Accounting for the Cyclical Volatility of Wages

Alisdair McKay* and Tamás Papp†

October 5, 2010

Abstract

Wage volatility, measured as the cross-sectional variance of wage changes, is positively correlated with the unemployment rate with a correlation coefficient of 0.61. We decompose this correlation into three main factors. During a recession, wage volatility increases substantially among those experiencing spells of unemployment. The cyclical changes in the variance within this group explain about 55% of the cyclical variation in wage volatility. The variance within the group not experiencing unemployment explains 18%. Finally, an increase in the fraction of workers experiencing unemployment explains 25%. We quantify the posterior uncertainty surrounding this decomposition and show the results are robust.

*Boston University. amckay@bu.edu

†Institute for Advanced Studies, Vienna. tpapp@ihs.ac.at

1 Introduction

The variance of idiosyncratic shocks to earnings is larger during recessions, a phenomenon commonly referred to as countercyclical volatility. This relationship between aggregate and idiosyncratic shocks plays a key role in the analysis of the welfare cost of business cycles in heterogeneous-agent macro models. While countercyclical volatility has been documented statistically, we know very little about the micro-foundations that generate it.[1] This lack of micro-foundations is important because one must take a stand on how idiosyncratic risk will evolve as the aggregate shock process changes in order to compute the welfare cost of aggregate fluctuations. Developing models with micro-founded links between aggregate and idiosyncratic risk is, therefore, an important part of the research agenda that seeks to use heterogeneous-agent macro models to study aggregate stabilization. The purpose of this paper is to provide a statistical decomposition of cyclical changes in idiosyncratic risk in order to shed light on the underlying mechanism.

A second motivation for our work comes from search-theoretic models of the labor market. Recent advances now allow us to solve models that feature both wage dispersion and aggregate risk.[2] Our results, which document changes in wage dynamics over the business cycle, could be useful to evaluate and extend these new theoretical results. Indeed we view models in this class as the natural starting point for the development of theoretical links between aggregate and idiosyncratic risk.

We begin by documenting the extent of countercyclical volatility in the data. To do so, we use a multilevel statistical model that relates the variance of wage shocks to the unemployment rate. We estimate this model using data on men's wages from the PSID and the national unemployment

[1] See Storesletten et al. (2004) for estimates of countercyclical volatility. There is also a large literature on earnings instability following Gottschalk et al. (1994) that documents an increase in the variance of transitory earnings shocks during recessions.

[2] See Moscarini and Postel-Vinay (2010), Menzio and Shi (2009), and Menzio and Shi (2010).

rate reported by the BLS. We find evidence of countercyclical volatility in annual wages as well as in annual earnings. It is not surprising that there is countercyclical volatility in earnings since the incidence and variability of unemployment shocks increase during a recession. Therefore, we focus on wages. Our chief object of interest is the cross-sectional variance of year-to-year changes in log wages, which we call wage volatility. Specifically, the main result of section 2 is the demonstration of a strong comovement between wage volatility and the unemployment rate, which we use as a cyclical indicator.

Our results in section 2 are related to work by Storesletten et al. (2004) who estimate an income process that allows the variance of shocks to differ between expansions and contractions. As our methodology differs from theirs, it is worth considering its advantages and disadvantages. Storesletten et al. point out that if the variance of persistent income shocks increases during a recession then cohorts who have lived through more recessions will have a larger variance of income levels. They then use variation in macroeconomic experiences across cohorts to estimate the extent to which the variance of income shocks increases in a recession. The real genius of their approach is that they recognize that data on income from 1967 onwards contain information about income shocks that occurred since 1930. The difficulty of this approach for our purposes is that we only have covariates, such as data on unemployment spells, since 1967 so we are unable to investigate the sources of income shocks before the start of our sample. In comparison to Storesletten et al., we use information on a more limited set of of aggregate fluctuations. Despite the fact that we only use information between 1967 and 1992, we identify a clear and significant countercyclical pattern in wage volatility. While the focus of our paper is on wages, we note in section 2.3 that when we apply our methods to data on incomes the resulting estimates are in line with those of Storesletten et al.

After documenting the magnitude of countercyclical wage volatility, the next step in our analysis is to investigate its sources. To do so, we partition the sample into groups according to whether an individual has experienced unemployment in the previous two years (i.e. the years over which we take the the first-difference in wages). We then perform a variance decomposition exercise that relates the total variance of wage growth (i.e. wage volatility) to the within group variances, the within group means and the group sizes and we explore how these components vary with the unemployment rate.

We show that the comovement of wage volatility and the unemployment rate is in part driven by a composition effect. The composition effect arises because more people experience unemployment in a recession and unemployment is associated with higher wage volatility in all aggregate conditions. Therefore, during a recession, the composition of the sample shifts towards the group with the larger variance. We estimate that this effect explains about 25% of the correlation between wage volatility and the unemployment rate. The other important factors are the cyclical changes in the within group variances. In particular, the variance of wage shocks among those experiencing unemployment increases strongly with the unemployment rate. This effect can explain about 55% of the cyclical movements in wage volatility. Increasing volatility among those not experiencing unemployment explains most of the remainder, about 18%.

In order to conduct the variance decomposition calculations we must have estimates of the mean and variance of wage growth in each unemployment-experience group and our decomposition results will be affected by the uncertainty surrounding these estimates. To quantify this uncertainty, we extend our model to estimate the mean and variance of wage growth in each group. We use a Bayesian estimation approach, which allows us to easily calculate the uncertainty about our decomposition results by using draws from the posterior distribution of the model. To our knowledge, this method of constructing standard errors for a variance decomposition has not previously been applied in the economics literature. We find that the posterior uncertainty about our results does not undermine the message of the point estimates reported above. For example, the 90% error band for the composition effect ranges from 21% to 29%.

Similar variance decomposition techniques have been used to understand secular trends in inequality. For example, Lemieux (2006) decomposes the trend in the variance of hourly wages from the CPS and shows that the shift towards an older, more educated workforce can explain a third to a half of the increase in residual wage inequality between 1973 and 2003.[3] Our approach differs from previous work because we focus on cyclical fluctuations as opposed to trends.

The paper is organized as follows: section 2 documents the magnitude of cyclical wage volatility, section 3 presents the variance decomposition methods and results, and section 4 concludes.

2 Wage Volatility and Unemployment

In this section we introduce our multilevel modeling approach and estimate the correlation between the volatility of wages and the unemployment rate. The data we use are from the Panel Study of Income Dynamics (PSID) covering income in years 1967 to 1992.[4] We measure wages as the ratio of annual labor income to annual hours worked and deflate to 1967 dollars using the CPI-Research

[3]The sociology literature has also explored trends in income inequality and recently Western and Bloome (2009) have shown how to construct standard errors for Lemieux's decomposition using Bayesian methods.

[4]Two considerations influence our choice of years to include in our sample. First, our object of interest is the first-difference of log annual wages for which we need data on wages in consecutive years. Therefore we cannot make use of PSID data after 1996 (survey year 1997) after which the PSID switches to a biannual frequency. Second, there is a structural break in PSID wage volatility around 1993, which has also been documented by Heathcote et al. (2010). The timing of this break coincides with the switch to a computer-based survey methodology although it is not clear whether this break represents an actual change in the data generating process or if it is an artifact of the methodological change. In our analysis, we have found that our findings survive if we model this break as a level shift in the wage volatility process. We choose, however, to end our sample in 1992 for the sake of simplicity and ease of exposition. Finally, the switch to the new survey methodology began in survey year 1993 and was completed in survey year 1994 so our 1992 data have some elements of the new methodology. The inclusion of 1992 does not exert a strong influence on our results.

price index. We restrict the sample to male heads who were between the ages of 25 and 60 and worked at least 320 hours per year. Students, business owners, and self-employed individuals are excluded from the analysis. We focus on the first difference of log wages across years so individuals must be present for two consecutive years in order to be included in our sample. Appendix A provides further details about our sample.

2.1 Multilevel Model

Our statistical model consists of three equations. For an individual i in year t, the model specifies the following distribution for wage growth

$$dw_{i,t} \sim \mathrm{N}\left(X_{i,t}\beta + \alpha_t, \sigma_t^2\right) \tag{1}$$

$$\alpha_t \sim \mathrm{N}\left(Z_t\gamma, v_\alpha\right) \tag{2}$$

$$\sigma_t^2 \sim \mathrm{N}\left(Z_t\delta, v_{\sigma^2}\right), \quad \sigma_t^2 > 0. \tag{3}$$

The first line states that the change in an individual's log wage, $dw_{i,t}$, is normally distributed. The mean of this distribution depends on the individual's demographic characteristics such as age and education, which are placed in the vector $X_{i,t}$. We assume that the coefficients on these demographic characteristics are common across years. In addition, the mean change in wages varies over time with the α_t term. Finally, we allow the variance of the innovation in wages to vary over time as captured by the σ_t^2 term.

The second and third lines show the multiple levels of our model as we impose structure on the parameters of the wage growth distribution and assume that they are drawn from their own distributions.[5] The α_t terms are drawn from a normal distribution, the mean of which is linearly

[5] See Gelman et al. (2004) and Gelman and Hill (2007) for a discussion of multilevel models.

related to aggregate variables, Z_t. We use the national unemployment rate as our measure of aggregate conditions and this, along with a constant, makes up the vector Z_t. Similarly, equation (3) relates the variance of wage growth to aggregate conditions. As in equation (2), the mean of this distribution is linearly related to the national unemployment. Our main object of interest is the component of δ that is the coefficient on the unemployment rate in equation (3), which we call δ_{unemp}. As the variance must be positive, we draw from a truncated normal distribution. In practice, however, the mass of the distribution below zero is so small that this truncation is practically irrelevant.

By estimating all of the parameters of the model jointly, the posterior uncertainty about δ_{unemp} reflects the uncertainty about σ_t^2. By contrast, one can imagine a two-stage estimation procedure in which one first estimates equation (1) and then computes σ_t^2 from the residuals of this first-stage regression and uses these estimates as "data" in estimating equation (3). The difficulty with this two-stage approach is that the standard errors in the second-stage regression do not reflect the fact that σ_t^2 is itself an estimate. The multilevel model avoids this difficulty by estimating both stages at once.

Another advantage of the multilevel model is that it provides sharper estimates of σ_t^2 than one would obtain from the first-stage regression. The reason is that the multilevel model is able to pool information across years if the data suggest that $dw_{i,t}$ in those years are generated by a similar process. Consider the meaning of the parameter v_{σ^2}. If the unemployment rate were the same in years t and s and $v_{\sigma^2} \approx 0$, then from equation (3) it follows that $\sigma_t^2 \approx \sigma_s^2$, which is to say that the variance of wage changes in years t and s is the same. If these variances are equal, we can estimate them more precisely by pooling the data from years t and s to estimate the single variance that applies to both years. Alternatively, if v_{σ^2} is very large, the implication is that even if the unemployment rate is the same in years t and s we have no reason to think that σ_t^2 and σ_s^2 should be related. Therefore, we should estimate σ_t^2 and σ_s^2 separately without pooling the data. In between these extremes, the model can partially pool the data across years by down-weighting the data from year s when estimating σ_t^2.

The amount of pooling that actually occurs in estimating σ_t^2 depends on the value of v_{σ^2} which is itself estimated at the same time. This is possible because the likelihood depends on the parameters of both levels of the model. If the data do not call for pooling, low values for v_{σ^2} have low likelihood because the data require (relatively) large errors in equation (3) or require that equation (1) be fit with similar variances, which is at odds with the data. Conversely, if the data call for pooling, high values of v_{σ^2} have low likelihood because the errors in equation (3) are small so the likelihood can be raised by reducing v_{σ^2}. By following this logic, the multilevel model is able to pool data across years to the extent called for by the data. In section 3.2.1 we discuss exactly how much sharper our estimates are as a result of this partial pooling.

By including the unemployment rate in equation (3), we allow the model to attribute some of the differences in σ_t^2 across years to changes in the unemployment rate. If the unemployment rate explains some of the variation in σ_t^2, then we can tighten our estimate of v_{σ^2}. Since equation (3) is our prior for estimating σ_t^2, a lower value of v_{σ^2} implies a sharper prior on σ_t^2. This prior, which is itself estimated from data across years, is the mechanism through which the model is able to use information from other years to inform the estimate of σ_t^2. When the prior becomes more precise, it has a larger effect on the estimate of σ_t^2 and more information is pooled across years. So if equation (3) fits better, v_{σ^2} falls and more information is pooled. In effect, the inclusion of predictors, such as the unemployment rate, in equation (3), allows the model to identify dimensions on which we expect σ_t^2 to differ between years and therefore pool information more effectively.

2.1.1 Prior Distributions

We need to specify prior distributions for β, γ, δ, v_α and v_{σ^2}. Since the first three are regression coefficients, it is natural to use the non-informative (reference) priors

$$p(\beta) \propto 1 \tag{4}$$

$$p(\gamma) \propto 1 \tag{5}$$

$$p(\delta) \propto 1 \tag{6}$$

There is sufficient sample size at each level of the model to make the posterior distribution proper.

In the context of an ordinary (single-level) linear regression, the usual choice for a non-informative reference prior for the variance v of the error term is $p(v) \propto v^{-1}$. In the context of multilevel models, however, Gelman (2006) demonstrates that this prior places infinite mass near $v = 0$, resulting in an improper posterior distribution. The same author constructs a proper, conditionally-conjugate prior distribution using the folded-noncentral-t family, but also notes that the improper prior distribution $p(v) \propto v^{-1/2}$ is usually a reasonable choice when there is sufficient data. In the generic language of multilevel modeling, each set of observations that are drawn from the same distribution is referred to as a group. As we have 25 years of data, we therefore have 25 groups. Specifically, Gelman shows that for three or more groups, $p(v) \propto v^{-1/2}$ yields a proper posterior, but this prior will also introduce a unreasonable amount of bias for fewer than six groups. Since we have 25 groups, this is not an issue for our model, and we follow Gelman's recommendation and use the prior distributions

$$p(v_\alpha) \propto v_\alpha^{-1/2} \tag{7}$$

$$p(v_{\sigma^2}) \propto v_{\sigma^2}^{-1/2}. \tag{8}$$

2.1.2 Estimation

We use a Gibbs sampler to draw from the posterior distribution of the parameters of the model. We partition the parameter space into blocks corresponding to β, α, γ, v_α, σ^2, δ, and v_σ^2 and sample each block in turn. Many of the sampling steps reduce to drawing from an ordinary linear regression. Appendix B contains details on the estimation methodology.

2.2 Results

	mean	5%	25%	50%	75%	95%
β_{age}	−0.0014	−0.0016	−0.0015	−0.0014	−0.0012	−0.0011
β_{edu}	0.0029	0.0019	0.0025	0.0029	0.0033	0.0038
γ_{constant}	0.0668	0.0428	0.0574	0.0666	0.0765	0.0907
γ_{unemp}	−0.6332	−0.9921	−0.7762	−0.6311	−0.4913	−0.2733
v_α ($\times 1000$)	0.2203	0.1018	0.1544	0.2039	0.2666	0.3970
δ_{constant}	0.0499	0.0360	0.0444	0.0500	0.0555	0.0638
δ_{unemp}	0.4204	0.2150	0.3365	0.4200	0.5043	0.6319
v_{σ^2} ($\times 1000$)	0.0850	0.0461	0.0631	0.0791	0.0996	0.1436

Table 1: Posterior means and quantiles for model parameters.

Table 1 shows the posterior means and quantiles of the parameters of our model. The posterior mean and median of δ_{unemp} are both 0.42 and the 90% error band extends from 0.22 to about 0.63. δ_{unemp} can be thought of as the slope of a least squares regression of σ_t^2 on the unemployment rate. As such, a positive coefficient implies a positive correlation.

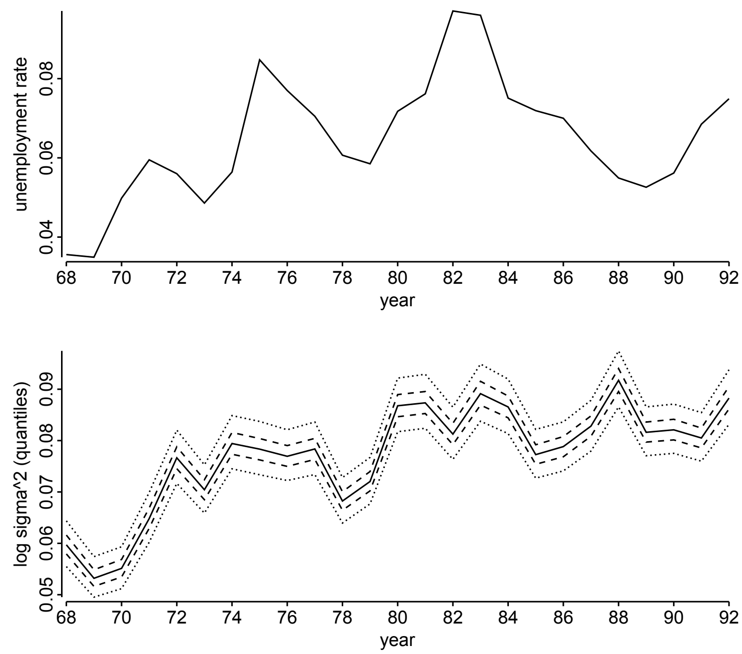


Figure 1: The top panel shows the unemployment rate. The bottom panel shows the posterior median of σ_t^2. The dashed lines show the 5%, 25%, 75%, and 95% quantiles from the posterior.

We now further document the correlation between wage volatility and the unemployment rate. When we draw from the model's posterior distribution we also draw σ_t^2 and these estimates are shown in Figure 1 along with the unemployment rate. For every posterior draw, we form the time-series of σ_t^2 and calculate the correlation with the unemployment rate. The median value of these simulated correlation coefficients is 0.61 and the 90% error band extends from 0.53 to 0.69. We believe that this is the clearest way of viewing the positive relationship between σ_t^2 and the unemployment rate and it is this quantity that we seek to decompose below.

To highlight the positive correlation between these two series, Figure 2 shows a scatter plot of the posterior median of the σ_t^2 against the unemployment rate. The vertical bars in the figure are 90% error bars for σ_t^2. The figure also shows posterior draws of the regression line from equation (3). The logic of the multilevel model is that we parameterize the prior distribution on σ_t^2 and

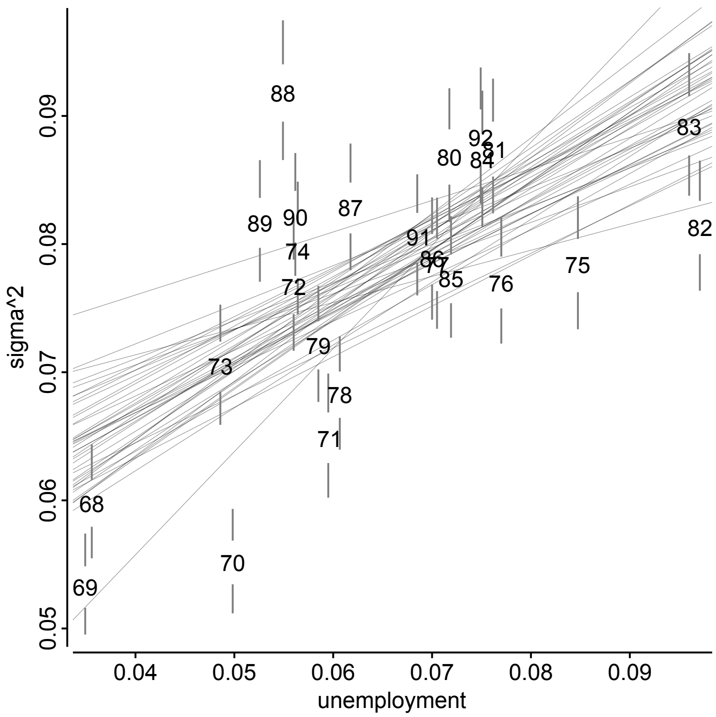


Figure 2: Scatter plot of posterior median of σ_t^2 against the unemployment rate in year t. Year labels are placed at the posterior median of σ_t^2 and vertical bars extend from the 25% to 5% quantiles and from the 75% to the 95% quantiles. The regression lines are draws from the posterior distribution of δ.

equation (3) represents the parameterization of this prior. The difference between the points in the scatter plot and the sample of regression lines is that the regression lines show the variation in the prior on σ_t^2 and the scatter plot shows the estimated values of σ_t^2, which are a compromise between the data and the prior.

We now briefly discuss some other features of our estimation results. γ_{unemp} is estimated to be negative, which implies that real wage growth is counter-cyclical. In the first-level of the model, the demographic effects show that wage growth declines with age and increases with education.

2.3 Relation to Previous Work

From the estimated relationship between the unemployment rate and σ_t^2, we can develop a sense of the changes in wage volatility over the business cycle with the following back of the envelope calculation. Over the course of a typical business cycle in the United States, the unemployment rate fluctuates by roughly 3 percentage points. Our estimate of δ_{unemp} then implies that σ_t^2 will increase by 0.013 as the economy moves from the peak of the cycle to the trough.

Many readers will compare our results to the work of Storesletten et al. (2004). Those authors use data on the income of households inclusive of transfers while we use wages of household heads so our results are not directly comparable to theirs. Bearing these differences in mind, one might still ask if our results are plausible in comparison to theirs. Their results imply that the variance of the first-difference in household earnings increases by about 0.041 as the economy moves from expansion to contraction.[6]

To facilitate the comparison of our results to theirs, we also estimate the model using the labor income as data instead of wages. All other features of our analysis remain the same. In this case,

[6] Appendix C explains how we reach this conclusion from their results.

the posterior mean of δ_{unemp} is 1.14 with a 90% error band extending from 0.76 to 1.54. Performing the same back of the envelope calculation as above produces a difference in income volatility of 0.034 as the unemployment rate increases by 3 percentage poitns. Given the important differences between their methodology and data and ours, we do not find it surprising that we obtain somewhat different results.[7] In particular, their analysis incorporates information about the Great Depression, which is to say that their sample includes larger fluctuations in the unemployment rate than ours does. In light of this difference, the fact that they find a bigger difference between expansions and contractions is perhaps in line with our finding.

2.4 Impact of Measurement Error

The PSID data on wages are surely affected by measurement error so it is important to consider how our results are affected. Suppose we measure $dw_{i,t} = \widehat{dw}_{i,t} + \varepsilon_{i,t}$ where $\widehat{dw}$ is the true wage growth and ε is measurement error distributed i.i.d. normal with some mean and variance, v_e. Then it follows from equation (1) that $\sigma_t^2 = \hat{\sigma}_t^2 + v_\varepsilon$, where $\hat{\sigma}_t^2$ is the variance of $\widehat{dw}_{i,t}$. So measurement error of this type will shift the intercept in equation (3), but δ_{unemp} is not affected. More generally, if the extent of measurement error varies over time it will affect our result to the extent that it covaries with the unemployment rate.

2.5 Wage Volatility Over the Business Cycle

One way to assess the role of business cycle fluctuations in driving our finding of a positive correlation between the unemployment rate and wage volatility is to compare years that are close in time

[7] The difference in methodology is that they look at the cross-sectional dispersion in earnings across cohorts who have lived through different macroeconomic experiences while we look at the dispersion of the *first-differences* in booms and recessions. The difference in data is that they look at total household income inclusive of transfers, while we look at just the labor income of male heads.

but differ in aggregate conditions. To do so, we use the NBER business cycle dates to identify peaks and then choose the years with the lowest and highest unemployment rates between peaks.[8] The pairs of peaks and troughs we identify are (1969, 1971), (1973, 1975), (1979, 1982) and (1989, 1992). Using posterior draws of σ_t^2, we calculate the difference in σ_t^2 for each pair and then compute the average of these differences across pairs. Table 2 shows the distribution of these average differences and the mean and median are both 0.0088. We also explore the increase in wage volatility per unit change in the unemployment rate. For each of the four pairs of years, we compute the ratio of the change in σ_t^2 over the change in the unemployment rate over those years. We then average over the four business cycles. We calculate the posterior distribution of this measure of the slope of the wage-volatility-unemployment relationship. The results in Table 3 show that the mean and median are both 0.31, which is comparable to the estimate of $\delta_{\text{unemp}} = 0.42$ that we obtained from the full sample. Thus we believe that there are important changes in wage volatility over the business cycle.

mean	5%	25%	50%	75%	95%
0.0088	0.0054	0.0074	0.0088	0.0102	0.0122

Table 2: Posterior distribution of difference in σ_t^2 between paired high- and low-unemployment years. Years are selected based on the highest and lowest unemployment rates between NBER peaks.

mean	5%	25%	50%	75%	95%
0.31	0.19	0.26	0.31	0.36	0.43

Table 3: Posterior distribution of $\Delta\sigma_t^2/\Delta u_t$ between paired high- and low-unemployment years. Years are selected based on the highest and lowest unemployment rates between NBER peaks.

[8] As we work with annual data, we treat 1980 to 1982 as a single recession.

3 Decomposition

We now investigate the forces that drive the positive correlation between wage volatility and the unemployment rate. In section 3.1 we layout our methodology for decomposing the correlation. The first key step in this decomposition is that we use observed covariates to partition the sample into groups within each year. We can then decompose the total variance of wage growth in a year into variances within groups and the variance between groups. Movements in the total variance over years are then driven by movements in these within- and between-group variances as well as shifts in the group sizes. The second key step is to model (statistically) the mean and variance of wage growth within each cell of the partition in each year. Doing so allows us to capture the posterior uncertainty about the components of the decomposition and therefore assess the uncertainty about the decomposition results themselves.

3.1 Methods

Suppose that we can partition individuals into J groups within each year based on observed covariates. When we apply this methodology below we form two groups according to whether an individual has experienced any unemployment spells in the preceding two years and so $J = 2$, but we choose to keep our discussion in general terms to emphasize the fact that this methodology could be applied to any partition of the data.

3.1.1 Decomposing the Correlation

We now explain our method for decomposing the correlation between wage volatility and the unemployment rate. Given our partition of the data, suppose we know the fraction of individuals in each group and the mean and variance for wage growth in each group. We can then construct the total variance of wage growth across all individuals as

$$\sigma_t^2 = \sum_{j=1}^{J} \pi_{j,t}\sigma_{j,t}^2 + \sum_{j=1}^{J} \pi_{j,t}(\overline{dw}_{j,t} - \overline{dw}_t)^2 \equiv W_t + B_t \tag{9}$$

where

$$\sigma_{j,t}^2 = \frac{1}{n_{j,t}} \sum_{i \in I_{j,t}} (dw_{i,t} - \overline{dw}_{j,t})^2$$

denotes the within-group variances, and $\overline{dw}_{j,t}$ is the within-group mean. In the above, $n_{j,t}$ is the number of individuals in group j in year t and $\pi_{j,t} = n_{j,t}/\sum_j n_{j,t}$ is the fraction of individuals in group j in year t. $I_{j,t}$ refers to those i in cell j in year t.

As it is customary, we call the two sums in equation (9) *within-* and *between-group* variances. There are two forces that can raise the contribution of the within-group variance. First, the group fractions, $\pi_{j,t}$ might shift towards groups with higher within-group variances. Second, the within-group variances might increase. In order to separate out these two effects, we further decompose the within-group variance term. We can write W_t as

$$W_t = \sum_j \bar{\pi}_j \bar{\sigma}_j^2 + \sum_j \bar{\pi}_j d\sigma_{j,t}^2 + \sum_j \bar{\sigma}_j^2 d\pi_{j,t} + \underbrace{\sum_j d\pi_{j,t} d\sigma_{j,t}^2}_{e_t} \tag{10}$$

where $\bar{\pi}_j$ and $\bar{\sigma}_j^2$ are time-series averages for $\pi_{j,t}$ and $\sigma_{j,t}^2$, respectively, and $d\pi_{j,t}$ and $d\sigma_{j,t}^2$ are deviations from these. Equation (10) is exact, but we can think of the last term as the second-order error term of a linear approximation, so we will denote it by e_t.

Our goal is to understand the correlation between σ_t^2 and the unemployment rate, call it u_t. To derive our decomposition, substitute equation (10) into equation (9) and take the covariance of both sides with respect to u_t to arrive at

$$\text{Cov}(\sigma_t^2, u_t) = \sum_j \text{Cov}\big(\bar{\pi}_j d\sigma_{j,t}^2, u_t\big) + \sum_j \text{Cov}\big(\bar{\sigma}_j^2 d\pi_{j,t}, u_t\big) + \text{Cov}(e_t, u_t) + \text{Cov}(B_t, u_t). \tag{11}$$

As we are interested in the *correlation* of σ_t^2 and u_t, we can divide both sides of equation (11) by the time-series standard deviations of σ_t^2 and u_t. The resulting expression decomposes the correlation coefficient into a sum of components. We find it useful, however, to normalize both sides by the correlation coefficient itself to express the decomposition in terms of the fraction of the correlation explained by each component. Therefore, we divide both sides of (11) by $\text{Cov}(\sigma_t^2, u_t)$. In the end we have

$$1 = \sum_j \underbrace{\frac{\text{Cov}\big(\bar{\pi}_j d\sigma_{j,t}^2, u_t\big)}{\text{Cov}(\sigma_t^2, u_t)}}_{\text{“}d\sigma_j^2\text{”}} + \sum_j \underbrace{\frac{\text{Cov}\big(\bar{\sigma}_j^2 d\pi_{j,t}, u_t\big)}{\text{Cov}(\sigma_t^2, u_t)}}_{\text{“}d\pi_j\text{”}} + \underbrace{\frac{\text{Cov}(e_t, u_t)}{\text{Cov}(\sigma_t^2, u_t)}}_{\text{error}} + \underbrace{\frac{\text{Cov}(B_t, u_t)}{\text{Cov}(\sigma_t^2, u_t)}}_{\text{between}}. \tag{12}$$

We use this equation to decompose the correlation of σ_t^2 and u_t. With this decomposition, we can study the contribution of each $d\sigma_{j,t}^2$ and each $d\pi_{j,t}$ to the correlation of σ_t^2 and u_t. However, since the group fractions are constrained to sum to one, it is not possible to consider changing just one $\pi_{j,t}$ in isolation. Therefore we analyze the sum of all the $d\pi_j$ terms to study their total contribution. Notice that the sum of the $d\pi_j$ terms is exactly the composition effect we mentioned above.

In order to calculate the right-hand side of equation (12), we need σ_t^2. Given our knowledge of the within-group means and variances and the fraction of individuals in each group, we use equation (9) to calculate σ_t^2.

3.1.2 Summarizing Uncertainty

Our decomposition can be applied directly to sample moments computed from the data. Using sample moments, however, does not give any sense of the uncertainty surrounding the results. To the extent that we partition the sample into small groups, our uncertainty about the sample moments can be substantial so it is important to account for this uncertainty.

In order to summarize the uncertainty surrounding the decomposition, we extend our statistical model to estimate the within-group means and variances. We then draw these means and variances from the posterior distribution and use these draws to compute the right-hand side of equation (12). With this procedure we can understand how our posterior uncertainty about the group means and variances translates into uncertainty about the sources of the correlation between σ_t^2 and u_t.

In this procedure, we get a new set of estimates for σ_t^2 based on the group-level estimates and equation (9). Since we partition the data, the hyperparameter regressions exert a slighly different influence on the variances compared to the results in section 2. However, we find that these differences are inconsequential for the correlation with the unemployment rate. In fact, we compute the correlation between σ_t^2 and then unemployment rate for both models and find the posterior mean correlation coefficients differ by just 0.0015.

3.1.3 Extending the Model

We now extend our model to the mean and variance of wage growth within each cell of the partition. For an individual i who is in cell j of the partition in year t, the extended model specifies the following distribution for wage growth

$$dw_{i,t} \sim \mathrm{N}\left(X_{i,t}\beta + \alpha_{j[i,t],t}, \sigma^2_{j[i,t],t}\right) \tag{13}$$

$$\alpha_{j,t} \sim \mathrm{N}\left(Z_t\gamma_j, v_{\alpha_j}\right) \quad \forall j \tag{14}$$

$$\sigma^2_{j,t} \sim \mathrm{N}\left(Z_t\delta_j, v_{\sigma^2_j}\right), \quad \sigma^2_{j,t} > 0 \quad \forall j. \tag{15}$$

Equation (13) states that the change in an individual's log wage, $dw_{i,t}$ is normally distributed. As before, the mean of this distribution depends on the individual's demographic characteristics and the coefficients on these demographic characteristics are assumed to be common across groups and across years. In addition, the mean change in wages varies over time and across groups with the $\alpha_{j[i,t],t}$ term. The notation $j[i,t]$ refers to the group index for the group that individual i is a member of at time t. We also allow the variance of the innovation in wages to vary over time and across groups as captured by the $\sigma^2_{j[i,t],t}$ term. Equations 14 and 15 model the within group means and variances, respectively. Again, these components are related to aggregate conditions captured by Z_t. Priors are independent for each group, and otherwise have the same form as in section 2.1.1.

3.2 Decomposing by Unemployment Experience

We apply our decomposition method to the PSID data partitioned by unemployment experience. In particular for an individual i at time t, $dw_{i,t}$ refers to wage growth between years $t-1$ and t. We assign an individual to the "no unemployment" group if this individual reports zero hours of unemployment for both years $t-1$ and t. Those reporting positive hours of unemployment in year $t-1$ or year t are assigned to the "unemployment" group.

As mentioned above, one advantage of the multilevel model is that it is able to partially pool information across years to produce sharper estimates than would be obtained without pooling. The role of equations (14) and (15) is to identify those years for which we expect the α's and σ's to be similar. In section 2 we allowed the model to capture comovement between the α's and σ's and the unemployment rate. In this context, however, we noticed that there is evidence of an upward trend in the variance within the unemployment group. Therefore we include a linear time trend in the vector of aggregate conditions, Z_t, for both groups.

3.2.1 Decomposition Results

Table 4 shows the results of our decomposition. The first row of the table shows the posterior mean for the fraction of the correlation explained by each component. Here, one can see that countercyclical fluctuations in the within group variances contribute 73% of the correlation with the unemployment group contributing the bulk of this (55%). Most of the remaining correlation comes from the composition effect that arises because the unemployment group has a larger variance at all times and the size of this group increases during a recession. This composition effect contributes 25% of the total. Finally, the error in the decomposition and the between variance contribute next to nothing. The table also shows quantiles for these fractions and one can see that the posterior uncertainty does not change the overall message.

	$d\sigma^2_{\text{no unemp}}$	$d\sigma^2_{\text{unemp}}$	composition	error	between
mean	0.1809	0.5518	0.2466	0.0026	0.0181
5%	0.0842	0.4739	0.2086	−0.0178	0.0047
25%	0.1438	0.5201	0.2284	−0.0052	0.0114
50%	0.1836	0.5508	0.2437	0.0027	0.0170
75%	0.2199	0.5837	0.2620	0.0108	0.0236
95%	0.2711	0.6318	0.2946	0.0220	0.0351

Table 4: Decomposition of correlation between wage volatility and the unemployment rate into components generated by within-group variances, between-group variances, the composition effect as group sizes change and a residual.

We view the results in Table 4 as our main results and we now present additional findings from the extended model to explain the forces that drive our results. Figure 3 plots our estimates of $\sigma^2_{j,t}$ for both groups. The top panels show $\sigma^2_{j,t}$ plotted against the unemployment rate and displayed on a common scale. From these plots, one can see that the comovement with the unemployment rate is much stronger in the unemployment group. While the covariance is low in the no unemployment group, this group is about four times the size of the no unemployment group and so this covariance is scaled up by a factor of about 0.8 when the term $d\sigma^2_{j,t}$ is multiplied by $\bar{\pi}_j$ in equation (12) vs 0.2 for the unemployment group.

Figure 3 also shows that the variance $\sigma^2_{j,t}$ is generally larger in the the unemployment group regardless of aggregate conditions. As the size of this group grows during a recession, the total variance, σ^2_t, increases. This is the source of the composition effect.

One way of understanding the benefit of using the multilevel model is to compare the posterior variance of the σ's to the posterior variance that would result without any pooling (e.g. from a model with a single level — we specify the details of this model in appendix B.3). This comparison shows how much posterior uncertainty is eliminated by partially pooling information across years. Figure 4 compares these variances year by year. For both groups, the multilevel model tightens our estimates of the $\sigma_{j,t}$. The fact that there is a larger benefit of pooling in the no unemployment group reflects the fact that there are fewer observations in this group and so information contained in the prior has a larger impact on the posterior.

4 Conclusion

Our results show that wage volatility increases with the unemployment rate and we have documented the role of unemployment episodes in generating this comovement. This analysis points to

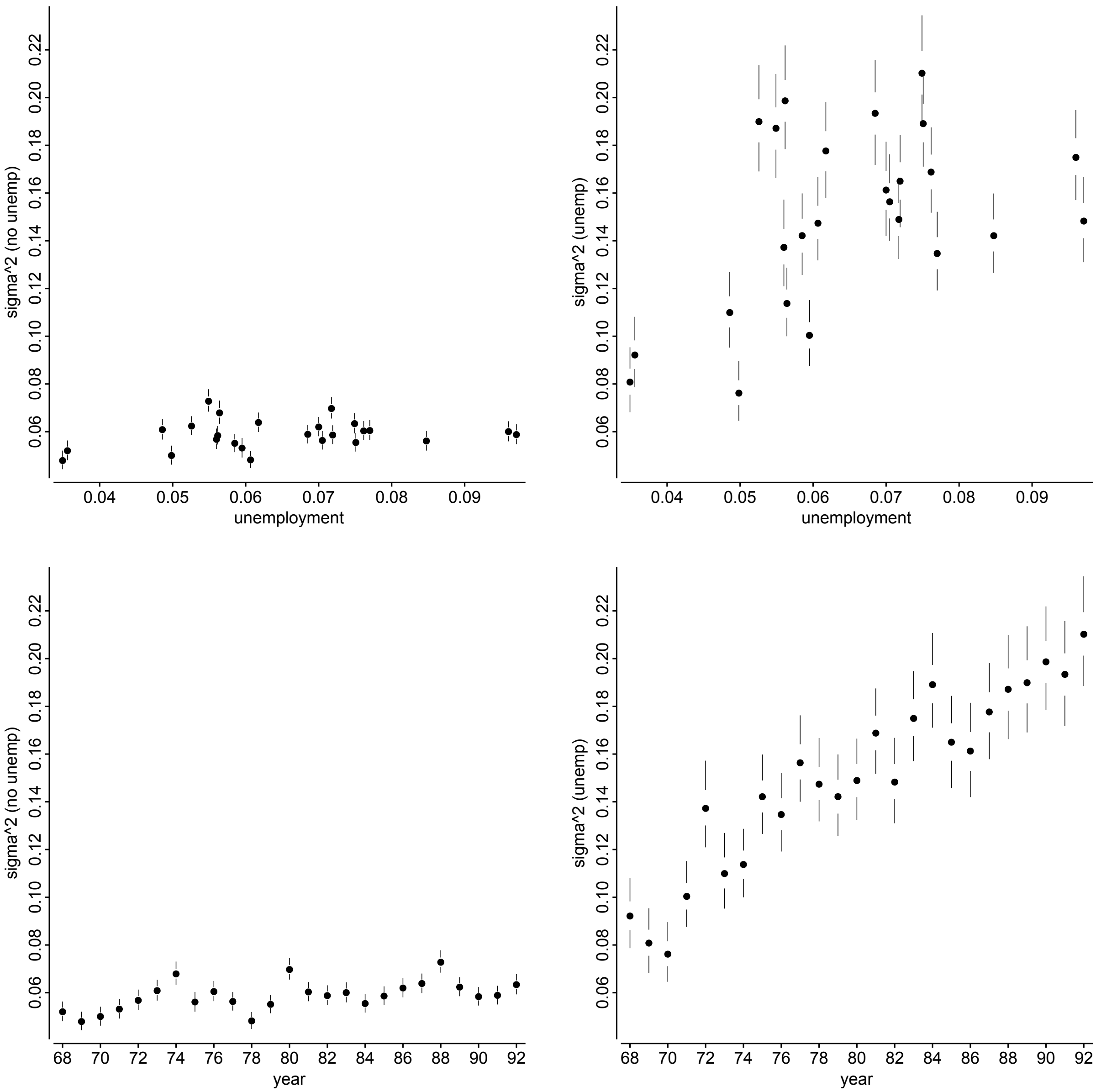
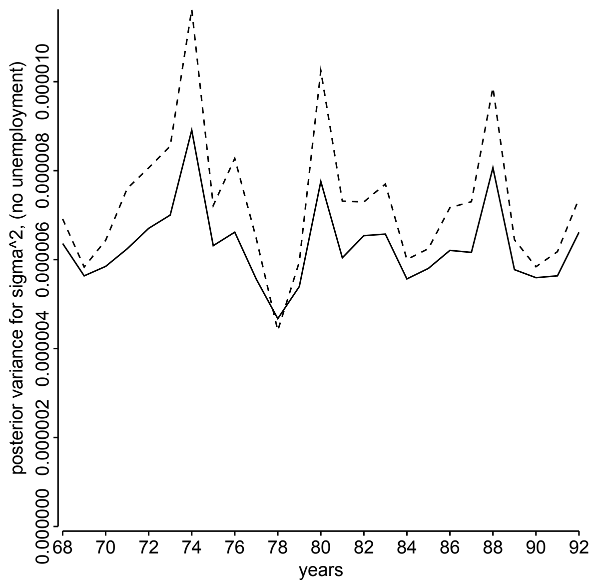
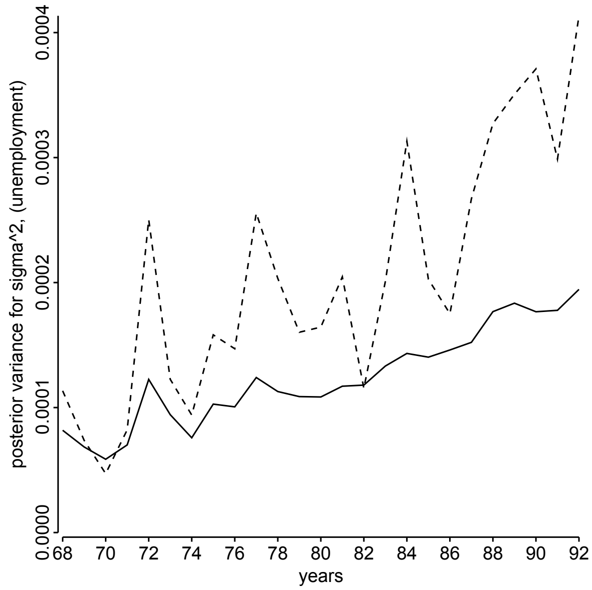


Figure 3: Posterior medians of $\sigma^2_{j,t}$ for no unemployment group (left panels) and unemployment groups (right panels) against the unemployment rate (top panels) and year (bottom panels). The vertical bars extend from the 25% to 5% quantiles and from the 75% to the 95% quantiles.



a) no unemployment group



b) unemployment group

Figure 4: Posterior variance of $\sigma_{j,t}$ by year for no unemployment group (top panel) and unemployment group (bottom panel). Each panel has two lines with the dashed line showing the posterior variance without pooling and the solid line showing the posterior variance from the multilevel model.

cyclical changes in the volatility of wages among those experiencing unemployment as an important source of countercyclical wage volatility.

The methods we have used in this paper can be applied to different partitions of the data to decompose this variance further, which may help us understand the economic mechanisms behind these findings. While our ability to partition the data is limited by the number of observations within cells, using multilevel modeling techniques to appropriately pool information across cells and switching to larger datasets will allow us and other researchers to study finer partitions of the data.

The persistence of wage shocks is an important determinant of the ability of households to self insure against these shocks. In this paper, we have assumed that wages are a random walk. To investigate the persistence of these shocks and how the persistence of shocks might change over the business cycle, our methods would have to be extended to include a richer time-series model of wage dynamics.

Finally, search-theoretic models of the labor market provide a set of predictions for wage distributions and the impact of unemployment spells on wages and these models have now been extended to incorporate aggregate fluctuations. Reconciling our findings with these predictions would bring us closer to understanding the theoretical links between aggregate and idiosyncratic risk.

Appendix

A Data Appendix

From the PSID, we use data on the annual labor income of male heads and annual hours of work to construct annual wages. In addition, we use data on age and education as covariates. We only include those heads that are between 25 and 60 years of age in both years over which the change in wages is calculated. We drop those with allocated labor income, students, business owners, self-employed individuals, and those with zero hours or income. We trim the top 1% of the income distribution in each year to remove the effect of changes in top-codes across years. Finally, we drop those with wages less than half of the federal minimum wage in that year and we drop those who work fewer than 320 hours in a given year. Our results on income changes in section 2.3 are based on the same sample, but we use annual labor income without dividing by annual hours.

To construct the unemployment-experience partition, we use data on annual hours of unemployment. Let H_t^U be the hours of unemployment in year t. The unemployment group at time t includes those individuals who report $H_t^U > 0$ or $H_{t-1}^U > 0$. The no-unemployment group is those individuals for whom H^U is equal to zero for both years.

For aggregate data, we use the national unemployment rate reported by the BLS and take the average value of the monthly series within each year.

B Methodological Appendix

B.1 The Gibbs sampler

Sampling from the posterior of our model is straightforward and follows Markov Chain Monte Carlo techniques that are commonly used in Bayesian statistics. There are several alternative methods for sampling from the posterior (see Gelman et al. (2004) for a summary). We found that a block Gibbs sampler is relatively fast, easy to implement and has good mixing properties (see Section B.2 for a discussion of convergence). A brief summary of the Gibbs sampler we used is given below.

Our model can be summarized by equations (13), (14), and (15). This model collapses to that in section 2 if one sets $J = 1$ with $j[i, t] = 1$ for all i and t. We partition the parameter space into blocks corresponding to β, α, γ, v_α, σ^2, δ, and v_σ^2.

B.1.1 Conditional posterior sampling for β, α, γ, and δ

Each of these parameters can be thought of as the coefficient η in a linear regression

$$b \sim \mathrm{N}(A\eta, \Phi) \tag{16}$$

with a known variance matrix $\Phi = \mathrm{diag}(\phi)$ and a given prior $\eta \sim \mathrm{N}(\Pi)$. Table 5 shows the mapping between (16) and the model parameters. The conditional posterior distribution is normal, and can be constructed and sampled from in a straightforward manner (Gelman et al., 2004, Sections 14.6 and 14.8).

η	b	A	ϕ	Π
β	$\left[dw_{i,t} - \alpha_{j[i,t],t}\right]_{i,t}$	$\left[X_{i,t}\right]_{i,t}$	$\left[\sigma^2_{j[i,t],t}\right]_{i,t}$	flat
α_j	$\left[dw_{i,t} - X_{i,t}\beta\right]_{i,t:j[i,t]=j}$	$\mathbf{1}$	$\left[\sigma^2_{j[i,t],t}\right]_{i,t:j[i,t]=j}$	v_α
γ_j	α	Z	$v_{\alpha,j}$	flat
δ_j	σ^2	Z	$v_{\sigma^2,j}$	flat

Table 5: Conditional posterior sampling for β, α, γ, and δ

B.1.2 Conditional posterior sampling for the group-level variances v_α and v_{σ^2}

As is well known, if $h_k \sim \mathrm{N}(0, v)$ (iid, $k = 1, \ldots, K$), then the likelihood is

$$p(v \mid h) \propto v^{-K/2} \exp\left(-\frac{\sum_{k=1}^K h_k^2}{2v}\right)$$

As discussed above (see Section 2.1.1), our prior for hyperparameter variances is $p(v) \propto v^{-1/2}$.

Then we sample from the conditional posterior using

$$v \sim \text{Inverse-Gamma}\left(\frac{K-1}{2}, \frac{\sum_{k=1}^K h_k^2}{2}\right) \tag{17}$$

Table 6 shows the correspondences between (17) and the model parameters.

h	v
$\alpha_{j,\cdot} - Z\gamma_j$	$v_{\alpha,j}$
$\sigma^2_{j,\cdot} - Z\delta_j$	$v_{\sigma^2,j}$

Table 6: Conditional posterior sampling for v_α, v_σ^2

B.1.3 Conditional posterior sampling for σ^2

Let us fix j and t, and only consider the observations in $I_{j,t} = \{(i',t') : t' = t, j[i',t'] = j\}$. Given (8) and (15), the conditional posterior for $\sigma^2_{j,t}$ is

$$p(\sigma^2_{j,t} \mid r, \alpha_{j,t}, Z_t, \delta_j) \propto \left(\sigma^2_{j,t}\right)^{-\frac{|I_{j,t}|}{2}} \exp\left(-\frac{\sum_{(i,t)\in I_{j,t}} (dw_{i,t} - X_{i,t}\beta - \alpha_{j,t})^2}{2\sigma^2_{j,t}}\right) \exp\left(-\frac{(\sigma^2_{j,t} - Z_t\delta_j)^2}{2v_{\sigma^2,j}}\right) 1_{\sigma^2_{j,t}\geq 0} \quad (18)$$

This does not correspond to any commonly used family of probability distributions, so we can only sample from it using general tools. After experimenting with rejection methods and obtaining poor acceptance rates, we settled on the slice sampling algorithm of Neal (2003) with excellent results.[9]

B.2 Convergence of the posterior sampler

We monitor the convergence of the Gibbs sampler by calculating the univariate potential scale reduction factor (PSRF) for each parameter value (Gelman and Rubin, 1992; Brooks and Gelman, 1998; Gelman et al., 2004). The PSRF uses variances within and between the parallel chains to estimate the factor by which the scale of the current posterior distribution for a given parameter might be reduced if we were to obtain a sample of infinite size. In practice, values below 1.1 are acceptable, unless very high precision is required.

Calculating the PSRF as the chain evolves for the second half of the chain is also a good way to monitor mixing. Figure 5 shows the evolution of the PSRF, which suggests that the mixing is excellent. We found that the mixing was greatly enhanced by subtracting the column means from X.

[9]In particular, we used the "stepping out" variant of the algorithm from Neal (2003, Figures 3 and 5).

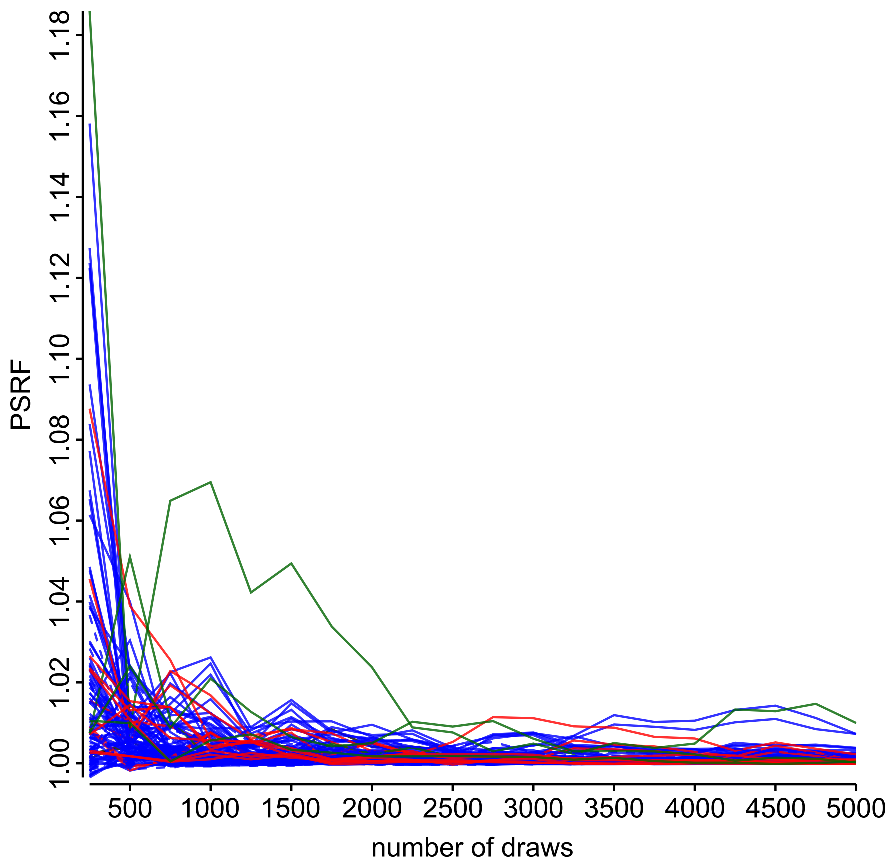


Figure 5: Evolution of univariate potential scale reduction factor for the model with 2 subgroups, using 3 parallel chains. *Solid* black: β, blue: α, red: γ, dark green: v_α. *Dashed* blue: σ^2, red: δ, dark green: v_{σ^2}.

B.3 The non-hierarchical model

We estimate a version of the model without hierarchical regressions for comparison.

$$dw_{i,t} \sim \mathrm{N}\big(X_{i,t}\beta + \alpha_{j[i,t],t}, \sigma^2_{j[i,t],t}\big)$$

$$p(\beta) \propto 1$$

$$p(\alpha_{j,t}) \propto 1 \; \forall j, t$$

$$p(\sigma^2_{j,t}) \propto (\sigma^2_{j,t})^{-1} \; \forall j, t$$

It is very straightforward to sample from the posterior of this model: first we sample $\beta \mid \alpha, \sigma^2, X$ (see Section B.1.1), then sample from $\alpha_{j,t}, \sigma^2_{j,t} \mid \beta, X$ by sampling from the posterior of the regression of $dw_{i,t} - \alpha_{j[i,t],t}$ on 1 with unknown variance and a reference prior for each j, t.

C Comparison to Storesletten et al. (2004)

The purpose of this appendix is to express Storesletten et al.'s results in the same terms as ours to show that the two are not vastly at odds with one another. Throughout, we take their results reported in their Table 2, Panel E because these results are calculated on the assumption that business cycles are defined by the unemployment rate as opposed to GNP growth or NBER cycles and thus closest to our work.

Storesletten et al. specify the following process for the residual of log earnings of individual i

$$u_{it} = \alpha_i + z_{it} + \varepsilon_{it}$$

$$z_{it} = \rho z_{i,t-1} + \eta_{it},$$

where $\alpha_i \sim \text{Niid}(0, \sigma_\alpha^2)$, $\varepsilon_{it} \sim \text{Niid}(0, \sigma_\varepsilon^2)$, and $\eta_{it} \sim \text{Niid}(0, \sigma_E^2)$ in an expansion and $\eta_{it} \sim \text{Niid}(0, \sigma_C^2)$ in a contraction. In this notation, our interest is in computing the variance of Δu_{it}, which is

$$\text{Var}\left[\Delta u_{it}\right] = (\rho - 1)^2 \text{Var}\left[z_{i,t-1}\right] + \text{Var}\left[\eta_{it}\right] + \text{Var}\left[\Delta \varepsilon_{it}\right]$$

Clearly, as $\rho \to 1$, the first term on the right-hand side goes to zero. This is relevant because Storesletten et al. estimate ρ to be close to 1. For now, suppose $\rho = 1$, but we return to the issue below. Since ε is distributed identically over time, the $\text{Var}\left[\Delta \varepsilon_{it}\right]$ term is a constant. Thus, the difference in $\text{Var}\left[\Delta u_{it}\right]$ between an expansion and a contraction is just the difference in $\text{Var}\left[\eta_{it}\right]$ or $\sigma_C^2 - \sigma_E^2$. Using Storesletten et al.'s estimates, this difference is $0.246^2 - 0.138^2 = 0.041$. In our back of the envelope calculation in section 2.3 we found a difference of 0.013 for wages and 0.034 for earnings.

In the calculation above we ignored the term $(\rho - 1)^2 \text{Var}\left[z_{i,t-1}\right]$, which will be countercyclical as the variance of z is countercyclical. As argued above, this term is small. To see this, consider two extreme economies, one that is always in an expansion and one that is always in a contraction. The unconditional variance of z in the expanding economy is then $\sigma_E^2/(1-\rho^2)$ and one can similarly calculate the unconditional variance of z in the contracting economy. The difference between these two variances is an upper bound on the cyclical fluctuations in $\text{Var}(z)$. Using this upper bound, we conclude that the contribution of the term $(\rho - 1)^2 \text{Var}\left[z_{i,t-1}\right]$ is at most 0.0013 or 3% of the 0.041 figure we found above.

References

Brooks, S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–533.

Gelman, A., Carlin, B., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall, 2nd edition.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Gottschalk, P., Moffitt, R., Katz, L., and Dickens, W. (1994). The growth of earnings instability in the US labor market. *Brookings Papers on Economic Activity*, 1994(2):217–272.

Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the united states: 1967-2006. *Review of Economic Dynamics*, 13(1):15–51.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *The American Economic Review*, 96(3):461–498.

Menzio, G. and Shi, S. (2009). Efficient search on the job and the business cycle. Working Paper 14905, National Bureau of Economic Research.

Menzio, G. and Shi, S. (2010). Block recursive equilibria for stochastic models of search on the job. *Journal of Economic Theory*, 145(4):1453 – 1494.

Moscarini, G. and Postel-Vinay, F. (2010). Stochastic search equilibrium. Discussion Paper CFDP 1754, Cowles Foundation, Yale University.

Neal, R. (2003). Slice sampling. *Annals of Statistics*, 31(3):705–741.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy*, 112(3):695–717.

Western, B. and Bloome, D. (2009). Variance function regressions for studying inequality. *Sociological Methodology*, 39(1):293–326.