growth across the population. In contrast, our estimates based on heterogeneous exposure to risk factors imply that the cost of business cycles is borne asymmetrically across the population depending on gender, age, the worker's earnings level, and industry. Therefore, monetary or fiscal policies that stabilize business cycles would also have heterogeneous benefits across the population.

Our findings also have implications for the importance of intergenerational risk sharing through a social security system (Allen and Gale 1997). In an overlapping generations economy, an important source of market incompleteness arises from the inability of generations that live in different periods to insure aggregate risk through financial markets. A government can improve welfare through a fully funded social security system that transfers income from lucky to unlucky generations. This transfer system could be improved with better knowledge of how aggregate income risk is distributed across the population.

Turning to another important issue, a theory of risk sharing under heterogeneous risk preferences implies that more risk-averse individuals should bear less aggregate consumption risk. One mechanism through which efficient risk sharing could be achieved is for more risk-averse individuals to choose jobs or occupations where *income* has less aggregate risk exposure (Schulhofer-Wohl 2011). When combined with estimates of risk preferences (e.g., from survey data), our estimates of aggregate risk exposure could be used to test this theory more precisely. Our finding that higher-paid workers have higher exposure to the employer-level risk than lower-paid workers could also be consistent with theories of risk sharing within firms (Guiso, Pistaferri, and Schivardi 2005). Lower-paid workers need more income insurance if they are more risk averse or have more limited self-insurance opportunities.

A theory of portfolio choice in the presence of risky labor income implies that the optimal allocation to stocks depends on the covariance of income growth with stock returns. More specifically, the formula for the optimal portfolio share in stocks is a weighted average of the mean-variance portfolio and the hedging portfolio (Campbell and Viceira 2002, equation 6.11). Hedging demand implies that the optimal portfolio share in stocks decreases with stock return beta, which is exactly what we estimate in this paper. Therefore, our estimates of stock return beta could be used for normative advice on how investors with different income risks should tilt their allocation to stocks. When combined with data on portfolio choice, our estimates of stock return beta could be used to test the theory of portfolio choice.

I. Administrative Data on Earnings

Our annual panel data on earnings are from the Master Earnings File of the Social Security Administration from 1978 to 2013. These administrative data are representative, complete, and free of measurement error because they are based on all employer filings of Form W-2 for all US workers with a Social Security number. Importantly, the earnings data are not top coded and include all wages, salaries, bonuses, and exercised stock options as reported in Box 1 of Form W-2.[1] For each worker, we aggregate earnings across all his/her employers in a given year. We deflate earnings to 2009 US$ using the GDP implicit price deflator.

[1] In addition to W-2 wages, the Master Earnings File contains self-employment income. However, we do not use self-employment income in our analysis because it was top coded prior to 1994.

In addition to earnings, we use demographic information from the Master Earnings File including gender, year of birth (or age), and the Standard Industrial Classification (SIC) code of the primary employer. Because of computing resource constraints, our analysis is based on a 10 percent representative sample of the Master Earnings File. We further limit our sample to workers who are in their prime working years from ages 26 to 65.

We compute real earnings growth as the difference in log real earnings between year t and $t-1$. As a proxy for *permanent income*, we also compute average real earnings over five years from year $t-6$ to $t-2$. When five years of earnings history are not available for a worker (primarily in the first four years of the panel from 1978 to 1981), we use the longest consecutive period (between one to four years) that is available. We emphasize that there is no overlap between the period over which earnings growth is computed (i.e., year $t-1$ to t) and the period over which average earnings are computed (i.e., year $t-6$ to $t-2$). This ensures that there is no mechanical correlation between our measures of earnings growth and average earnings. The data requirements for computing earnings growth and average earnings imply that to enter our sample, a worker must have positive earnings in years t, $t-1$, and at least one year between $t-6$ and $t-2$.

In each year, we group our sample into four age groups: 26–35, 36–45, 46–55, and 56–65. We also group our sample into 12 earnings percentiles (i.e., tenth to ninetieth, ninety-ninth, and 99.9th) conditional on gender and age group, based on average earnings that we described above. Finally, we group our sample into ten industries based on the four-digit SIC code of the primary employer: construction (1521–1799); nondurable manufacturing (111–1499, 2011–2399, 2611–3199, and 3951–3999); durable manufacturing (2411–2599 and 3211–3949); transportation (4011–4971); retail and wholesale (5012–5999); finance (6011–6799); services (7011–7999, 8111, and 8322–8999); health and education (8011–8099 and 8211–8299); and other industries (9111–9999 and missing SIC code).

Table A1 of the online Appendix summarizes our sample by gender and age. An advantage of our administrative data is that our sample is much larger than in typical studies of household finance that are based on surveys. For example, we have 5.073 million observations of males aged 36–45 who fall between the fiftieth and sixtieth percentiles of the earnings distribution. The median earnings for this group is \$45,000. We also have 457,000 observations of males aged 36–45 who fall between ninety-ninth and 99.9th percentiles of the earnings distribution, where median earnings is \$333,000. We even have 51,000 observations above the 99.9th percentile, where median earnings is \$1.073 million.

II. GDP and Stock Return Beta

A. *GDP Beta*

Let $\Delta y_{n,t}$ be the log real earnings growth of individual n in year t, and let Δy_t be the log real GDP growth in year t. Our main regression specification is

$$\Delta y_{n,t} = \alpha_g + \beta_g \Delta y_t + \epsilon_{n,t}. \tag{1}$$

We estimate the coefficients α_g and β_g by pooled OLS, separately by gender, four age groups, and 12 earnings percentile bins. The assumption, for example, is that males aged 36–45 whose earnings fall between the fiftieth and sixtieth percentiles have the same GDP beta.

Figure 1 reports GDP beta across the earnings distribution at age 36–45 by gender. For both males and females, GDP beta is U-shaped in the earnings level. That is, workers at the tails of the earnings distribution have the highest aggregate risk exposure. In particular, males at the 99.9th percentile of the earnings distribution have a GDP beta of 3.70. Parker and Vissing-Jorgensen (2009) also find that income is most cyclical at the top of the earnings distribution. Throughout the earnings distribution, males have higher GDP beta than females. For example, males at the fiftieth percentile of the earnings distribution have a GDP beta of 1.09, compared with 0.69 for females.

Figure 1 also reports GDP beta across the earnings distribution for males by age group. Within each age group, GDP beta is U-shaped in the earnings level. Below the ninetieth percentile of the earnings distribution, younger males have higher GDP beta than older males. For example, males aged 26–35 at the fiftieth percentile of the earnings distribution have a GDP beta of 1.55, compared with 0.30 for males aged 56–65.

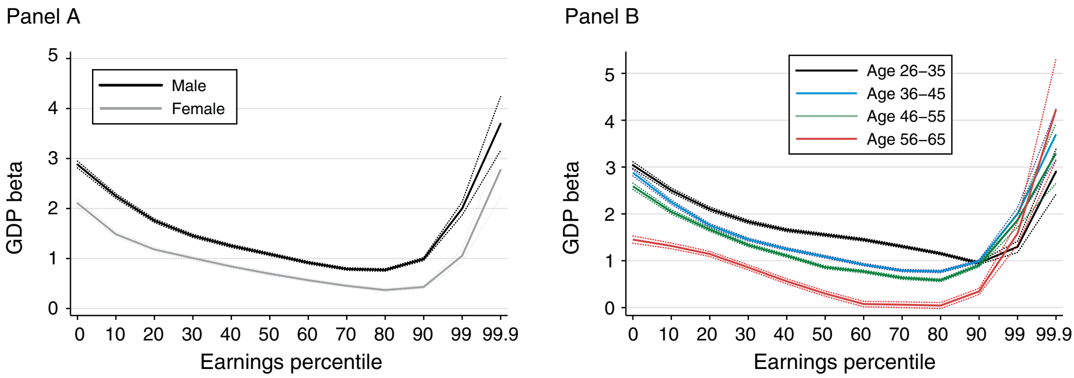

Figure 1. GDP Beta at Age 36–45 by Gender and for Males by Age Group

Notes: Earnings percentiles (conditional on gender and age group) are based on average real earnings from year $t-6$ to $t-2$, while real earnings growth used to estimate GDP beta is from year $t-1$ to t. The dotted lines represent the 95 percent confidence interval.

Above the ninetieth percentile of the earnings distribution, however, this relation reverses so that older males have higher GDP beta than younger males. For example, males aged 56–65 at the 99.9th percentile of the earnings distribution have a GDP beta of 4.23, compared with 2.90 for males aged 26–35.

Figure B1 of the online Appendix reports GDP beta across the earnings distribution for males aged 36–45 by industry. There are significant differences in GDP beta across industries. At the fiftieth percentile of the earnings distribution, the industries (and corresponding GDP beta) ranked from the most to least cyclical are construction (2.31); durable manufacturing (1.97); services (1.17); retail and wholesale (1.05); nondurable manufacturing (0.88); finance (0.87); transportation (0.47); and health and education (0.23). This ranking should not surprising, except for the fact that finance is one of the less cyclical industries. However, the cyclicality of earnings in the finance industry is highly dependent on the earnings level. At the ninety-ninth percentile of the earnings distribution, finance is actually the most cyclical industry with a GDP beta of 3.05.

B. *Stock Return Beta*

We now repeat regression (1) with real stock returns instead of real GDP growth as the explanatory variable. Real stock returns are the Center for Research in Securities Prices value-weighted index deflated by the GDP implicit price deflator. In aligning earnings growth with stock returns, we use the beginning-of-period timing convention, which leads to a higher correlation between GDP growth and stock returns than the end-of-period timing convention. That is, we align earnings growth from year $t-1$ to t with stock returns during year $t-1$.

Figures C1 and C2 of the online Appendix report results that are analogous to Figures 1 and B1 for stock returns. In our sample from 1980 to 2013, the correlation between real stock returns and real GDP growth is 0.59. Therefore, it should not be surprising that our main findings for stock return beta are similar to those for GDP beta.

III. Employer and Industry Beta

Variation in earnings growth that remains after taking out the aggregate exposure to GDP growth need not be purely idiosyncratic. In particular, earnings growth could be correlated across workers within the same employer or industry. To examine the importance of factor structure in earnings growth at the employer and industry levels, we modify regression (1) to include employer- and industry-level factors.

Let $\Delta y_{e\backslash n,t}$ be the log real growth rate of average earnings for other workers working for worker n's employer (defined by the Employer

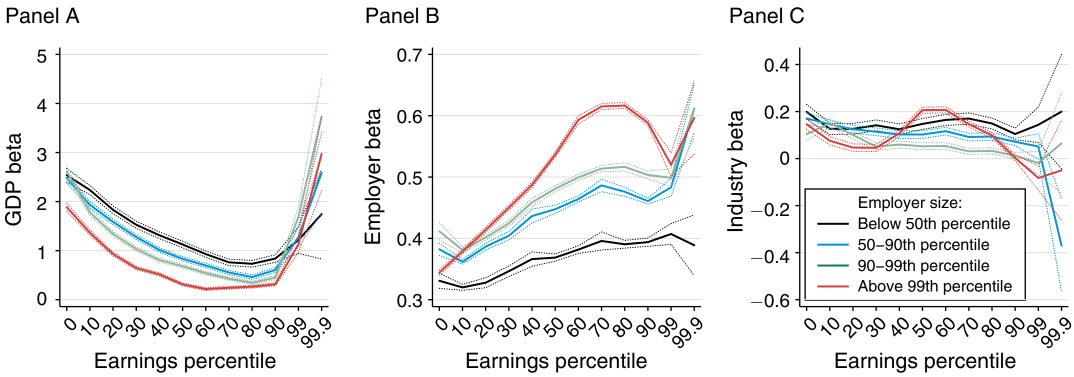

FIGURE 2. GDP, EMPLOYER, AND INDUSTRY BETA FOR MALES

Notes: Earnings percentiles (conditional on gender and age group) are based on average real earnings from year $t-6$ to $t-2$, while real earnings growth used to estimate GDP, employer, and industry beta is from year $t-1$ to t. The dotted lines represent the 95 percent confidence interval.

Identification Number) in year t, where we exclude worker n in computing average earnings. By excluding worker n from the average, we avoid any mechanical correlation in earnings growth between the worker and the employer. Similarly, let $\Delta y_{i\backslash n,t}$ be the log real growth rate of average earnings for worker n's industry in year t, where we again exclude worker n from the average. Our regression specification is

$$(2) \quad \Delta y_{n,t} = \alpha_e + \beta_g \Delta y_t + \beta_{g,e} \Delta y_{e\backslash n,t} + \beta_{g,i} \Delta y_{i\backslash n,t} + \nu_{n,t},$$

where α_e are employer fixed effects.

To isolate meaningful factors in employer-level earnings, we limit our sample to employers with at least ten observations in our sample. In each year, we group our sample by employer size with cutoffs at fiftieth, ninetieth, and ninety-ninth percentiles. Our four groups correspond to median employer sizes of 13, 31, 154, and 1,210 observations in our sample. Because we start with a 10 percent representative sample, these employers on average have 130, 310, 1,540, and 12,100 workers. To estimate regression (2), we first take out the employer fixed effects by cross-sectionally demeaning. We then estimate the coefficients by pooled OLS, separately by gender, 12 earnings percentile bins, and 4 employer-size groups.

Figure 2 reports GDP, employer, and industry beta across the earnings distribution for males by employer-size groups. Two important facts emerge. First, GDP beta decreases in employer size, while employer beta increases in employer size. GDP beta decreases from 1.13, 0.83, 0.68, to 0.31 by employer size at the fiftieth percentile of the earnings distribution. At the same time, employer beta increases from 0.37, 0.45, 0.48, to 0.54 in employer size. There is no such monotonic pattern in the industry beta. Second, employer beta increases in the earnings level. This means that within an employer, higher-paid workers absorb a higher share of the employer-level risk.

In summary, we have uncovered interesting heterogeneity in risk exposure across workers coming from aggregate-, employer-, and industry-level sources. A further exploration of the implications of systematic earnings risk for macro, labor, and financial economics is an exciting area of research, which we intend to undertake in future work.

REFERENCES

Allen, Franklin, and Douglas Gale. 1997. "Financial Markets, Intermediaries, and Intertemporal Smoothing." *Journal of Political Economy* 105 (3): 523–46.

Campbell, John Y., and Luis M. Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice* *for Long-Term Investors*. Oxford: Oxford University Press.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi. 2005. "Insurance within the Firm." *Journal of Political Economy* 113 (5): 1054–87.

Lucas, Robert E., Jr. 2003. "Macroeconomic Priorities." *American Economic Review* 93 (1): 1–14.

Parker, Jonathan A., and Annette Vissing-Jorgensen. 2009. "Who Bears Aggregate Fluctuations and How?" *American Economic Review* 99 (2): 399–405.

Schulhofer-Wohl, Sam. 2011. "Heterogeneity and Tests of Risk Sharing." *Journal of Political Economy* 119 (5): 925–58.