Employment Reconciliation and Nowcasting*

Eiji Goto
University of Missouri-St. Louis
egoto@umsl.edu

Jan P.A.M. Jacobs
University of Groningen, CAMA and CIRANO
j.p.a.m.jacobs@rug.nl

Tara M. Sinclair
U.S. Department of the Treasury and The George Washington University
tsinc@gwu.edu

Simon van Norden
HEC Montréal, CIRANO and CIREQ
simon.van-norden@hec.ca

January 2023

Abstract

We construct a latent employment estimate for the U.S. which both reconciles the information from separate payroll and household surveys, and incorporates the preliminary data revision process of the payroll data. We find that our reconciled latent employment series looks somewhat different than the initial release of payroll employment and is closer to the fully-revised data that is benchmarked to a near census of employment. A real-time exercise, however, suggests that the reconciled employment estimate is remarkably similar to the initial release of payroll employment with near zero weight on the household survey information.

*The authors thank Matthew Zahn (whose undergraduate thesis inspired this project), Elmar Mertens, Todd Clark, James Mitchell and other participants at the FriendlyFaces seminar. We also thank participants at the Real-time Economics Conference 2021, the the Federal Reserve Bank of Philadelphia virtual brown bag seminar, the ESCOE Conference on Economic Measurement 2022, the Society for Economic Measurement 2022 Conference, the 2023 Federal Reserve Bank of St. Louis Applied Time Series Econometrics Workshop, and the 2023 IAAE Annual Conference. Finally, thanks to the editor and three referees for feedback that improved this version of the paper. The views expressed here are the authors' personal views and do not necessarily reflect those of the Department of the Treasury or the U.S. Government.

JEL classification: C22, C53, C82, E24
Keywords: employment, United States, real-time data, news, noise

1 Introduction

Employment is a key economic indicator that is closely watched by governments, businesses, journalists, financial analysts, and many others. In the U.S. in particular, the releases by the Bureau of Labor Statistics (BLS) have a following of reporters and analysts who celebrate "jobs day" each month with a race to make the most interesting charts and analyses based on the latest data. The monthly BLS Employment Situation news release includes two different estimates of employment from two different surveys. The payroll employment estimates are based on the monthly survey of businesses and government agencies, whereas the household employment estimates are based on a monthly survey of households (see `https://www.bls.gov/web/empsit/ces_cps_trends.htm` for more details on the two surveys). The data are typically released at 8:30 am ET on the first Friday following the reference month. In the short term, the household estimates are not revised but the payroll estimates are revised to incorporate additional data over the next two months. In the longer term, the payroll estimates are revised each year in a benchmarking process that re-anchors the employment estimates to the full population counts based on unemployment insurance records which cover approximately 97 percent of employment (see `www.bls.gov/opub/mlr/2017/article/benchmarking-the-current-employment-statistics-national-estimates.htm`). There are also annual population control adjustments that are introduced each January into the household series without historical revisions. Some household numbers are affected by annual revisions to the seasonal adjustment factors. Wu (2004) discusses the pros and cons of the two different surveys for tracking employment in real time.

An obvious question arises with two different surveys focused on the same underlying economic variable—is there a way to combine the information from the surveys to obtain an improved estimate of employment in real time?[1] This question is particularly important in the US employment case because the payroll survey is generally the one that is given more attention due to its larger sample size, but the revision patterns in the payroll data may make it not as useful in real time.[2] For example, Neumark and Washer (1991) find

[1]Justin Wolfers has suggested a rule of thumb on Twitter with 80 percent on the payroll survey, 20 percent on the household survey: `https://twitter.com/JustinWolfers/status/431783260524265472?s=20`.

[2]A third series, the ADP employment report (`https://adpemploymentreport.com/`), is an estimate of US employment from a private payroll processing company and is released a couple of days before the

that preliminary payroll employment estimates are not “efficient forecasts” of the revised estimates, Phillips and Nordlund (2012) find evidence of a cyclical bias in the early payroll employment estimates, and Owyang and Vermann (2014) document systematic bias in the revisions of the payroll employment data. Furthermore, although we are focused on shorter term revisions, Haltom et al. (2005) find that previous (longer term) benchmark data revisions help predict future employment benchmark data.

Several studies have tried to determine which measure should be preferred in various contexts, but there is a growing literature focusing instead on reconciliation of multiple measures in order to incorporate more information. Much of the argument for reconciliation has focused on different measures of output, for example, Stone et al. (1942), Weale (1992) and Diebold (2010). Early models of reconciliation relied on the assumption that measurement errors are “noise”, which in turn forces the reconciled estimate of the latent variable to be less variable than any of the individual series being reconciled. Aruoba et al. (2013) consider the problem from a forecast combination perspective, assuming “news” errors and imposing priors to address identification, while Aruoba et al. (2016) consider alternative identifying assumptions and propose the addition of an instrumental variable. Almuzara et al. (2021) investigate a dynamic factor model (DFM) with cointegration restrictions while Anesti et al. (2022) propose a mixed-frequency Release-Augmented DFM. See Jacobs et al. (2020) for details. Our contribution to this literature is to extend the analysis to focus on employment instead of output and to incorporate the unique features of the payroll and household surveys used to measure U.S. employment.

In this paper we construct a latent employment estimate which reconciles the information from the two separate surveys as well as incorporating the preliminary data revision process of the payroll data. Our model builds upon a version of Jacobs et al. (2020). We find, similar to the Council of Economic Advisers (2017), that the household survey is in general not very informative, but we also find that our reconciled latent employment series looks somewhat different than the initial release of payroll employment and is closer to the

BLS estimates. Until recently the publicly available series was a forecast of employment rather than the microdata from the firm. Researchers from the Federal Reserve have access to the microdata and have combined it with the payroll data in Cajner et al. (2019). On August 31, 2022, ADP released a newly revised methodology, but the data only goes back to 2010 and is not in real time. This is not a long enough sample to use in our analysis.

benchmarked data. This is only true, however, for the full sample. Once we move to a real-time exercise, our findings suggest that the reconciled employment estimate is remarkably similar to the initial release of payroll employment. We also find that the payroll series is predominantly news, with much of the news in the first estimate, whereas the household series is almost all noise. These results confirm the approach of focusing attention on the preliminary release of payroll data when analyzing the US labor market.

In Section 2, we describe our data and estimation method. Results are shown in Section 3 and Section 4 concludes. An Appendix provides additional results and a detailed explanation of the Bayesian estimation method.

2 Assumptions, Data, and Estimation

2.1 Some assumptions and notation

We have two data series: Payroll (PR) employment x_t and Household (HH) employment z_t. We model them in first differences, since that is how employment is typically discussed. Working with first differences also ensures that both series are stationary and mitigates problems associated with benchmark or historical revisions. See e.g. Croushore (2011). Estimating the models in growth rates produced similar results.

Both series are reported at the same frequency and are released at the same time. Our Household survey data's first vintages start in February 1961 (total employment starts in January 1961, so its monthly changes start in February 1961). Our last available data vintage ends in June 2022.

We use seasonally adjusted data[3] and ignore seasonal factor changes to treat z_t as if it is not revised (although we do use original vintage data in the estimates). We explicitly model the more substantive revisions of x_t which occur in the first three releases.[4] Let x_t^j

[3]Employment data are typically seasonally adjusted with the Census X13-ARIMA-SEATS method, the combination of Census X12-ARIMA and TRAMO-Seats which has become the industry standard (Department of Commerce Census Bureau `https://www.census.gov/data/software/x13as.html`). The method produces the best trend/cycle and season estimates in 'normal' circumstances. However its performance during crises is strange. For details, see Abeln and Jacobs (2023).

[4]Our modeling framework allows for l different releases of x_t. There's nothing that requires these l releases to be consecutive, but they need to be in order. For example, when $l = 3$ we could use releases $(1, 2, 3)$ or $(1, 3, 12)$ but not $(1, 12, 3)$.

be the jth published estimate of x_t and $\boldsymbol{X}_t \equiv [x_t^1, x_t^2, x_t^3]$.

Both the HH and PR series are measures of the same macroeconomic concept but do so with some degree of error as well as differences in definition. We call the macroeconomic concept "employment" m_t and treat it as a latent variable. The Bureau of Labor Statistics (BLS) also releases (on the same day) a version of the HH survey that is modified to match the payroll definition, but that data is only available back to 1994 and includes data revisions which affect the real-time estimates, so for our analysis we use the unmodified HH data. We also estimated models using the nonagricultural wage and salary employment series from the household survey, which the BLS suggests is the closest proxy series available for the full sample, which we found gave similar results. An anonymous referee pointed out that the HH series includes potential level shifts in January of each year due to changes in population controls. We thus also estimated models treating those January observations as missing and again find similar results to what is reported here. (These additional results are available from the authors upon request.)

Due to differences in the coverage of the HH and PR employment measures, we include a constant (μ_Z) and a slope coefficient (β_Z) in the relationship between HH and the latent variable, and interpret the latent variable as having the same units as the payroll definition.

The measurement error in PR for each release j (where $j = 1, 2, 3$) is $e_t^j \equiv x_t^j - m_t$. Each e_t^j may be the sum of a pure news (ν_t^j) and a pure noise (ϵ_t^j) component, both of which have means of zero. The measurement error in HH ($e_t^H \equiv z_t - m_t$) is similarly defined as the sum of a pure news (ν_t^H) and a pure noise (ϵ_t^H) component, both of which have mean zero. These two components could be correlated (to varying degrees) with ν_t^j or ϵ_t^j respectively.

We evaluate the quality of the latent employment estimate by comparing it to the benchmarked Establishment Survey Estimates: `https://www.bls.gov/web/empsit/cesbmart.htm`. At the time of writing, the most recent benchmarked data was released on February 4, 2022 (`https://www.bls.gov/news.release/archives/empsit_02042022.htm`) and provides data through March 2021. We treat these as final numbers, although small revisions due to seasonal adjustment happen for 5 years, and occasional benchmark or other revisions may change the whole vintage.

2.2 Data

The need to reconcile employment estimates in real time requires that we take account of the revisions process of our series. All our data are from the Employment Situation releases from the Bureau of Labor Statistics as provided by ALFRED, the Federal Reserve Bank of St. Louis' ArchivaL Federal Reserve Economic Data. We focus on two seasonally adjusted data series: (1) Payroll employment (PR): total nonfarm payroll employment from the Current Employment Statistics (i.e. the Establishment Survey) and (2) Household employment (HH): civilian employment from the Current Population Survey (i.e. the Household Survey). We focus on the initial release of the (change in) employment from both the household and the payroll surveys. Since the payroll survey follows a regular pattern of revisions in the following two months, we also include the second and third releases of payroll survey data.

Our analysis focuses on data from 1961 through 2019. In the appendix, we extend the analysis to include data through early 2022 to explore the robustness of our results in light of the unprecedented effects of the COVID-19 pandemic on employment growth.

Figure 1 presents the initial releases of the change in employment from the two surveys. From this figure we can see the greater volatility of the household numbers (reflecting its smaller sample size.) This greater volatility can also be seen in the descriptive statistics reported in Table 1; although the means and medians of the various series are broadly similar, the standard deviation for the household data is substantially larger than that for the payroll data. There is also a slight tendency for the series to increase as they are revised; the mean and median for the benchmarked data are both higher than initial release data, with a statistically significant difference in mean of about 20,000. This difference becomes less for the second revision (but still statistically significant) and by the third revision the difference is less than 3,000 and statistically insignificant.

[Figure 1 about here.]

[Table 1 about here.]

2.3 Estimation

We estimate the model with one release of HH (z_t), and three releases for PR (x_t^j).

State Vector:

$$\boldsymbol{\alpha}_t = \left[m_t, e_t^H, e_t^1, e_t^2, e_t^3\right]', \tag{1}$$

where the measurement errors e_t^H (for the household survey) and e_t^j, $j = 1, 2, 3$ (for the three payroll survey releases) are the sum of news and noise errors; $e_t^i = \nu_t^i + \varepsilon_t^i$ where $i = H, 1, 2, 3$.

Measurement Equation:

$$\begin{bmatrix} z_t \\ x_t^1 \\ x_t^2 \\ x_t^3 \end{bmatrix} = \begin{bmatrix} \mu_z \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \beta_z & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \boldsymbol{\alpha}_t. \tag{2}$$

State Equation:

$$\boldsymbol{\alpha}_t = \boldsymbol{T} \cdot \boldsymbol{\alpha}_{t-1} + \boldsymbol{R} \cdot \boldsymbol{\eta}_t, \tag{3}$$

where

$$\boldsymbol{T} = \begin{bmatrix} \rho & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{R} = \begin{bmatrix} \sigma_\nu^H & \sigma_\nu^1 + \sigma_\nu^{H1} & \sigma_\nu^2 + \sigma_\nu^{H2} & \sigma_\nu^3 + \sigma_\nu^{H3} & 0 & 0 & 0 & 0 \\ -\sigma_\nu^H & -\sigma_\nu^{H1} & -\sigma_\nu^{H2} & -\sigma_\nu^{H3} & \sigma_\epsilon^H & \sigma_\epsilon^{H1} & \sigma_\epsilon^{H2} & \sigma_\epsilon^{H3} \\ 0 & -\sigma_\nu^1 & -\sigma_\nu^2 & -\sigma_\nu^3 & 0 & \sigma_\epsilon^1 & 0 & 0 \\ 0 & 0 & -\sigma_\nu^2 & -\sigma_\nu^3 & 0 & 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & 0 & -\sigma_{\boldsymbol{\nu}}^3 & 0 & 0 & 0 & \sigma_\epsilon^3 \end{bmatrix},$$

$$\eta_t = \left[\nu_t^H, \nu_t^1, \nu_t^2, \nu_t^3, \epsilon_t^H, \epsilon_t^1, \epsilon_t^2, \epsilon_t^3\right]'.$$

2.4 Bayesian estimation

For our preferred specification, we estimate the parameters using Bayesian methods. We generate draws from the posterior distributions using a random walk Metropolis-Hastings algorithm.[5] A detailed explanation of the Bayesian estimation is in Appendix B. We make 50,000 draws in total but throw away the first 5,000 draws as a burn-in period. Thereafter, we draw the median, 25th percentile, and 75th percentile of the draws from the posterior distribution. Table 2 summarizes the specification of the prior distribution. Given the median parameter draw, we estimate the state variables using the Kalman filter.

[Table 2 about here.]

3 Results

3.1 Results for the full 1961–2019 sample

Table 3 provides the parameter estimates for the model estimated over the full sample of February 1961 through December of 2019. There we see that the news component for the first release of the PR series (given by $\sigma_\nu^1 = 33.847$) is slightly smaller than that of its noise component (given by $\sigma_\epsilon^1 = 49.168$.) However the two subsequent releases add substantially more news ($\sigma_\nu^2 = 35.913$ and $\sigma_\nu^3 = 22.166$) and less noise ($\sigma_\epsilon^2 = 8.672$ and $\sigma_\epsilon^3 = 13.416$.)

At first glance, the HH series might appear to be a more reliable guide, with a larger news component ($\sigma_\nu^H = 96.931$) and only somewhat more noise ($\sigma_\epsilon^H = 53.760$.) However, that ignores the contributions of both news and noise that is correlated with their respective components in the PR series (σ_ν^{H1} through σ_ν^{H3} and σ_ϵ^{H1} through σ_ϵ^{H3}, respectively.) These add only modestly to the overall news component (due in large measure to the substantial negative component found in σ_ν^{H3}) while greatly increasing the noise component (particularly with $\sigma_\epsilon^{H2} = 283.791$.)

[5] Chib (1995) contains a comprehensive introduction of the Metropolis-Hastings algorithm. This method has been intensively used to estimate state space models (see e.g. Smets and Wouters (2007) and Aruoba et al. (2016)).

[Table 3 about here.]

The Kalman gains implied by the above parameter estimates are reported in Table 4. They show that the model puts almost all weight on the second revision (third release) of the payroll data. This may seem unsurprising since the objective of the revisions is to improve the estimates, but it is notable that our reconciliation model picks this up so clearly without being provided any final target. In the presence of large noise component in the final estimate, for example, one might expect the model to put more weight on other releases to mitigate the effect of noise errors.

[Table 4 about here.]

Figure 2 presents our smoothed estimates of the latent employment change series based on our preferred model along with the initial payroll estimate.[6] These two series appear similar, but there are some differences, particularly in the early 1980s and again around the Great Recession.

[Figure 2 about here.]

3.2 Real-time results

The above smoothed estimates based on full sample parameter estimates do not show whether or not we can improve upon initial releases of payroll employment in real time. To address this question we examine our model's nowcasts starting from January of 1990. These are estimates of the reconciled series at month $t-1$ as could be constructed at the start of month t. These are filtered (rather than smoothed) estimates, and take into account the missing values $dx^2_{t-1}, dx^3_{t-1}, dx^3_{t-2}$. They use parameter values estimated using an expanding window based on the data vintages as published in month t. We find two key results from these estimates. First, the nowcast series closely resembles the initial payroll release, as can be seen from Figure 3. Second, as can be seen from the Kalman gains reported in Figure 4, the household series dz_t never receives much weight, despite the absence of recent revised estimates of the payroll series.

[6] A smoothed estimate for period t is the expected value of the latent series m_t conditional on the parameter estimates $\hat{\theta}$ and all our data series from $t = 1$ to T.

[Figure 3 about here.]

[Figure 4 about here.]

3.3 Comparing the reconciled series with benchmarked employment

An interesting feature of the payroll series is that the data are eventually benchmarked to a near census of employment. As discussed in the Data section above, the latest benchmarked employment data as of this analysis is through March of 2021, meaning we have the near "truth" to evaluate the results we have for the data through 2019. In order to evaluate the performance of our latent series, we compare our series with the benchmarked data.

Table 5 presents the mean absolute errors (MAE) of using the latent series for the benchmarked series based on our full sample and nowcast models. It also shows the results of using initial payroll release data as our latent series. This analysis is consistent with our previous findings that with the full sample we can find some improvements from our model, but for nowcasts we might just as well use just the initial payroll release.

[Table 5 about here.]

We also considered the performance of the different models in different states of the economy.[7] The NBER only announces the peak and trough dates of the business cycle with a lag so this information is not available in real time, but an ex post analysis shows that generally the models perform roughly 25% worse (in terms of MAE for the fully revised data) in recessions than in expansions.

4 Conclusion

Unlike earlier work using payroll and household survey data to forecast employment growth, we've introduced a structural model of news and noise measurement errors to explore how these two data sources may best be reconciled to produce combined estimates of

[7]Results available from the authors upon request.

employment growth. The state-space framework used shows how those estimates should be formulated when initial estimates are released, and how they should be updated as revised payroll estimates become available.

Our results are consistent with conclusions from work on employment forecasting and strikingly different from those on reconciling expenditure and income estimate of real GDP. We find that data from the household survey receive only trivial weight in reconciliation, regardless of whether or not a full set of revised payroll estimates are available. Our reconciled estimate closely tracks the latest available estimate of the payroll series, which has notable deviations from the initial estimates at times, particularly in the recessions of the early 1980s and 2008-09. However, in real time our reconciliation of the household and establishment surveys puts near-exclusive weight on the best available payroll estimates.

These findings suggest that the employment data from the household survey does not improve real time employment estimates, but that does not mean that the household survey is not useful. It provides information available only from households (e.g. on demographics, unemployment, etc.). Furthermore, the information reported in ratios from the household survey (such as the unemployment rate) are useful in other contexts.

Conflict of Interest Statement

In addition to her academic affiliations, Tara Sinclair had an employment relationship with job search site Indeed during the writing of this paper, but neither Indeed nor the Indeed Hiring Lab have any stake in this research. Sinclair is now with the U.S. Department of Treasury. The views expressed here are the authors' personal views and do not necessarily reflect those of the Department of the Treasury or the U.S. Government. The remaining authors have no conflicts of interest relevant to this research.

References

Abeln, Barend and Jan P.A.M. Jacobs (2023). *Seasonal Adjustment without Revisions: A Real-Time Data Approach*. SpringerBrief. [forthcoming]. Berlin: Springer.

Almuzara, Tincho, Gabriele Fiorentini and Enrique Sentana (2021). *U.S. aggregate output measurement: A common trend approach*. Staff Report 962. FRB of New York.

Anesti, Nikoleta, Ana Galvão and Silvia Miranda-Agrippino (2022). 'Uncertain Kingdom: Nowcasting GDP and its revisions'. In: *Journal of Applied Econometrics* 37, pp. 42–62.

Aruoba, S. Boragan, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide and Dongho Song (2013). 'Improving US GDP measurement: A forecast combination perspective'. In: *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*. Ed. by X. Chen and N. Swanson. New York: Springer, pp. 1–26.

— (2016). 'Improving GDP measurement: A measurement-error perspective'. In: *Journal of Econometrics* 191.2, pp. 384–397.

Cajner, Tomaz, Leland D Crane, Ryan A Decker, Adrian Hamins-Puertolas and Christopher Kurz (2019). *Improving the Accuracy of Economic Measurement with Multiple Data Sources: The Case of Payroll Employment Data*. Working Paper 26033. National Bureau of Economic Research.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino and Elmar Mertens (2021). *Addressing COVID-19 outliers in BVARs with stochastic volatility*. Tech. rep. No. DP15964. CEPR.

Chib, Siddhartha and Edward Greenberg (1995). 'Understanding the Metropolis-Hastings algorithm'. In: *The American Statistician* 49.4, pp. 327–335.

Council of Economic Advisers (CEA) (2017). *Assessing the State of the Economy in Real Time Using Headline Economic Indicators*. Issue Brief.

Croushore, Dean (2011). 'Frontiers of real-time analysis'. In: *Journal of Economic Literature* 49, pp. 72–100.

Diebold, Francis X. (2010). 'Comment'. In: *Brookings Papers on Economic Activity* Spring, pp. 107–112.

Dowd, Kevin, John Cotter, Christopher Humphrey and Margaret Woods (June 2008). 'How Unlucky Is 25-Sigma?' In: *Journal of Portfolio Management* 34, pp. 76–80. DOI: 10.3905/jpm.2008.709984.

Haltom, Nicholas L., Vanessa D. Mitchell and Ellis W. Tallman (2005). 'Payroll employment data: Measuring the effects of annual benchmark revisions'. In: *Economic Review-Federal Reserve Bank of Atlanta* 90.2, pp. 1–23.

Jacobs, Jan P.A.M., Samad Sarferaz, Jan-Egbert Sturm and Simon van Norden (2020). 'Can GDP measurement be further improved? Data revision and reconciliation'. In: *Journal of Business Economics & Statistics* 40, pp. 423–431.

Lenza, Michele and Giorgio E. Primeceri (2022). 'How to estimate a vector autoregression after March 2020'. In: *Journal of Applied Econometrics* 37, pp. 688–699.

Neumark, David and William L Wascher (1991). 'Can we improve upon preliminary estimates of payroll employment growth?' In: *Journal of Business & Economic Statistics* 9.2, pp. 197–205.

Ng, Serena (2021). *Modeling Macroeconomic Variations after COVID-19*. arXiv:2130:02732.

Owyang, Michael and E. Katarina Vermann (2014). 'Employment Revision Asymmetries'. In: *Economic Synopses* 11.4, pp. 1–2.

Phillips, Keith R. and James Nordlund (2012). 'The efficiency of the benchmark revisions to the current employment statistics (CES) data'. In: *Economics Letters* 115.3, pp. 431–434.

Schorfheide, Frank and Dongho Song (2020). *Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic*. Working Paper 20-039. PIER.

Smets, Frank and Rafael Wouters (2007). 'Shocks and frictions in US business cycles: A Bayesian DSGE approach'. In: *American Economic Review* 97.3, pp. 586–606.

Stone, Richard, D. G. Champernowne and J. E. Meade (1942). 'The precision of national income estimates'. In: *Review of Economic Studies* 9, pp. 111–125.

Weale, Martin (1992). 'Estimation of Data Measured with Error and Subject to Linear Restrictions'. In: *Journal of Applied Econometrics* 7, pp. 167–174.

Wu, Tao (2004). *Two measures of employment: How different are they?* FRBSF Economic Letter. Federal Reserve Bank of San Francisco.

Appendices

A Extending the analysis through 2022

Based on the results during the Great Recession, it may be the case, at least in sample, that our latent series performs differently than the initial release of the PR series in recessions. Since we experienced a dramatic and unusual recession in 2020, we re-estimate the model using data through 2022, on both the full-sample and real-time estimation, to compare the estimates of the latent series.

[Figure 5 about here.]

To put these results into context, note that they are based on adding 30 new observations to arrive at a sample size of 737 months. However, owing to the start of the Covid-19 pandemic in early 2020, some of those observations are extreme, as Figure 5 illustrates. The values for April 2020 are notable, as those shown for the PR series are roughly 100 standard deviations below the series mean.[8] Figure 5 compares the real-time estimates of our latent series and the initial payroll releases. Until the beginning of the pandemic, the two series are very close to each other, since the latent series put almost the entire weight on the initial PR release. However, after April 2020, the deviations between the two series clearly appears, indicating the distortion of the weight structure of the latent series.

[Table 6 about here.]

Adding the post-2019 data changes the estimated parameters as shown in Table 6. Despite adding only 30 more observations, we find that the serial correlation in m_t drops dramatically compared to the earlier estimates and becomes nearly zero. We also see

[8] The sample mean of the 3rd release of the PR series up to the end of 2019 is 136,200 while its sample standard deviation is 201,800. The value for April 2020 is -20,787,000, which is 103.7 standard deviations below the mean. To put this into perspective, Dowd et al. (2008) calculate that (if we assume normality) the probability of observing an event at least 8 sigma below the mean is roughly equal to 1 divided by the number of days since the Big Bang, while observing an event 20 sigma below the mean is roughly equal to 1 divided by the number of particles in the universe. The probability of a 38-sigma event is roughly 10^{-316} and beyond this point the limitations of IEEE double-precision values (which cannot represent anything closer to zero than 4.94066×10^{-324}) complicate calculation of the probability. Using arbitrary-precision computations provides a estimated probability for events at least 103.7 standard deviations below the mean of roughly 6×10^{-2337}. Ng (2021) independently makes a similar observation.

some changes in terms of the role of independent news and noise shocks. The PR series adds more news components ($\sigma_\nu^1 = 52.765$, $\sigma_\nu^2 = 7.482$, and $\sigma_\nu^3 = 191.140$) than the noise component ($\sigma_\epsilon^1 = 33.609$, $\sigma_\epsilon^2 = 2.259$, and $\sigma_\epsilon^3 = 45.363$). The news component of the HH series ($\sigma_\nu^H = 296.309$) is larger than the noise component ($\sigma_\epsilon^H = 150.932$).

The contributions of the correlated news and noise have changed remarkably compared to the earlier estimates. These add substantially to the overall news component (particularly with $\sigma_\nu^{H3} = 631.660$), while negatively influencing the overall noise component (particularly with $\sigma_\epsilon^{H2} = -141.981$).

[Table 7 about here.]

These dramatic changes in the parameters alter the importance of the HH series relative to the PR series as well as the relative importance of each release within the PR series. Table 7 shows the Kalman gains based on the parameter estimates through 2022. Compared to the earlier estimate in the third column, the weight of the HH series increases from near zero per cent to a non-negligible weight of 25 per cent. Also, the weights of the first and second releases of the PR series increase. On the other hand, the weight of the third release of the PR series drops from nearly 90 per cent to -46 per cent.

[Figure 6 about here.]

The weight structure negatively affects the performance of our latent series in terms of closeness to the benchmarked employment. As seen in Table 5, the MAE of the earlier estimates from 1961 to 2019 is 72.013. However, using the parameter estimates through 2022, the MAE from 1961 to 2019 now becomes 105.99.[9]

Several recent papers propose methods for dealing with the unusual volatility seen during the early phases of the COVID pandemic, (e.g., Lenza and Primiceri, 2022, Schorfheide and Song, 2020, Carriero et al., 2021). To deal with the extreme changes in employment in 2020, we treat the data from March 2020 through August 2020 as missing and re-estimate the model.[10] The results from the alternative "dummying out COVID" estimates are com-

[9]The results are available upon request.

[10]To determine which months to drop, we take the last five years of payroll data and calculate the pre-pandemic average and standard deviation. Then we identify the months with changes over 10 sigma from that pre-pandemic mean. This approach suggests we dummy out March 2020 through August 2020.

parable to the earlier estimates of data from 1961 to 2019 in terms of parameter estimates, Kalman gains, and MAE.

Figure 6 shows the deviations of the first release of the PR series (**PR1**) and three measures of the reconciled series: the smoothed estimate (using parameter estimates from the expanded 1961-2022 sample), the real-time estimate, and the alternative smoothed estimate (using parameter estimates from the expanded 1961-2022 sample without the March-August 2020 observations.) Due to the changes in the parameter estimates, the deviation of the updated smoothed series estimates large prior to the pandemic. However, this deviation dramatically decreases when the alternative smoothed series is used. Focusing on the alternative smoothed series, we see that the big differences from the other two estimates are in March 2020 through May 2020.

B A Detailed Explanation of the Bayesian Estimation

Given the assumption that the error terms of the state space model follow the Gaussian distribution, the density of the data $f(\boldsymbol{Y}_t|\boldsymbol{\alpha}_t)$, where $\boldsymbol{Y}_t = \left[z_t x_t^1 \dots x_t^l\right]'$, is:

$$f(\boldsymbol{Y}_t|\boldsymbol{\alpha}_t) = (2\pi)^{-\frac{1}{2}}|\boldsymbol{f}_{t,t-1}|^{-\frac{1}{2}}exp(-0.5\boldsymbol{u}_{t,t-1}\boldsymbol{f}_{t,t-1}^{-1}\boldsymbol{u}_{t,t-1}),$$

where $\boldsymbol{u}_{t,t-1}$ is the predictive error and $\boldsymbol{f}_{t,t-1}$ is the variance of the predictive error of the Kalman filter. The likelihood function of the model is:

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} f(\boldsymbol{Y}_t|\boldsymbol{\alpha}_t),$$

where $\boldsymbol{\theta} = \{\rho, \sigma_\nu^H, \sigma_\nu^1, \sigma_\nu^2, \sigma_\nu^3, \sigma_\nu^{H1}, \sigma_\nu^{H2}, \sigma_\nu^{H3}, \sigma_\epsilon^H, \sigma_\epsilon^1, \sigma_\epsilon^2, \sigma_\epsilon^3, \sigma_\epsilon^{H1}, \sigma_\epsilon^{H2}, \sigma_\epsilon^{H3}, \mu_z, \beta_z\}$.

We conduct the random walk Metropolis-Hastings approach in the following steps:
Step 1: Specify a starting value $\boldsymbol{\theta}_0$ and variance of the shock $\boldsymbol{\Sigma}$.
Step 2: Draw a new parameter vector from the random walk equation:

$$\boldsymbol{\theta}_{NEW} = \boldsymbol{\theta}_{OLD} + \boldsymbol{e} \qquad \boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

Step 3: Compute the acceptance probability:

$$\alpha = \min\left(\frac{f(\boldsymbol{Y}|\boldsymbol{\theta}_{NEW})p(\boldsymbol{\theta}_{NEW})}{f(\boldsymbol{Y}|\boldsymbol{\theta}_{OLD})p(\boldsymbol{\theta}_{OLD})}, 1\right),$$

where $p(\boldsymbol{\theta}_i)$ is the prior density.

Step 4: If $\alpha > a \sim \mathcal{U}(0,1)$, obtain $\boldsymbol{\theta}_{NEW}$. Otherwise $\boldsymbol{\theta}_{NEW} = \boldsymbol{\theta}_{OLD}$.

Repeat steps 2, 3, and 4 50,000 times. The first 5,000 draws are a burn-in period. After the burn-in period, we draw the median, 25th percentile, and 75th percentile of the draws from the posterior distribution.

List of Figures

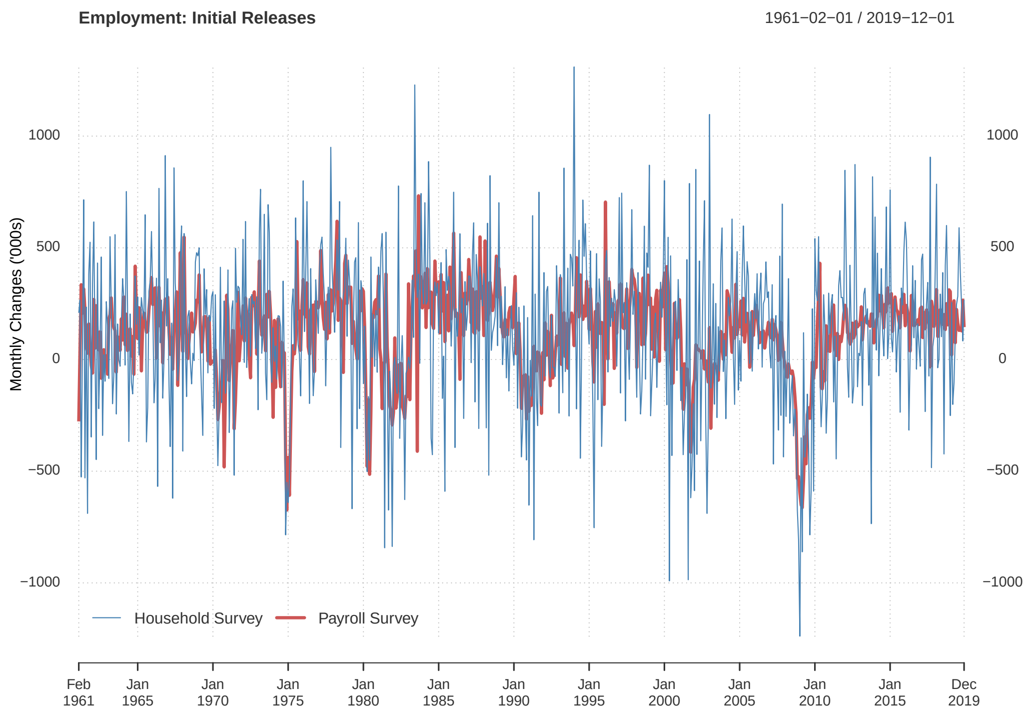

Figure 1: U.S. Employment: Initial Release of Monthly Changes, Thousands of Persons, Feb 1961–Dec 2019

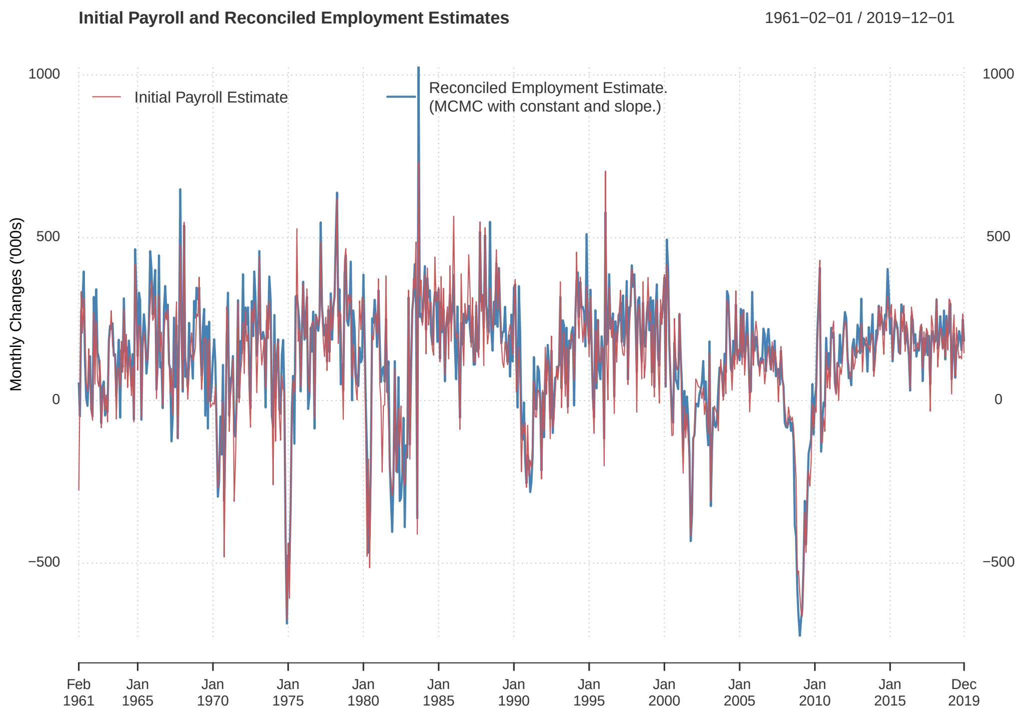

Figure 2: Initial Release of Monthly Payroll Changes and Smoothed Full Sample Latent (Reconciled) Estimate, Thousands of Persons, Feb 1961–Dec 2019

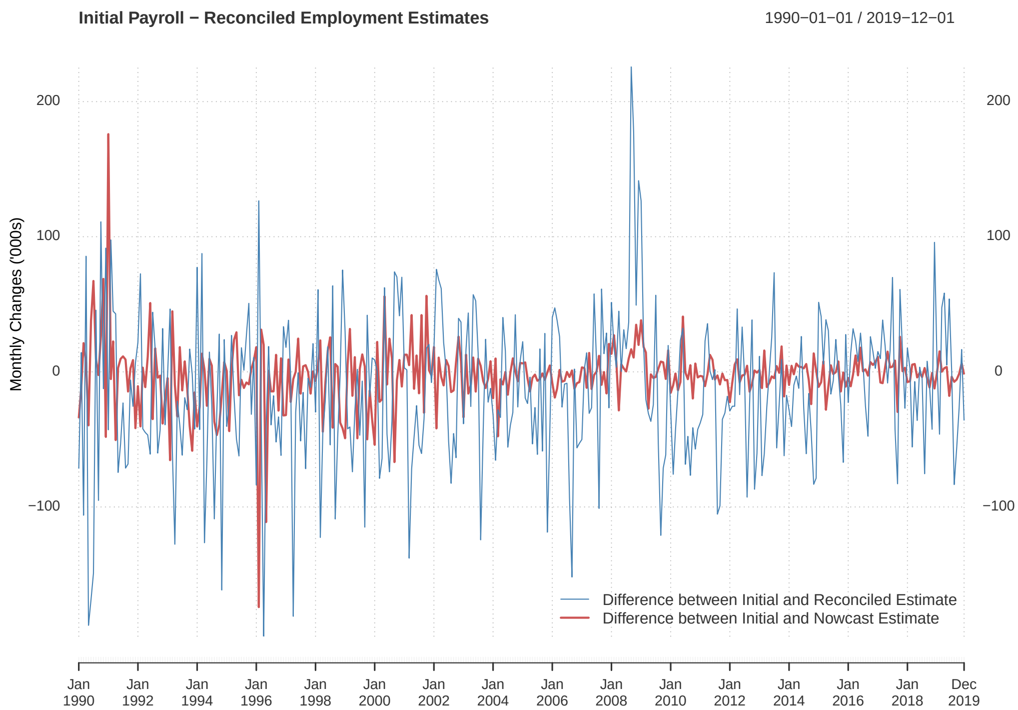

Figure 3: Comparing Difference Between Latent and Initial Release and Nowcast and Initial Release of Employment Change, Thousands of Persons, Jan 1990–Dec 2019

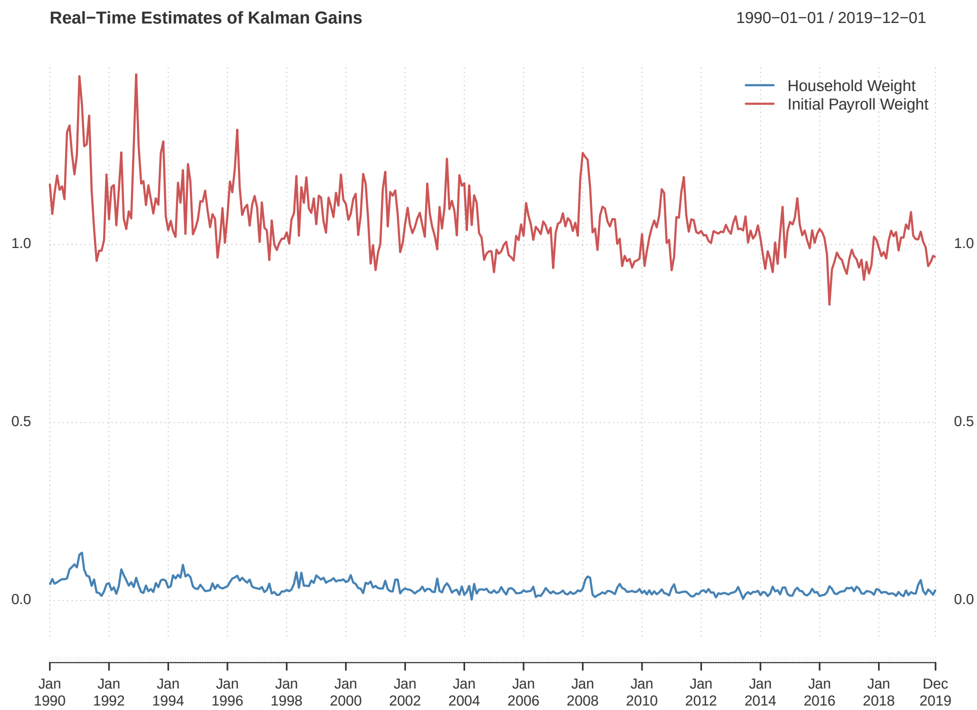

Figure 4: Real-Time Estimates of Kalman Gains

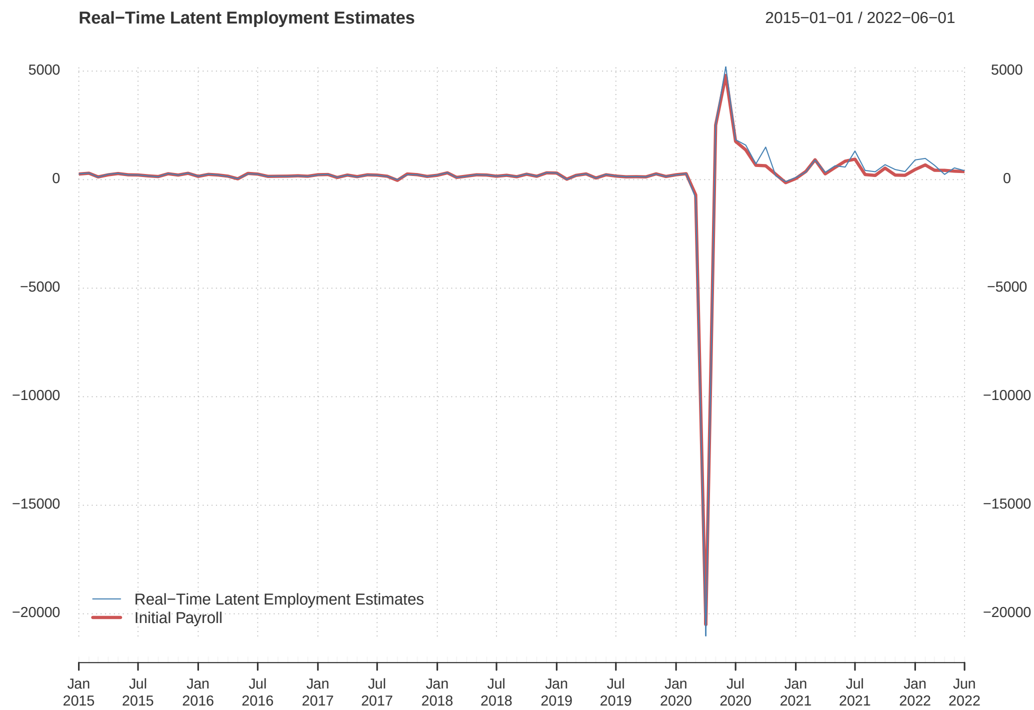

Figure 5: The COVID-19 Pandemic in Perspective

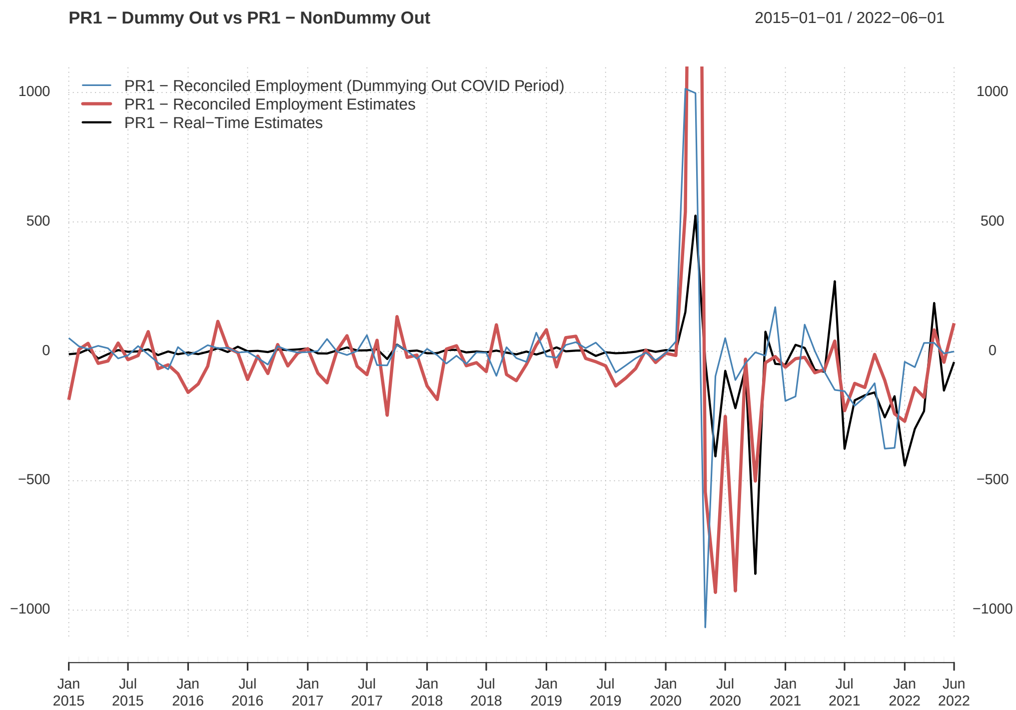

Figure 6: Performance of Dummying Out COVID Period

List of Tables

Table 1: Descriptive statistics for monthly change in employment Feb 1961–Dec 2019

	Initial Household dz	Initial Payroll dx1	First Revision Payroll dx2	Second Revision Payroll dx3	Benchmarked Payroll dfinal
Mean	125	119	124	136	139
Median	157	145	157	165	172
Standard Deviation	351	190	197	202	197
Max	1310	733	1018	1103	1118
Min	−1239	−674	−699	−741	−800
Observations	707	707	707	707	707

Change in employment in '000s.

Table 2: Prior Specification

Parameter	Density	Parameter 1	Parameter 2
ρ	Normal	0.5	0.2
σ_ν^H	Uniform	0	2000
σ_ν^1	Uniform	0	2000
σ_ν^2	Uniform	0	2000
σ_ν^3	Uniform	0	2000
σ_ν^{H1}	Uniform	-1000	1000
σ_ν^{H2}	Uniform	-1000	1000
σ_ν^{H3}	Uniform	-1000	1000
σ_ϵ^H	Uniform	0	2000
σ_ϵ^1	Uniform	0	2000
σ_ϵ^2	Uniform	0	2000
σ_ϵ^3	Uniform	0	2000
σ_ϵ^{H1}	Uniform	-1000	1000
σ_ϵ^{H2}	Uniform	-1000	1000
σ_ϵ^{H3}	Uniform	-1000	1000
μ_z	Normal	$\bar{x^3} - \bar{z}$	100
β_z	Normal	1	0.2

Note: Parameter 1 is the mean of the normal distribution and the minimum value of the uniform distribution. Parameter 2 is the standard deviation of the normal distribution and the maximum value of the uniform distribution.

Table 3: Parameter estimates for the full Feb 1961–Dec 2019 sample

Parameter	Estimate
ρ	0.629
	(0.598,0.672)
σ_ν^H	96.931
	(55.005,124.752)
σ_ν^1	33.847
	(26.281,46.902)
σ_ν^2	35.913
	(34.231,37.594)
σ_ν^3	22.166
	(9.738,39.445)
σ_ν^{H1}	68.697
	(28.188,107.247)
σ_ν^{H2}	8.215
	(2.594,12.124)
σ_ν^{H3}	-57.578
	(-95.899,-22.966)
σ_ϵ^H	53.760
	(25.096,87.645)
σ_ϵ^1	49.168
	(37.770,53.843)
σ_ϵ^2	8.672
	(7.295,10.150)
σ_ϵ^3	13.416
	(8.833,17.278)
σ_ϵ^{H1}	65.929
	(49.649,105.863)
σ_ϵ^{H2}	283.791
	(272.397,295.691)
σ_ϵ^{H3}	51.278
	(17.760,82.267)
μ_z	6.925
	(3.063,11.255)
β_z	1.314
	(1.286,1.342)
log likelihood	-17,167.901

Notes. Model applied to data in first differences. Values between parentheses represent 25th% and 75th% draws for MCMC estimates.

Table 4: Kalman Gain Full Sample Estimates Feb 1961–Dec 2019

Series	Kalman Gain
Initial Household (dz)	0.003
Initial Payroll (dx1)	-0.058
First Revision Payroll (dx2)	0.100
Second Revision Payroll (dx3)	0.895

Table 5: Mean Absolute Errors Comparing with Benchmarked Employment

Estimation	MAE 1961-2019	MAE 1990-2019
PR First Release	82.505	72.597
Smoothed Full Sample	72.013	65.367
Real Time Jan 1990-Dec 2019		75.999

Table 6: Parameter estimates comparing samples

Sample	1961–2022	1961-2019
ρ	-0.034	0.629
	(-0.044,-0.016)	(0.598,0.672)
σ_ν^H	296.309	96.931
	(141.233,439.278)	(55.005,124.752)
σ_ν^1	52.765	33.847
	(45.310,59.190)	(26.281,46.902)
σ_ν^2	7.482	35.913
	(4.344,10.463)	(34.231,37.594)
σ_ν^3	191.140	22.166
	(168.490,216.613)	(9.738,39.445)
σ_ν^{H1}	147.368	68.697
	(118.915,187.398)	(28.188,107.247)
σ_ν^{H2}	314.097	8.215
	(95.857,413.101)	(2.594,12.124)
σ_ν^{H3}	631.660	-57.578
	(541.371,708.090)	(-95.899,-22.966)
σ_ϵ^H	150.932	53.760
	(82.539,234.344)	(25.096,87.645)
σ_ϵ^1	33.609	49.168
	(20.007,42.502)	(37.770,53.843)
σ_ϵ^2	2.259	8.672
	(1.240,3.562)	(7.295,10.150)
σ_ϵ^3	45.363	13.416
	(44.329,46.652)	(8.833,17.278)
σ_ϵ^{H1}	-22.466	65.929
	(-40.658,2.349)	(49.649,105.863)
σ_ϵ^{H2}	-141.981	283.791
	(-235.816,-28.883)	(272.397,295.691)
σ_ϵ^{H3}	56.583	51.278
	(47.935,65.568)	(17.760,82.267)
μ_z	5.420	6.925
	(0.675,9.744)	(3.063,11.255)
β_z	1.751	1.314
	(1.701,1.793)	(1.286,1.342)
log likelihood	-19,319.350	-17,167.901

Notes.
Model applied to data in first differences.
Figures in parentheses give 25th% and 75th% draws for MCMC estimates.

Table 7: Kalman Gain Sample Estimates Dummying Out the COVID Period Feb 1961–Jun 2022

Series	1961-2022	1961-2019	1961-2022 (Dummy Out COVID)
Initial Household (dz)	0.256	0.003	-0.010
Initial Payroll (dx1)	0.522	-0.058	0.060
First Revision Payroll (dx2)	0.862	0.100	0.103
Second Revision Payroll (dx3)	-0.461	0.895	0.899