A Measure of Core Wage Inflation
Martin Almuzara, Richard Audoly, and Davide Melcangi
Federal Reserve Bank of New York Staff Reports, no. 1067
July 2023
JEL classification: C33, E24

Abstract
We recover the persistent (“core”) component of nominal wage growth over the past twenty-five years in the United States. Our approach combines worker-level data with time-series smoothing methods and can disentangle the common persistence of wage inflation from the persistence specific to some subgroup of workers, such as workers in a particular industry. We find that most of the business cycle fluctuations in wage inflation are persistent and driven by a common factor. This common persistent factor is particularly important during inflationary periods, and it explains 75 to 90 percent of the post-pandemic surge in wage inflation.

Key words: wage inflation, persistence, factor models
1 Introduction

The post-pandemic surge in inflation has renewed interest in the evolution of aggregate nominal wages.\textsuperscript{1} Because wage inflation is perceived as tightly linked to the evolution of prices, it is one of the key indicators monitored for the conduct of monetary policy.\textsuperscript{2} It also provides a signal about the state of the labor market, and it is an important input for households’ and firms’ decisions. But which measure of wage inflation is appropriate for these purposes? While extensive work has been done to define measures of price inflation that are purged from noise and short-run fluctuations, surprisingly little research has been devoted to the underlying dynamics of aggregate nominal wage growth.\textsuperscript{3}

This paper describes a framework to isolate the persistent (“core”) latent component of wage inflation.\textsuperscript{4} Our approach combines worker-level data with time series filtering and smoothing methods. We estimate a dynamic factor model with time-varying parameters, in which wage inflation is the sum of a persistent component common to all workers, a persistent component specific to some subgroup of workers, and transitory shocks. We find that most of the business cycle fluctuations in wage inflation are persistent and that the evolution of this persistent component is driven by its common component. This common latent factor is especially predominant during the wage inflation surge of 2021, regardless of the various cross-sections we consider, such

\textsuperscript{1}In the remainder of the paper, we use the terms “nominal wage growth” and “wage inflation” interchangeably.

\textsuperscript{2}In a recent speech, Fed Chairman J.H. Powell declared: [...] “for wage growth to be sustainable, it needs to be consistent with 2 percent inflation” [Powell, 2022].

\textsuperscript{3}See Stock and Watson [2016] for a summary of the price inflation literature. On the evolution of aggregate wages, Daly, Hobijn, and Wiles [2011] examine the time-varying cyclicality of real wages. Huh and Trehan [1995] estimate a vector error correction model with prices, wages and productivity. A broader literature studies the link between nominal wage growth and price inflation, which is beyond the scope of our paper. For recent work on this, see Kiley [2023].

\textsuperscript{4}Similarly to measures of core price inflation, our goal is to capture the persistent part of inflation by abstracting from its more volatile components. As we describe in the paper, our measure is conceptually distinct from core price inflation, which simply excludes food and energy.
as industry and education.

Our analysis employs monthly worker-level information on hourly wages from the Current Population Survey (CPS). Our statistical model lends itself naturally to this dataset as it allows us to break down aggregate changes in nominal wages across a large set of cross-sectional variables. In line with the price inflation literature, our main focus is on the industry cross-section, but we also estimate the model using other cross-sections based on demographic and geographic variables. Although observed each month, nominal wage growth in our worker-level data is measured as a 12-month rate. One of the contributions of our paper is to build a model that explicitly accounts for the monthly-to-yearly temporal aggregation, thus extending the framework that Stock and Watson [2016] previously applied to price data, and allowing us to retrieve the persistent component of unobserved monthly growth.\(^5\) This component is our key object of interest, and we refer to it as Core CPS Wage Inflation (CoWI).

To validate our approach as a successful characterization of persistent wage dynamics, we conduct a real time forecast comparison against a random walk benchmark. These forecasts are known to be hard to improve upon in the price inflation literature (see, e.g., Atkeson and Ohanian [2001], Stock and Watson [2008])). We find that our model outperforms the random walk alternative, especially at the 3–month horizon. At the industry level, our model is substantially more accurate than the random walk at all horizons. In addition, we find that changes in CoWI are significantly correlated with changes in the unemployment rate, and that this relationship is stronger than for raw aggregate wage growth. This finding confirms that CoWI, by purging transitory variation, can provide reliable signals for turning points in wage inflation.

Armed with this model, we set out to empirically document the dynamics of wage inflation and its drivers, thus making progress on a relatively understudied topic.

\(^5\)This methodological contribution is novel and potentially relevant for the empirical analysis of other datasets subject to temporal aggregation.
As a first empirical exercise, we partition our sample by broad industry groups and estimate our model to recover the dynamics of persistent wage growth. While wage growth is characterized by substantial monthly variation, the largest movements in CoWI coincide with the episodes of greatest macroeconomic significance in our sample. Indeed, CoWI falls substantially during the 2001 and 2008 recessions, and it rises sharply at the beginning of 2021, the onset of the post-pandemic inflation surge episode.

A key empirical question is how strongly correlated nominal wage growth is across workers. The size of the cross-sectional correlation provides a measure of the relative importance of economy-wide versus idiosyncratic shocks in driving wage inflation. Our model draws an explicit distinction between common and sector-specific components in CoWI and is, therefore, particularly suitable to make this assessment. We document that major fluctuations in CoWI cannot be ascribed to specific industries. In fact, most changes in the persistent component of nominal wage growth are common across industries, especially during the recent pandemic surge. By contrast, the estimated sector-specific component appears to capture lower frequency movements and has been declining since the 1990s, albeit tentatively surging after COVID-19.⁶

In the last part of the paper, we document that our findings are robust to alternative cross-sectional variables that represent likely sources of heterogeneity in nominal wage growth. We re-estimate the model many alternative worker characteristics to partition our sample, such as race, gender, education, age, and other demographics. Regardless of the cross-section we consider, CoWI still accounts for a substantial share of the large swings in aggregate wage growth. Most importantly, we find that none of these alternative cross-sections is quantitatively important for the evolution of CoWI, but

⁶In an early related contribution, Watson and Engle [1981] find that a highly persistent single common factor explains most of the variation across industries using wage data for the metropolitan area of Los Angeles. Ahn, Chen, and Kister [2020] use a dynamic factor model on quarterly data from the Current Employment Survey (CES) to extract the common component of inflation in average hourly earnings across a large number of industries. In addition to the data, an important difference is that CoWI measures both common and sector-specific persistent components in wage inflation while their measure targets only the common components (both persistent and transitory).
instead that a common latent factor drives most of its evolution.

We formally quantify the importance of the common component of CoWI for the three episodes of greatest macroeconomic significance in our sample: the 2001 and 2008 recessions, and the wage inflation surge of 2021. Common factors explain a quantitatively large share of the fall in CoWI during the two recessions. This is particularly true for the Great Recession. For example, industry or region–specific factors account for less than 20% of the fall in CoWI in 2008-09. By comparison, the common component of CoWI accounts for between 45% and 85% of the 2008-2009 drop across the various cross-sections we consider.

The estimated common factor matters even more during the 2021 inflation surge. We find that it accounts for at least 80% of the increase in CoWI over this period, regardless of the cross-sectional variable we use in the estimation. This finding suggests that a form of asymmetry might be at play between recessionary and inflationary episodes, with potentially more worker heterogeneity in downward than in upward nominal wage rigidity. Importantly, our results indicate that this heterogeneity, if present, is not of first-order importance for characterizing wage inflation episodes at the macro level: the common component is the main driver of CoWI during inflationary episodes. Additional estimates using data going back to the 1960s suggest that our results are not specific to the post-pandemic period: the common factor also explains the majority of CoWI fluctuations during the inflationary episodes of the 1970s.

The rest of the paper proceeds as follows. Section 2 describes the data. Section 3 describes the model and estimation method. Section 4 reports the estimation results, and Section 5 concludes.
2 Data

We use monthly data on nominal wages from the Current Population Survey (CPS). An objective of this paper is to propose a reliable measure of wage inflation that can be monitored on a regular basis. Using data from the CPS has two key advantages in that regard. Compared to proprietary data, such as the ADP data used in Grigsby, Hurst, and Yildirmaz [2021], the CPS goes back more than two decades, is representative of the US population, and is publicly available. Compared to other measures of nominal wage growth, such as the Employment Cost Index (ECI), the CPS is available on a monthly basis.

We follow the well-established approach of the Atlanta Fed Wage Growth Tracker and define wage inflation as the median growth in the hourly wage of individuals observed twelve months apart [Atlanta Fed, 2023]. Letting \( t \) denote a month, \( i \) denote some subgroup of the data, the 12-month rate of wage inflation \( w_{i,t} \) in subgroup \( i \) is defined as

\[
\ln W_{i,j50,t} - \ln W_{i,j50,t-12}
\]

where \( j50 \) denotes the individual with median 12-month wage growth in group \( i \) and \( W \) denotes hourly wages.

The rationale behind considering wage changes within individual is to measure changes in the price of a homogeneous unit of labor. Similarly to standard measures of price inflation, which aim to measure price changes in the same goods, the goal is

\(^7\)We use data between 1997m1 and 2023m2. Our measure can be updated each month when the CPS is released. We start our sample in 1997 because there are gaps in the wage variables we are using in earlier periods.

\(^8\)We show in Appendix D that our results are little affected, albeit with less precise estimates, if we use average growth. The hourly wage is for the person’s “usual” weekly earnings and “usual” hours, which include the variable part of compensation (overtime pay, tips, commissions).
to abstract from changes in the composition of labor demand.\textsuperscript{9} One solution to this problem is to consider wage changes within narrowly defined job types. This is the approach underlying the ECI.\textsuperscript{10} The within-worker CPS definition we use can be seen as the worker side counterpart to that within-job approach.

We consider individual wage changes from observations twelve months apart due to the interview structure of the CPS. This directly adjusts for seasonal factors. But it also implies that this measure of wage inflation tends to lag actual month-on-month changes. To see this, consider the case of a worker with no month-on-month wage growth for eleven months and 0.1\% in the twelfth month. Their twelve-month apart wage change is 0.1\%, but their annualized month-on-month change in wages, the object of interest to monitor wage inflation, is $12 \times 0.1\% = 1.2\%$. We fully account for this time aggregation lag in the measurement framework introduced in Section 3.

A particularly attractive feature of building this measure from worker-level data is that aggregate wage inflation can be decomposed by both job and demographic characteristics. For most of Section 4, we break down aggregate changes in nominal wages by industry. We consider the seven broad industry groups defined by the Atlanta Fed Wage Growth Tracker, which ensures that the sample size within each cell is not too small: Construction and Mining, Education and Health, Finance and Business Services, Leisure and Hospitality, Manufacturing, Public Administration, and Trade and Transportation. This classification refers to workers’ current industry, irrespective of their industry in their previous interview. In Section 4.2, we split the sample by many other worker characteristics, following definitions by the Atlanta Fed Wage Growth Tracker: occupation, education, region, wage quartile, and age. Appendix A.1

\textsuperscript{9}We show in Appendix A.2 that alternative measures of nominal wage growth that do not account for composition changes are markedly more volatile. This also applies to alternative nominal wage growth concepts derived from the CPS, such as growth in the average hourly wage measured in each month.

\textsuperscript{10}A job type in the ECI is a small sector-industry-occupation cell. See here for definitions: \url{https://www.bls.gov/opub/hom/ncs/home.htm}. 
discusses the definition of these variables in detail.

3 Model and estimation

We describe our modeling approach and our proposed measure of core wage inflation in this section. A challenge posed by the data we analyze is that nominal wage growth, although observed every month, is measured as a 12-month rate. We are interested in the persistence of monthly wage growth and we therefore need to explicitly account for the monthly-to-yearly temporal aggregation.

Let \( w_{it} \) be the 12-month rate of wage inflation for sector \( i \) and month \( t \) defined in (1). We decompose \( w_{it} \) into the contribution of a persistent (or trend) component and a transitory (or noise) component:

\[
(2) \quad w_{it} = \frac{1}{12} \sum_{\ell=1}^{12} \tilde{\tau}_{i,t+1-\ell} + \tilde{\varepsilon}_{it}.
\]

In Equation (2), \( \tilde{\tau}_{it} \) is the persistent component for the unobserved monthly (annualized) rate of wage growth, while \( \tilde{\varepsilon}_{it} \) is the transitory error in the observed 12-month rate. To capture the potential cross-sectional correlation across sectors, we further decompose trend and noise into the sum of common (indexed by \( c \)) and sector-specific components:

\[
\tilde{\tau}_{it} = \alpha_{\tau,ct} \tau_{ct} + \tau_{it},
\]

\[
\tilde{\varepsilon}_{it} = \alpha_{\varepsilon,ct} \varepsilon_{ct} + \varepsilon_{it},
\]

and we model trends as random walks and transitory errors as low-order moving averages,

\[
\tau_{ct} = \tau_{c,t-1} + \sigma_{\Delta \tau,ct} \eta_{\Delta \tau,ct},
\]
$$\tau_{it} = \tau_{i,t-1} + \sigma_{\Delta \tau, it} \eta_{\Delta \tau, it},$$

$$\varepsilon_{ct} = (1 + \theta_c L + \cdots + \theta_{cp} L^p) \sigma_{\varepsilon, ct} \eta_{\varepsilon, ct},$$

$$\varepsilon_{it} = (1 + \theta_i L + \cdots + \theta_{iq} L^q) \sigma_{\varepsilon, it} \eta_{\varepsilon, it}.$$  

Here $L$ is the lag operator and $\eta_{\Delta \tau, ct}, \eta_{\Delta \tau, it}, \eta_{\varepsilon, ct}, \eta_{\varepsilon, it}$ are serially and mutually independent $N(0, 1)$, and independent of the loadings $\{\alpha_{\tau, it}, \alpha_{\varepsilon, it}\}_{i=1}^n$ and volatilities $\sigma_{\Delta \tau, ct}, \{\sigma_{\Delta \tau, it}\}_{i=1}^n, \sigma_{\varepsilon, ct}, \{\sigma_{\varepsilon, it}\}_{i=1}^n$.

**Time aggregation** The monthly-to-yearly transformation is an important feature of our framework and a key difference with models of price inflation such as Stock and Watson [2016] in which monthly rates of change are observed directly. To interpret what we do, write

$$\frac{W_{i,j,50,t}}{W_{i,j,50,t-12}} = \frac{W_{i,j,50,t}}{W_{i,j,50,t-1}} \times \frac{W_{i,j,50,t-1}}{W_{i,j,50,t-2}} \times \cdots \times \frac{W_{i,j,50,t-11}}{W_{i,j,50,t-12}}.$$  

Consider the case $\tilde{\varepsilon}_{it} = 0$ with no noise component in (2). Then,

$$\tilde{\tau}_{it} = 12 \ln \left( \frac{W_{i,j,50,t}}{W_{i,j,50,t-1}} \right) \approx \left( \frac{W_{i,j,50,t} - W_{i,j,50,t-1}}{W_{i,j,50,t-1}} \right)^{12},$$

that is, $\tilde{\tau}_{it}$ would measure the unobserved monthly annualized growth in median wage.

In practice, $W_{i,j,50,t}$ is an estimate based on survey data and expected to contain measurement error. Due to the sampling design of the CPS, the workers who contribute to $W_{i,j,50,t}/W_{i,j,50,t-12}$ and to $W_{i,j,50,s}/W_{i,j,50,s-12}$ differ whenever $t \neq s$. If workers are sampled at random, the measurement error in $w_{it}$ and $w_{is}$ will then be independent for $t \neq s$. The measurement error in $w_{it}$ will be independent of that in $w_{jt}$ for $j \neq i$ by a similar argument. This suggests having no common transitory error and white noise dynamics for the sector-specific transitory components. For added flexibility, we do include $\varepsilon_{ct}$ and we allow $\varepsilon_{it}$ to be serially dependent through $q > 0$ and the presence of time-
varying volatility. Consistent with interpreting $\tilde{\epsilon}_{it}$ as measurement error, we find $\epsilon_{ct}$ to be small and $\theta_{it}$ to be close to zero. Importantly, $\sigma_{\epsilon_{it}}$ appears to evolve in proportion to the inverse of the square root of the number of workers surveyed each month, which is exactly what one would expect if $\tilde{\epsilon}_{it}$ represented the error in estimating median wage growth in the population using its sample counterpart.\(^{11}\)

What about genuine transitory shocks to monthly wage growth? To accommodate them we would either enrich $\tilde{\tau}_{it}$ to be itself the sum of persistent and transitory parts or model $\tilde{\epsilon}_{it}$ as an MA(12). In Appendix D we show that the estimates from this extended model are similar to the more parsimonious model with random walk $\tau_{ct}, \{\tau_{it}\}_{i=1}^{n}$ and low $p,q$. Given the robustness of our results, the empirical analysis reported in Section 4 is based on the more parsimonious model introduced above.

**Objects of interest** In addition to the sector-level trend $\tilde{\tau}_{it}$ we are interested in the aggregate wage growth trend that we define for sectoral shares $\{s_{it}\}_{i=1}^{n}$ as

\[
\bar{\tau}_{t} = \sum_{i=1}^{n} s_{it} \tilde{\tau}_{it} = \left( \sum_{i=1}^{n} s_{it} \alpha_{i,t} \right) \tau_{ct} + \left( \sum_{i=1}^{n} s_{it} \tau_{it} \right).
\]

We thereafter refer to $\bar{\tau}_{t}$ as Core CPS Wage Inflation (CoWI).

Similar to sector-level trends, CoWI is driven both by common persistence across sectors and by a weighted average of sector-specific trend movements. We report on this decomposition in Section 4.

A technical point to consider is that the contribution of common and idiosyncratic components to the *level* of the trend is only pinned down by an arbitrary normalization.\(^{12}\) The contributions of common and idiosyncratic components to the *changes* in

\(^{11}\)To be more precise, if $n_{it}$ is the number of workers surveyed during month $t$ and $\tilde{\sigma}_{\epsilon_{it}}$ is the (posterior median estimate of the time-varying) standard deviation of the weighted-average transitory error $\tilde{\epsilon}_{it} = \sum_{i=1}^{n} s_{it} \tilde{\epsilon}_{it}$, we find that most pairs $(n_{t}, \tilde{\sigma}_{\epsilon_{it}})$ lie remarkably close to a line with no intercept, $\tilde{\sigma}_{\epsilon_{it}} = 14.1 / \sqrt{n_{t}}$, and that $\text{corr}(1 / \sqrt{n_{t}}, \tilde{\sigma}_{\epsilon_{it}}) = 0.9$. See Appendix D for further discussion.

\(^{12}\)This is because the locations of $(\alpha_{i,t}, \tau_{ct})_{i=1}^{T}$ and $(\tau_{it})_{i=1}^{T}$ can be changed by simply adding and sub-
the trend, on the other hand, are not subject to the same caveat. Hence, we focus on changes as opposed to levels when discussing this decomposition below.

**Time-varying parameters** In specifying the dynamics of the loadings and (log) volatilities, we follow Del Negro and Otrok [2008] and Stock and Watson [2016] and model them as random walks with small variances, i.e.,

\[
\Delta \alpha_{m,it} = \gamma'_{\alpha,m,i} \nu_{\alpha,m,it}, \quad m = \tau, \varepsilon, \quad i = 1, \ldots, n,
\]

\[
\Delta \ln \sigma_{m,jt}^2 = \gamma'_{\sigma,m,j} \nu_{\sigma,m,jt}, \quad m = \Delta \tau, \varepsilon, \quad j = 1, \ldots, n,
\]

with innovations \(\nu_{\alpha,\tau,\tau,it}, \nu_{\alpha,\varepsilon,\tau,it}, \nu_{\sigma,\Delta\tau,\tau,it}, \nu_{\sigma,\varepsilon,\Delta\tau,it}\) serially and cross-sectionally independent \(N(0,1)\). This is a standard approach to allow for parameters that drift slowly over time. It can accommodate trends in the correlations across sectors and noisiness of the different series.

**Estimation** Our model has data \(w_t = \{w_{it}\}_{i=1}^n\), time-invariant parameters

\[
\theta = (\{\theta_{ct}\}_{1 \leq t \leq p}, \{\theta_{it}\}_{1 \leq t \leq q, 1 \leq i \leq n}), \quad \gamma = (\{\gamma'_{\alpha,m,i}\}_{m=\tau,\varepsilon, 1 \leq i \leq n}, \{\gamma'_{\sigma,m,j}\}_{m=\Delta\tau,\varepsilon, 1 \leq j \leq n}),
\]

time-varying parameters

\[
\lambda_t = (\{\alpha_{\tau,it}\}, \alpha_{\varepsilon,it}, \{\sigma_{\Delta\tau,ct}\}, \{\sigma_{\varepsilon,ct}\})
\]

and latent variables

\[
\xi_t = (\tau_{ct}, \varepsilon_{ct}, \{\varepsilon_{it}\}_{i=1}^n).
\]

tracting an arbitrary constant without affecting the data. We impose \(\tau_{c1} = 0\) to eliminate this ambiguity. Similarly, \(\alpha_{\tau,it}\tau_{ct}\) and \(\alpha_{\varepsilon,it}\varepsilon_{ct}\) are invariant to multiplying \(\alpha_{\tau,it}\) or \(\alpha_{\varepsilon,it}\) by a constant and dividing \(\tau_{ct}\alpha_{\Delta\tau,ct}, \varepsilon_{ct}\sigma_{\varepsilon,ct}\) accordingly. Hence we fix \(\sigma_{\Delta\tau,ct} = \sigma_{\varepsilon,ct} = 1\). See Del Negro and Otrok [2008] for further discussion on the normalizations.
We use Bayesian methods to estimate the parameters and make inferences about the latent components. Specifically, we formulate a prior on \((\theta, \gamma)\) and we simulate a Markov chain \(\{\theta(s), \gamma(s), \lambda(s), \xi(s)\}_{s=1}^{S}\) that possesses as invariant distribution the joint posterior given the data \(\{w_t\}\). We can then use the draws \(\{\xi(s)_{st}\}_{s=1}^{S}\) to form point and set estimates of the aggregate trend \(\bar{\tau}_t\), the sector-level trends \(\{\bar{\tau}_{it}\}_{ni=1}^{ni}\), and decomposition of trend changes over time in terms of common and sector-specific contributions. The priors and the Markov Chain Monte Carlo (MCMC) algorithm used for estimation and filtering are described in detail in Appendix B.

**Validation** We conduct a real time forecast comparison against a random walk benchmark to validate our framework. As noted above, two features distinguish our approach: (i) the explicit treatment of temporal aggregation in the persistent component and (ii) the use of the cross-sectional dimension. The forecasting exercise we present here is designed to assess both.

We construct a total of 300 datasets, each of them extending from January 1997 to \(T\) where \(T\) ranges from January 1998 to December 2022. In each dataset (indexed by \(T\)) we run our model and recover point estimates (the posterior median) of the path of the trend in wage growth at the aggregate \(\bar{\tau}_{Tt}\) and at the sectoral level \(\{\bar{\tau}_{it}\}_{ni=1}^{ni}\). We then define the following forecasts based on our model

\[
\begin{align*}
    f_T(h) &= \frac{\sum_{\ell=0}^{12-h} \bar{\tau}_{T,t-\ell} + h\bar{\tau}_{Tt}}{12}, \\
    f_{iT}(h) &= \frac{\sum_{\ell=0}^{12-h} \bar{\tau}_{iT,t-\ell} + h\bar{\tau}_{it}}{12},
\end{align*}
\]

if the forecasting horizon \(h < 12\) and \(f_T(h) = \bar{\tau}_{Tt}\), \(f_{iT}(h) = \bar{\tau}_{it}\) if the forecasting horizon \(h \geq 12\). Thus, \(f_T(h)\) and \(f_{iT}(h)\) are point forecasts of 12-month wage growth at time \(T+h\) given information available at \(T\).

We compare \(f_T(h)\) and \(f_{iT}(h)\) with the random-walk forecasts \(f_{RW_T}(h) = w_T\) and \(f_{RW_{iT}}(h) = w_{iT}\), where the 12-month rate of wage inflation \(w\) is defined as in Equation
We choose a random-walk benchmark because these forecasts are known to be hard to improve upon in the price inflation literature (see, e.g., Atkeson and Ohanian [2001], Stock and Watson [2008]). Since our model assumes random-walk behavior for the trend in monthly nominal wage growth, this exercise can be interpreted as a validation of our temporal aggregation approach. Table 1 shows that $f_T(h)$ consistently outperforms $f_{RW}^T(h)$ for a selection of short and long horizons $h$, as evidenced by the root mean square forecast error (RMSFE). The improvement is particularly clear for aggregate wage growth at the 3-month horizon, decreasing with $h$ thereafter.

At the sector level, $f_T^i$ is a substantially more accurate forecast than $f_{RW}^T$. The improvements occur for all sectors and horizons, but they are largest for the sectors that generally experience more variability in wage growth. Table 1 illustrates this for construction & mining and for leisure & hospitality, two sectors that received a lot of attention during the Great Financial Crisis and the COVID-19 Pandemic, and coincidentally the two sectors with the most volatile wage growth.

Overall, the comparison provides strong evidence in favor of our model and, in particular, of the way we relate monthly to 12-month wage growth. This matters for the joint dynamics of wage growth with other labor market indicators. For example, the contemporaneous correlation between monthly changes in the unemployment rate and monthly changes in CoWI is $-0.18$ (p-value< 0.01) compared to $-0.09$ for the raw data (p-value = 0.12). This finding confirms how CoWI, by purging transitory variation, can provide reliable signals for turning points in aggregate wage growth dynamics. In the next Section, we describe the evolution of CoWI over the past 25 years.

---

13 We omit March 2020 and April 2020 from this computation because of the spike in unemployment at the start of the pandemic.
### TABLE 1. Forecast comparison

<table>
<thead>
<tr>
<th></th>
<th>$h = 3$</th>
<th>$h = 6$</th>
<th>$h = 12$</th>
<th>$h = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f^T(h)$</td>
<td>0.420</td>
<td>0.548</td>
<td>0.839</td>
<td>1.138</td>
</tr>
<tr>
<td>$f_{RW}^T(h)$</td>
<td>0.500</td>
<td>0.619</td>
<td>0.880</td>
<td>1.153</td>
</tr>
<tr>
<td>Difference</td>
<td>−0.080</td>
<td>−0.070</td>
<td>−0.040</td>
<td>−0.014</td>
</tr>
<tr>
<td>p-value</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.07)</td>
</tr>
</tbody>
</table>

**Construction and Mining**

<table>
<thead>
<tr>
<th></th>
<th>$h = 3$</th>
<th>$h = 6$</th>
<th>$h = 12$</th>
<th>$h = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f^T(h)$</td>
<td>1.573</td>
<td>1.670</td>
<td>1.765</td>
<td>2.063</td>
</tr>
<tr>
<td>$f_{RW}^T(h)$</td>
<td>2.077</td>
<td>2.112</td>
<td>2.127</td>
<td>2.389</td>
</tr>
<tr>
<td>Difference</td>
<td>−0.504</td>
<td>−0.443</td>
<td>−0.363</td>
<td>−0.326</td>
</tr>
<tr>
<td>p-value</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

**Leisure and Hospitality**

<table>
<thead>
<tr>
<th></th>
<th>$h = 3$</th>
<th>$h = 6$</th>
<th>$h = 12$</th>
<th>$h = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f^T(h)$</td>
<td>1.420</td>
<td>1.510</td>
<td>1.739</td>
<td>2.026</td>
</tr>
<tr>
<td>$f_{RW}^T(h)$</td>
<td>1.845</td>
<td>1.835</td>
<td>2.052</td>
<td>2.324</td>
</tr>
<tr>
<td>Difference</td>
<td>−0.425</td>
<td>−0.324</td>
<td>−0.313</td>
<td>−0.298</td>
</tr>
<tr>
<td>p-value</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

NOTES. We report p-values for a one-sided test of the null hypothesis that the RMSE of our model forecast is not lower than the RMSE for the random-walk forecast. We computed them by block bootstrap with blocks of size 8.

### 4 Empirical analysis

In this section, we discuss two main sets of empirical results. First, we follow the price inflation literature and aggregate the micro data at the industry level. We describe the evolution of CoWI over time and the relative importance of its common and industry-specific component. Second, we analyze the evolution of CoWI around key macroeconomic events in our sample period: the 2001 Recession, the Great Recession, and the post-pandemic spike in inflation. We re-estimate the model using a series of alternative cross-sections other than industry, such as gender, education, and race. We document that most of the post-pandemic surge in nominal wage growth is common across workers, regardless of the cross-section considered, whereas we find...
FIGURE 1. Wage Growth and Its Persistent Component

(a) Whole sample
(b) Recent sample (2019-2023)

NOTES. Blue shaded areas denote 68 percent probability bands. Grey shaded areas in the left panel denote recessions.

relatively more, albeit limited, support for group-specific trends during the 2001 and 2008 recessions.

4.1 CoWI over the last 25 years: industry decomposition

We estimate the model breaking down wage inflation into seven broad industry groups. $w_{it}$ is defined as in Equation (1) and thus is median nominal wage growth in industry $i$. Figure 1 shows our estimated CoWI (solid blue line) together with the realized twelve-month wage growth from the CPS data (black line). A 68 percent posterior probability band is given by the blue shaded areas around CoWI.

We highlight two main takeaways. First, the underlying data exhibit a lot of monthly variation that is purged by our approach. Most of the high-frequency variation in nominal wage growth is ascribed to measurement error in our model (the $\tilde{\epsilon}_{it}$ terms), as discussed in Section 3. Figure 1a makes clear that substantial movements in the persistent component of wage inflation coincide with significant macroeconomic economic
events over the period, such as the Great Recession or the inflation surge episode starting in 2021. As shown in Section 3, our measure also strongly correlates with the unemployment rate, which confirms its informative power for macroeconomic behavior.

Second, we can analyze the wage inflation surge episode starting in 2021 through the lens of our model. Figure 1b shows that CoWI oscillates between 3.2 percent and 3.7 percent between 2016 and 2020, and then shoots up in early 2021, nearly doubling over the course of the year. As such, the model assigns almost all of the nominal wage growth observed in the data to the persistent component. The estimated CoWI peaks at the beginning of 2022 and slowly declines thereafter, although it seems to have plateaued in the last months.

We use our framework to further investigate whether fluctuations in CoWI are the result of sector-specific shocks. Recall from Equation (3) that CoWI can be written as the sum of a common component \(\sum_{i=1}^{n} s_i \alpha_{s_i,t} \tau_{s_i,t}\) and a sector-specific component \(\sum_{i=1}^{n} s_i \tau_{s_i,t}\). Figure 2 depicts the evolution of each of these components over the sample. We find that most of the variation in CoWI is accounted for by the common persistent component of nominal wage growth, both during the two NBER recessions and the inflation surge episode in the sample. By contrast, the estimated sector-specific persistent component of the model captures lower frequency movements. We find that it steadily falls over the period, which implies an increase in the cross-sector correlation of nominal wage growth. This may be the result of slow-moving compositional changes, with relatively high-wage growth sectors becoming smaller over the sample period (e.g., manufacturing). The sector-specific component of CoWI appears to in-

---

14As discussed in Section 3, the trend extracted by the model is expressed in terms of annualized monthly wage growth, which explains why it leads the actual year-over-year wage growth series in the chart.

15This finding is consistent with the analysis of sectoral mismatch during the Great Recession in Şahin, Song, Topa, and Violante [2014]. They find limited evidence of sectoral mismatch between job seekers and job openings. While wages are not specified in their framework, this result is in line with sectoral wage inflation being driven by a common trend.
FIGURE 2. Persistence of Wage Growth: Common or Sector-Specific?

NOTES. Shaded areas denote 68 percent probability bands. The dashed black line is nominal wage growth, the red line is the common component $P_{n,t} = \alpha_{t} \tau_{t}$, while the blue line is the sector-specific component $P_{n,t} = \tau_{t}$. Each of these three series are plotted as a cumulative change from their average over the period 2017–2019, which is therefore centered at 0.

crease following the pandemic (although the probability bands are large) but remains below its level from the early 2000s.

To sum up, our framework assigns most of the business cycle variation in nominal wage inflation to CoWI. In turn, CoWI is mostly driven by the component that is common across industries. Variations in the component of CoWI specific to each industry are second-order. This finding holds even for industries that can be expected to be specifically impacted at some point in our sample, such as construction during the Great Recession and leisure and hospitality following the COVID-19 pandemic (see Appendix C). Next, we show how our conclusions are broadly unaffected when considering different cross-sections of workers.
4.2 Heterogeneity and the aggregate dynamics of CoWI

In this section, we investigate the robustness of our findings by considering a series of alternative cross-sections that represent likely sources of heterogeneity in nominal wage growth. In particular, we investigate whether the aggregate dynamics of CoWI is driven by specific groups of workers or employers by considering several alternative cross-sections to industries.

Our measure of wage inflation comes from the aggregation of changes in nominal wages at the worker level, which do not occur frequently. Many explanations for these nominal wage rigidities have been put forward in the literature. Examples include employers insuring workers against firm-level shocks, fairness considerations related to efficiency, and institutionalized wage setting, such as employers setting wages at the national level or minimum wage institutions.\textsuperscript{16} All of these explanations are likely to result in different degrees of nominal wage stickiness across distinct groups of workers. As just one example, national wage setting by employers or minimum wage regulations are more likely to apply to low-skill workers.\textsuperscript{17}

To gauge whether this micro-level heterogeneity translates into group-specific trends in wage inflation, we re-estimate our model for a series of cross-sections that we derive from the CPS. Besides the industry cross-section discussed in Section 4.1, we consider cross-sections based on occupation, education level, region, wage quartile, age, gender, and race.\textsuperscript{18} For each cross-section, we obtain estimates for CoWI and its common and idiosyncratic components given in Equation (3).

We summarize our estimation results in Figure 3, by focusing on the three largest


\textsuperscript{17}See Avouyi-Dovi, Fougère, and Gautier [2013] and Minton and Wheaton [2022] for recent work on minimum wages and downward wage rigidity. Hazell, Patterson, Sarsons, and Taska [2022] show how national wage setting increases wage rigidity.

\textsuperscript{18}We follow the definitions used in the Atlanta Fed Wage Growth Tracker. See Appendix A.1 for details.
FIGURE 3. CoWI across models

NOTES. The three episodes refer to the following periods: 2001m3-2001m11, 2007m12-2009m6, and 2021m2-2022m1. Each model uses the data cut described in the legend; the variables follow the Atlanta Fed wage tracker definition and are detailed in Appendix A.1. Black square markers indicate the peak-to-trough change in CoWI for each model and episode. The diamond markers indicate the peak-to-trough change for the common component \( \left( \sum_{i=1}^{n} s_i \alpha_{i,t} \right) \tau_{t,t} \) in each model and episode. Vertical lines show the 68 percent probability bands.

changes in CoWI in the sample: the 2001 Recession (2001m3-2001m11), the Great Recession (2007m12-2009m6), and the post-pandemic wage inflation surge episode (2021m2-2022m1).\(^{19}\) For each episode and each model estimated on a separate cross-section, we plot the cumulative change in CoWI, denoted by a black square marker. It is clear that the estimated cumulative CoWI change is very robust across specifications. For instance, CoWI cumulative change during the 2021 wage inflation surge episode is estimated between 2.9 and 3.3 percent across models.

Figure 3 also reports the cumulative change in the common component of CoWI for

\(^{19}\)The recession episodes follow NBER dating – with start date one month before the onset of the recession. The results are robust to considering slightly different start and end dates.
each episode, which are denoted by diamond markers. The vertical bars delineate 68 percent posterior probability bands. The figure suggests that, across models, most of the cumulative variation in CoWI in these three episodes can be ascribed to its common component, as can be seen by how close the diamond markers are to the black square markers.

To formally quantify these contributions, we report in Table 2 the fraction of the cumulative change in CoWI accounted for by its common component over the duration of each episode. Overall, these results invite us to qualify the results obtained based on the industry group cross-section. We find that the common component remains an important determinant of CoWI across the various cross-sections, but to a lesser extent during the 2001 recession (42% to 91%) and the Great Recession (53% to 91%). Industries are measured to have a larger contribution of the common component of wage inflation than any other cut of the data across the three episodes considered (greater than 90% overall). These results suggest that industries may not be the most relevant level of disaggregation to study the implications of heterogeneity in wage stickiness at the macro level.

Notably, we find that the common component of wage inflation accounts for most of the increase in CoWI across the various cross-sections during the 2021 inflation surge episode. The contribution of the common component is greater than 77% across subgroups during that episode, typically higher than during the two recessions in the sample.

This finding suggests the existence of a form of asymmetry between recessionary

---

20 To obtain the probability bands, we apply the algorithm described in Appendix B to each cross-sectional cut. This produces draws from the posterior distribution of cumulative changes in the common component, the empirical quantiles of which form the bands we report. In that regard, the probability bands have coverage pointwise in both episode and data cut.

21 We have outlined in Section 3 why this object is identified while the levels are not.

22 This result is consistent with Le Bihan, Montornès, and Heckel [2012] and Barattieri, Basu, and Gottschalk [2014], who find little industry (and occupation) heterogeneity in the frequency of nominal wage adjustments.
### TABLE 2. Contribution of common trend to CoWI during selected episodes

<table>
<thead>
<tr>
<th></th>
<th>2001 Recession</th>
<th>Great Recession</th>
<th>2021 Inflation Surge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry</td>
<td>0.91</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>(0.69, 1.15)</td>
<td>(0.77, 1.03)</td>
<td>(0.80, 1.04)</td>
</tr>
<tr>
<td>Occupation</td>
<td>0.52</td>
<td>0.69</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>(0.17, 0.85)</td>
<td>(0.48, 0.88)</td>
<td>(0.69, 1.06)</td>
</tr>
<tr>
<td>Education</td>
<td>0.65</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(0.33, 0.95)</td>
<td>(0.56, 0.94)</td>
<td>(0.68, 1.03)</td>
</tr>
<tr>
<td>Region</td>
<td>0.88</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>(0.63, 1.11)</td>
<td>(0.79, 1.02)</td>
<td>(0.82, 1.04)</td>
</tr>
<tr>
<td>Wage quartile</td>
<td>0.68</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>(0.23, 1.93)</td>
<td>(0.60, 0.95)</td>
<td>(0.68, 1.04)</td>
</tr>
<tr>
<td>Age</td>
<td>0.47</td>
<td>0.69</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>(0.04, 0.92)</td>
<td>(0.43, 0.91)</td>
<td>(0.60, 1.07)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.42</td>
<td>0.53</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>(0.03, 0.85)</td>
<td>(0.20, 0.81)</td>
<td>(0.47, 1.01)</td>
</tr>
<tr>
<td>Race</td>
<td>0.49</td>
<td>0.60</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>(0.08, 0.92)</td>
<td>(0.34, 0.85)</td>
<td>(0.55, 1.04)</td>
</tr>
</tbody>
</table>

**NOTES.** Each row refers to a different model, using the cross-sectional variable reported, defined following the Atlanta Fed wage tracker definition and detailed in Appendix A.1. The three episodes in the columns refer to the following periods: 2001m3-2001m11, 2007m12-2009m6, and 2021m2-2022m1. For each cross-section and episode, we compute the change in the common component of CoWI between the start and end of the period, as well as the change in the overall CoWI over the same horizon, and report the median of this ratio. We report a 68 percent posterior probability interval in parentheses. 

and inflationary episodes, with wages potentially more rigid downward than upward. To investigate whether this conclusion is specific to the pandemic, we have also estimated our model using average hourly earnings data from the Current Employment Statistics (CES), which, despite the compositional issues documented in Appendix A.2, has the advantage to start in 1964. Our results, reported in Appendix E, show that the common persistent component is also the main driver of wage inflation during the inflation episodes of the 1970s. For instance, 78% of the 4.5 percentage point increase in wage inflation between 1965 and 1981 is estimated to be common across industries,
while industry-specific variation played a quantitatively bigger role during the Great Recession.

In conclusion, no specific subgroup, in the eight cross-sections that we consider, seems to play a disproportionate role in driving the overall trend in nominal wage growth, especially during inflationary periods. A useful case study is the model estimated with wage quartiles. As recently studied by Autor, Dube, and McGrew [2023], the COVID-19 period has seen a substantial wage compression, with nominal wage growth at the bottom of the distribution being relatively stronger than in the upper echelons. While our estimated model suggests that nominal wage growth in the bottom wage quartile is initially driven by group-specific forces, this has a relatively minor role for the evolution of wage growth in the aggregate.

We wish to stress that our findings relate to the importance of worker and job heterogeneity in accounting for wage inflation at the macro level. Our results do not necessarily imply that there is no heterogeneity in wage stickiness at the micro level, including during inflationary episodes. Instead, our framework suggests that this heterogeneity, if present, is not materially important for characterizing wage inflation episodes at the macro level.

5 Conclusion

We estimate the latent persistent component of aggregate nominal wage growth by combining worker-level data with time series filtering and smoothing methods. This measure, which we label CoWI, is purged from noise and short-run fluctuations. We propose a novel approach to account for the monthly-to-yearly temporal aggregation present in CPS data. In a real-time forecast comparison, we show that it outperforms a

23 Blanchet, Saez, and Zucman [2022] construct high-frequency distributional indicators and highlight how real wages grew more rapidly at the bottom of the distribution in 2021.
random walk benchmark and, as such, that it can provide reliable signals for turning points in wage inflation. CoWI moves substantially during the three episodes of largest macroeconomic relevance in our sample: the 2001 and 2008 recessions, and the wage inflation episode of 2021. In all episodes, we find that changes in CoWI are driven by its common component. In accounting for these changes, no specific job or worker characteristics matters to a first order. Finally, our results suggest that the common component of CoWI seems especially important during inflationary episodes. It contributes more than 75% of the increase in aggregate nominal wage growth during the post-pandemic episode of inflation.
References


Supplemental Appendix to

“A Measure of Core Wage Inflation”

by Martín Almuzara†, Richard Audoly‡, and Davide Melcangi§

A Data Appendix

A.1 Variable definition

The cross-sectional variables we use in the estimation follow the definition by the Atlanta Fed wage growth tracker.

**Industries (7 groups)** Construction and Mining, Education and Health, Finance and Business Services, Leisure and Hospitality, Manufacturing, Public Administration, and Trade and Transportation.

**Occupations (3 groups)** high-skill (Managers, Professionals, Technicians), middle-skill (Office and Administration, Operators, Production, Sales), and low-skill (Food Preparation and Serving, Cleaning, individual Care Services, Protective Services).

**Race (2 groups)** White and Nonwhite.

**Education (3 groups)** High school or less, Associates degree, and Bachelor degree or higher.

---

†Federal Reserve Bank of New York. Email: martin.almuzara@ny.frb.org. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

‡Federal Reserve Bank of New York. Email: richard.audoly@ny.frb.org.

§Federal Reserve Bank of New York. Email: davide.melcangi@ny.frb.org.
**Age (3 groups)** 16–24 years old, 25–54, and 55+.

**Gender (2 groups)** Male and Female.

**Wage quartiles (4 groups)** The quartiles are based on the average between workers’ current hourly wage and their wage 12 months prior (when their wages are last recorded).

**Region (9 groups)** The nine Census Divisions: New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont), Mid-Atlantic (New Jersey, New York, Pennsylvania), East North Central (Illinois, Indiana, Michigan, Ohio, Wisconsin), West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota), South Atlantic (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, West Virginia), East South Central (Alabama, Kentucky, Mississippi, Tennessee), West South Central (Arkansas, Louisiana, Oklahoma, Texas), Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming), Pacific (Alaska, California, Hawaii, Oregon, Washington).

**A.2 Alternative measures of nominal wage growth**

Figure 1 shows four alternative measures of nominal wage growth. The first two measures, the Atlanta Fed Wage Growth Tracker (WGT) and Employment Compensation Index (ECI), adjust for composition changes of the workforce by measuring nominal wage growth respectively within worker and within job. The other two measures, Average Hourly Earnings (AHE) and Compensation per Hour (CpH), are aggregate measures subject to any composition changes of the workforce.

As shown in figure 1, these composition changes limit what can be learned about the underlying growth in the price of labor. Average Hourly Earnings and Compensation
per Hour are strikingly more volatile. This volatility is mostly at odds with the nominal wage growth definitions adjusting for composition changes. For instance, both Average Hourly Earnings and Compensation per Hour exhibit large swings during the Covid pandemic. The Atlanta Fed Wage Growth Tracker and Employment Cost Index, by contrast, are flat over the same period.

FIGURE 1. Alternative wage inflation measures
\section*{B \hspace{0.2cm} Details of model and estimation approach}

Recall our notation for the data $w_t = \{w_{it}\}_{i=1}^n$, the time-invariant parameters

\[ \theta = \left\{ \{\theta_{c\ell}\}_{1 \leq \ell \leq p}, \{\theta_{i\ell}\}_{1 \leq \ell \leq q, 1 \leq i \leq n} \right\}, \]

\[ \gamma = \left\{ \{\gamma_{a,m,j}\}_{m=\tau,\epsilon, 1 \leq i \leq n}, \{\gamma_{o,m,j}\}_{m=\Delta \tau, \epsilon, j=c,1,\ldots,n} \right\}, \]

the time-varying parameters

\[ \lambda_t = \left\{ \{\alpha_{t,ijt}, \alpha_{e,ijt}\}_{i=1}^n, \sigma_{\Delta \tau,clt}, \{\sigma_{\Delta \tau,ji} \}_{i=1}^n, \sigma_{\epsilon,clt}, \{\sigma_{\epsilon,ji} \}_{i=1}^n \right\}, \]

and the latent components

\[ \xi_t = \left( \{\tau_{clt}, \{\tau_{ijt}\}_{i=1}^n, \epsilon_{clt}, \{\epsilon_{ijt}\}_{i=1}^n \right). \]

To conduct Bayesian inference, we begin by formulating a prior on $(\theta, \gamma)$.

\paragraph*{Choice of priors} The MA coefficients $\theta$ are prior independent of each other with $\theta_{jt} \sim N(0, \nu_{jt}^2)$ for $j = c, 1, \ldots, n$. That is, we shrink the model towards one with white noise transitory errors and the strength of the shrinkage is determined by the choice of $\nu_{jt}$. In our baseline model we set $p = 0$ and $q = 3$, and we put higher penalties on the more distant lags as in the Minnesota prior of Doan, Litterman, and Sims [1983]. We achieve that by setting $\nu_{jt} = 1/(10t^2)$.

The standard deviations $\gamma$ control the amount of time-variation in loadings and volatilities. Unless they are small, the model may be excessively flexible causing overfitting. Our approach is to put a reasonably tight prior centered around small values to shrink the model towards no time-variation in the parameters. Specifically we use independent inverse gamma priors of the form $\gamma_{k,m,j}^2 \sim \Gamma^{-1}(d_k/2, 2/(d_k\sigma_k^2))$ for
\( k = \alpha, \sigma \). The location parameters are set to \( s^2_\alpha = 0.0001 \) and \( s^2_\sigma = 0.001 \), and the degree-of-freedom hyperparameters are set to \( d_\alpha = d_\sigma = 60 \).

**Estimation and filtering** Inference about parameters and latent variables is implemented via Gibbs sampling. This is a type of Markov Chain Monte Carlo (MCMC) algorithm suitable to approximate the joint posterior distribution of parameters and latent variables by simulation in state-space models.

The Gibbs sampler constructs a Markov Chain \( \{\theta^{(s)}(t), \gamma^{(s)}(t), \lambda_t^{(s)}, \xi_t^{(s)}\}_{s=1}^S \) having as invariant distribution the posterior

\[
P(\theta, \gamma, \lambda_t, \xi_t | \{w_t\}).
\]

This allows us to estimate the posterior of our objects of interest, e.g.

\[
P(\{\bar{\tau}_t, \{\bar{\tau}^{(s)}_i\}_{i=1}^n\} | \{w_t\}),
\]

using the draws \( \{\theta^{(s)}, \gamma^{(s)}, \lambda_t^{(s)}, \xi_t^{(s)}\}_{s=1}^S \) to form \( \{\bar{\tau}_t, \{\bar{\tau}^{(s)}_i\}_{i=1}^n\}_{s=1}^S \) and taking the simulation frequencies of the objects as estimates of posterior probabilities. If the Markov chain converges (in a suitable sense) and \( S \) is large, the approximation error will be small.

One advantage of the Bayesian approach is that posterior calculations already integrate both the sampling uncertainty from parameter estimation and the signal-extraction uncertainty about the latent components. When reporting the path over time of a latent time series in our empirical analysis, we use credible intervals with fixed credibility level pointwise in \( t \).\(^1\)

An alternative would be to estimate \((\theta, \gamma)\) by maximum likelihood. It is straight-

\(^1\)It is conceptually straightforward and computationally feasible to compute pathwise credible regions along the lines of, e.g., Inoue and Kilian [2016].
forward to modify our MCMC algorithm to approximate the maximum likelihood estimator by stochastic EM — simply replace the posterior updates of \( \theta \) and \( \gamma \) by the solutions to the corresponding complete-data score equations. However, inferences about latent variables (and, in particular, about our objects of interest) that arise from that procedure would not necessarily account for the estimation uncertainty in \((\theta, \gamma)\).


Relative to Del Negro and Otrok [2008], Stock and Watson [2016] incorporate outliers in the transitory shocks. Compared to Stock and Watson [2016], we allow for temporal aggregation in the persistent components and for MA dynamics in the transitory components. For simplicity, we discuss estimation of a model without outliers.\(^2\) We find only a negligible role for them in the data we analyze.

The Gibbs sampler exploits the fact that with a careful grouping of parameters and latent variables, the conditional distributions of each block given the rest can be simulated by well-known algorithms. In our model, there are three big blocks with many sub-blocks, namely:

(A) \( P(|\xi_t|, \lambda_t, \theta, \gamma, \{w_t\}) \). Conditional on time-varying parameters \( \lambda_t \) and the MA coefficients \( \theta \), the data \( w_t \) and the latent variables \( \xi_t \) are related by a linear state-space model with time-varying matrices. We apply the simulation smoother algorithm proposed by Durbin and Koopman [2002] to efficiently sample \( \{\xi_t\} \).

To accommodate the MA dynamics of the common and sector-specific transitory errors, we include \( \varepsilon_{c,t}, \varepsilon_{c,t-1}, \ldots, \varepsilon_{c,t-p+1}, \varepsilon_{i,t}, \varepsilon_{i,t-1}, \ldots, \varepsilon_{i,t-q+1} | i = 1 \) as additional state variables.

---

\(^2\)The outliers in Stock and Watson [2016] are introduced by assuming \( \eta_{c,j} = s_{\beta} \times \tilde{\eta}_{c,j} \) for \( j = c, 1, \ldots, n \) where \( s_{\beta} = 1 \) with probability \( p_j \) and \( s_{\beta} \sim U(1, 10) \) with probability \( 1 - p_j \) while \( \tilde{\eta}_{c,j} \sim N(0, 1) \).
(B) $P(\lambda_t|\xi_t, \theta, \gamma, \{w_t\})$. This can be further partitioned into the following blocks:

(i) $P\left(\left\{\alpha_{z,t,i}, \alpha_{e,t,i}\right\}_{i=1}^{n}\left|\left\{w_{it}, \tau_{it}\right\}_{i=1}^{n}, \{\tau_{c,i}, \epsilon_{c}\}, \left\{\sigma_{e,t,i}\right\}_{i=1}^{n}, \{y_{a,\tau,\nu}, Y_{a,\tau,\nu}\}_{i=1}^{n}\right\}\right)$. It is the result of a multivariate regression with time-varying coefficients and MA error terms. It can be dealt with using linear state-space techniques. Thus, we apply the simulation smoothing algorithm of Durbin and Koopman [2002] to the corresponding state-space representation in order to sample $\{\left[\alpha_{t,i}, \alpha_{e,t,i}\right]\}_{i=1}^{N}$.

(ii) $P\left(\left\{\alpha_{m,i,j}\right\}|\left\{m_{jt}\right\}, \gamma_{0,\nu}\right)$ for $m = \Delta\tau, \epsilon$ and $j = c, 1, \ldots, n$. Given $\gamma_{0,\nu}$ and $m_{jt}$, $\sigma_{m,i,j}$ follows a stochastic-volatility model with observation equation $\ln m_{jt}^2 = \ln \sigma_{m,i,j}^2 + \ln \eta_{m,i,j}^2$ and transition equation $\Delta \ln \sigma_{m,i,j}^2 = \gamma_{0,\nu}v_{0,\nu}$. We then use the algorithm proposed in Kim, Shephard, and Chib [1998] and Omori, Chib, Shephard, and Nakajima [2007] that consists of approximating the log-$\chi^2$ distribution of $\ln \eta_{0,\nu}^2$ with a 10-component normal mixture and applying linear state-space techniques to that approximation.

(C) $P(\theta, \gamma|\xi_t, \{\lambda_t\}, \{w_t\})$. This can also be partitioned into subblocks:

(i) $P\left(\gamma_{0,\nu}|\Delta\alpha_{m,\nu}\right)$ for $m = \tau, \epsilon$. We draw the reciprocal of the square root of a gamma random variable with $d_{\alpha} + T$ degrees of freedom and mean $(d_{\alpha}s_{\alpha}^2 + \sum_{t=1}^{T} \Delta \alpha_{m,\nu})/(d_{\alpha} + T)$ for $m = \tau, \epsilon$.

(ii) $P\left(\gamma_{0,\nu}|\Delta \ln \sigma_{m,\nu}^2\right)$ for $m = \Delta\tau, \epsilon$ and $j = c, 1, \ldots, n$. We draw the reciprocal of the square root of a gamma random variable with $d_{\omega} + T$ degrees of freedom and mean $(d_{\omega}s_{\omega}^2 + \sum_{t=1}^{T} \Delta \ln \sigma_{m,\nu}^2)/(d_{\omega} + T)$ for $m = \tau, \epsilon$ and $j = c, 1, \ldots, n$.

(iii) $P\left(\theta|\nu_{jt}, \sigma_{\nu,jt}\right)$ for $j = c, 1, \ldots, n$ where $\theta_c = (\theta_{c1}, \ldots, \theta_{cp})'$ and $\theta_i = (\theta_{i1}, \ldots, \theta_{iq})'$ for $i = 1, \ldots, n$. This problem can be treated separately for each $j$. We do the derivation for $j = i = 1, \ldots, n$ (the case $j = c$ is identical except that $p$ should take the place of $q$). Define $\nu_{jt} = \sigma_{\nu,jt}^{-1}$. Conditioning on $q$ initial observations
\( \epsilon_{i0}, \ldots, \epsilon_{iq-1} \) we obtain the likelihood term

\[
P\left( \{\epsilon_{it}^{T}\}_{t=1}^{T} | \{\theta_{it}\}_{t=1}^{q}, \{\epsilon_{i1-\ell}^{q}\}_{t=1}^{q}, \{\sigma_{\epsilon, it}\} \right) = \prod_{t=1}^{T} P\left( \epsilon_{it} | \theta_{it}, \{\epsilon_{i1-\ell}^{q}\}_{t=1}^{q}, \{\sigma_{\epsilon, it}\} \right) \\
= \prod_{t=1}^{T} P\left( \epsilon_{it} | \theta_{it}, \{\epsilon_{i1-\ell}^{q}\}_{t=1}^{q}, \sigma_{\epsilon, it} \right) \\
= \prod_{t=1}^{T} \frac{1}{\sigma_{\epsilon, it}} \phi \left( \frac{\epsilon_{it} - \left( \sum_{\ell=1}^{q} \theta_{it} \epsilon_{i1-\ell} \right)}{\sigma_{\epsilon, it}} \right)
\]

where \( \phi \) is the standard normal density. This is the likelihood from a regression of \( \epsilon_{it} \) on \( (\epsilon_{i1-1}, \ldots, \epsilon_{i1-q})' \) with heteroskedastic Gaussian errors or, equivalently, (up to a scaling constant) from a regression of \( y_{it} = \epsilon_{it}/\sigma_{\epsilon, it} \) on \( x_{it} = (\epsilon_{i1-1}, \ldots, \epsilon_{i1-q})'/\sigma_{\epsilon, it} \) with i.i.d. \( N(0,1) \) errors. Since the prior is \( (\theta_{i1}, \ldots, \theta_{iq})' \sim N(0_{Q \times 1}, V_{\theta}) \) where the variance is \( V_{\theta} = \text{diag}(v_{1}, \ldots, v_{q}) \), the posterior follows from the usual regression formula

\[
\{\theta_{it}\}_{t=1}^{q} \mid \{\epsilon_{it}\}, \{\sigma_{\epsilon, it}\} \sim N \left( \frac{V_{\theta}^{-1} + \sum_{t=1}^{T} x_{it}' x_{it}}{V_{\theta}^{-1} + \sum_{t=1}^{T} x_{it}' x_{it}} \right)^{-1} \sum_{t=1}^{T} x_{it} y_{it}, \left[ V_{\theta}^{-1} + \sum_{t=1}^{T} x_{it}' x_{it} \right]^{-1}.
\]

Conditioning on the initial observations \( \{\epsilon_{i1}, \ldots, \epsilon_{i1-q}\} \) has at most a small effect when \( T \) is large. As an alternative, we can include \( \{\epsilon_{i1}\} \) as state variables in step (A).

**Implementation, numerical accuracy and tests** We do \( S = 12,000 \) draws retaining one every two after burning the first 6,000. The result is a chain for the parameters \( (\theta, \gamma) \) with low enough autocorrelations that the posterior expectations have negligible Monte Carlo standard errors. We also monitor the behavior of the latent variables and stochastic volatilities, the paths of which seem to stabilize within a small region well before the burn-in period ends. We ran the posterior simulator test suggested by
Geweke [2004] and an extensive Monte Carlo simulation study, finding no indication against our implementation.

C Additional empirical results

FIGURE 2. Aggregate and group-specific trend by industry

(a) Construction and Mining  (b) Education and Health

(c) Finance and Business Services  (d) Leisure and Hospitality
NOTES. The figure shows for each industry the raw nominal wage growth data, the common trend component \((\alpha_t, \tau_t)\) and the sector specific trend \((\tau_{it})\) over the sample period.
D Robustness checks

In this appendix, we verify the robustness of the main results of the paper to three choices we make in the empirical analysis. The first is to use the median instead of the mean of year-over-year wage growth as the observable \( w_{it} \) in our model. The second choice is to use the unweighted median as opposed to the median weighted by the survey weights as \( \tilde{w}_{it} \). Third, we do not allow for \( \tilde{\tau}_{it} \) to be itself the sum of a persistent and a transitory component. Figures 3, 4 and 5 show that both the historical behavior of the persistent component in wage growth and the relatively high importance of common variation across industries are insensitive to these choices.

FIGURE 3. Estimates based on mean year-over-year wage growth

Despite mean year-over-year wage growth being more volatile than median wage growth, our model traces a remarkably similar historical evolution of the persistent component (CoWI), with the largest swings located around the same episodes we discussed in Section 4 (i.e., the 2001 and 2008 recessions, and the post-pandemic inflation spike). CoWI is somewhat higher when using the mean instead of the median due to the positive skewness in the wage growth distribution, but this seems to imply merely a level shift in the persistent component. The cumulative changes in panel (b) of figure 3, for example, are quantitatively very close to our baseline results.

Differences in our estimates when using the weighted instead of the unweighted
median of wage growth as $w_{it}$ are imperceptible, as shown in figure 4.

FIGURE 4. Estimates based on weighted median wage growth

![Graph showing wage growth and inflation](image)

(a) Persistent component of wage growth  
(b) Common and sector-specific contribution

Figure 5 illustrates a point made in section 3. Our empirical analysis interprets the transitory component of year-over-year wage growth $\bar{\epsilon}_{it}$ as being largely measurement error. Therefore, $\bar{\tau}_{it}$ is interpreted as the unobservable monthly growth rate of nominal wages that could be recovered with a perfect error-free survey. Because we rely on time series smoothing techniques, the assumption that $\bar{\tau}_{it}$ is well approximated by a random walk is important in order to filter the survey measurement error out. If instead $\bar{\tau}_{it}$ is the sum of two components,

$$\bar{\tau}_{it} = \bar{\tau}_{it}^{\text{pers}} + \bar{\tau}_{it}^{\text{tr}}$$

where $\bar{\tau}_{it}^{\text{pers}}$ is now a random walk and $\bar{\tau}_{it}^{\text{tr}}$ is white noise, our baseline model with the choice of moving average orders $p = q = 12$ would estimate $\bar{\tau}_{it}^{\text{pers}}$ instead of $\bar{\tau}_{it}$. Comparing estimates from this extended model and the results in section 4 provides a sense of how important the genuine transitory shock $\bar{\tau}_{it}^{\text{tr}}$ is. Figure 5 indicates that $\bar{\tau}_{it}^{\text{tr}}$ plays at most a minor role and that the more parsimonious model used in our paper captures sufficiently well the most salient movements in aggregate wage growth, which tend to be very persistent.

A final piece of evidence supporting our interpretation of $\bar{\epsilon}_{it}$ as measurement error
FIGURE 5. Estimates based on a more flexible specification

(a) Persistent component of wage growth
(b) Common and sector-specific contribution

is shown in figure 6. Consider the transitory component of aggregate wage growth, which we define as

$$\tilde{\varepsilon}_t = \sum_{i=1}^{n} s_{it} \tilde{\varepsilon}_{it} = \sum_{i=1}^{n} s_{it} w_{it} - \bar{\tau}_t$$

where $s_{it}$ is the employment share of cross-section $i$ in month $t$. The variance of $\tilde{\varepsilon}_t$ is given by

$$\tilde{\sigma}^2_{\varepsilon,t} = \left( \sum_{i=1}^{n} s_{it} \alpha_{\varepsilon, it} \right)^2 \sigma^2_{\varepsilon, it} + \sum_{i=1}^{n} s^2_{it} \sigma^2_{\varepsilon, it},$$

If $\tilde{\varepsilon}_{it}$ is the error made in using the sample median of year-over-year wage growth $w_{it}$ from a sample of $n_{it}$ workers to estimate the population growth rate $\sum_{\ell=1}^{12} \tilde{\tau}_{it+1-\ell}/12$ in each sector $i$, then the standard deviation $\tilde{\sigma}_{\varepsilon,t}$ should be proportional to $1/\sqrt{n_t}$ where $n_t = \sum_{i=1}^{n} n_{it}$ is the survey sample size in month $t$. Figure 6 shows precisely that: a scatter plot of (the posterior median estimate of) $\tilde{\sigma}_{\varepsilon,t}$ against $n_t$ in which most of the points lie close to the line $\tilde{\sigma}_{\varepsilon,t} = \hat{\alpha}/\sqrt{n_t}$.\(^3\) We find a similar pattern if we consider the correlation between sample size $n_{it}$ and the standard deviation $\sigma_{\varepsilon, it}$ for a specific industry $i$.

We also find, consistent with our interpretation, that for every $i$ the path of $\alpha_{\varepsilon, it}\tilde{\sigma}^2_{\varepsilon, it}$

\(^3\)In fact, the correlation between $\tilde{\sigma}_{\varepsilon,t}$ and $1/\sqrt{n_t}$ is 0.9.
always contains the zero line, indicating a negligible role for cross-sectional correlation across $\tilde{\epsilon}_{it}$.

E Additional evidence using CES

This appendix presents estimates of the persistent component of month-on-month growth rates in nominal wages using data from the Current Employment Statistics (CES).\footnote{This is formally defined as average hourly earnings.} Because the data already provides month-on-month changes, denoted by $W_{it}$ below, we estimate our model without temporal aggregation. In other words, instead of (2), our measurement equation is

$$W_{it} = \tilde{z}_{it} + \tilde{\epsilon}_{it}$$
with the persistent component $\bar{\tau}_{it}$ and the transitory component $\bar{\varepsilon}_{it}$ modeled as in Section 3. The cross-sectional dimension is industries since the CES is a survey of establishments.\footnote{We consider 10 industries: Construction, Financial Activities, Information, Leisure and Hospitality, Manufacturing, Mining and Logging, Other Services, Private Education and Health Services, Professional and Business Services, Trade-Transportation-Utilities.}

As noted in Section 2, the CES measure of wages is subject to compositional issues. However, the CES spans a longer period (in this case beginning in 1964), which allows us to empirically study additional recessions and the inflationary episodes of the late 1960s and 1970s. Figure 7 contains the trend estimates and its decomposition into common and sector-specific drivers. Figure 7a is the CES equivalent to Figure 1 and Figure 7b is comparable to Figure 2.

Figure 7a shows that the model attributes most of the high-frequency variation in nominal wage growth in the CES to the transitory variation term $\bar{\varepsilon}_{it}$. The two largest changes in the persistent component of wage inflation correspond to the inflation episodes in the 1970s and the post-pandemic inflation surge. From the 1980s, most NBER recessions tend to be associated with a drop in CoWI.

In addition, Figure 7b confirms our findings that the sector-specific persistent component captures very low frequency movements. In contrast, the common latent factor plays a prominent role during large swings in aggregate nominal wage growth, and especially in the inflationary periods. This is visually clear in the 1970s, thus suggesting that our results are not specific to the post-pandemic period. For example, about 80% of the 4.5 percentage point increase in wage inflation between 1965 and 1981 is common across industries. Looking at shorter periods, the common factor explains 78% and 85% of the increase in wage inflation between 1973 and 1975 and between 1980 and 1982, respectively.
FIGURE 7. Estimates using CES data

(a) Persistent component of wage growth

(b) Common and sector-specific contribution
**References: Supplemental Appendix**


