## **Editor’s Main Concerns**

The referees are split, and this is a very close call; after careful consideration, I decided to leave the door open to a revision -- but noting that this is closer to a "reject and resubmit" decision, with a still uncertain outcome. Everybody, myself included, thinks that the paper deals with an important topic, and is in principle potentially a good fit for the journal. A successful revision, however, will need to do the following:

1. **Address the concerns regarding the changing monetary policy regimes** (R1's main comment 1 and R2’s comment 1.h). I think the paper needs to do more on this. Maybe one possible approach is to use higher frequency data for the more recent period and exploit cross-sectional variation across countries to show that the state-dependent effects are present within a specific monetary regime.

2. Better explain where the identification comes from (R1's main comment 1)

3. Show robustness / explain a few puzzling results (R1's Other comments + R2's comment 1.f, 1.i)

4. Address R1's Expositional and Minor comments

5. Change the anchored language (R2's comment 2.d and 3)

6. Try to speak to R2's comment 4

## **Main Comments**

**Referee 2**

1. Data:
   1. **How much do you miss for not having shorter business cycles? Could you resort instead to a shorter time series and use quarterly or monthly data, and expand your panel?**
      1. Honestly, I am not sure I am missing much for not having shorter business cycles because the key message about my paper is focused on the long-run impact of monetary policy. As you can see from the key Figures 4, 5, and 6, I am trying to look at the trade-off in the medium to longer run: 3 to 10 years horizon. In order to do that, the trade-off between having a bigger set of countries but a shorter time horizon is clear – I need a longer time horizon.
      2. Moreover, the novelty of my findings hinges on having a longer time horizon that allows me to explore the wage inflation-unemployment relation in periods other than the post-World War II years which has been extensively researched.
      3. Finally, by using these long run data series, I can reduce the results’ sensitivity to the data vintage and mitigate the large amount of sampling uncertainty coming from different researchers using different data vintages to compute Phillips curves, as Mavroeidis et al. 2014 point out.
   2. **Figure 1 suggests a significant change in the series for wages before and after mid-1900s. It looks like there was a substantial change to the nature of the series since 1970s. This could be a change in wage inflation or simply a change to the methodology. I suspect it's the latter since the variation in mean wages is much larger in periods "where inflation is low".**
      1. I agree. The reason why I think that is the case is due to the data availability. As it is evident from the online data appendix, I only have truly comparable wage series data across countries and over time post World War II where I rely on the IMF International Financial Statistics database as my main source. Before that I need to rely on subsamples of the population for each country or relying on research articles that would measure the average wage across the economy.
      2. It is also true that methodologies change so the way one would compute average wages or unemployment for a subset of the population in the early 1990s is very different from the sophisticated methods that are employed nowadays.
      3. This is one of the many caveats of using this long-run data series, but it is also one caveat I try to mitigate by using the most comparable series available and by chain linking the old data on the new. No extrapolation exercise is done which allows me to be as confident as possible on the final data series.
   3. **Figure 2 again suggest a substantial change in the volatility of the series in the early part of the sample relative to the post-1920's. How much of this is due to a change in the methodology versus a change in the relationships?**
      1. Previous papers (cite whom?) have documented that the Gold Standard period was very stable in terms of economic growth and inflation. So, I think this rather confirms that the historical data is accurately capturing the expected behavior, and thus capturing a change in the relationships.
      2. However, I can’t be 100% sure because there are reasons supporting an artificial change.
      3. A possibility comes from the effect of the two World Wars on the computation of the 10-year rolling window.
      4. One other possibility for the substantial change might also be the change in the sample composition because all countries are equally weighted when data is available. To plot Figure 2, I am relying on a matched sample, that is, only using data from countries with available information on both unemployment and wages. From Table A.1, one can see that 12 out of 18 countries only have unemployment data starting in the 20th century, which might partially explain the puzzling behavior.
      5. Notwithstanding, allow me to mitigate eventual concerns arising from this discussion. The key result of this paper in Figure 5 still goes through if one excludes the Gold Standard period or uses a constant set of countries (Figures ?? and ??) and Tables (?? And ??).
      6. Moreover, Figure ?? plots the same graph but now with a weighted average based on each country’s GDP and Figure ?? plots an estimation of a 20-year rolling window instead. Both help understanding that the puzzle is partially due to using a rolling window and weighting all countries equally. Nevertheless, the negative co-movement between the two variables is always present.
   4. **How do you define periods in which inflation is low? Your first table shows a significant heterogeneity across countries., with wage ranging from -10 to 11 in the first subsample, for example.**
      1. This exercise is presented in subsection 4.2.2. I rewrote the exposition paragraph where I define low price. The periods of low price inflation are different across countries and vary over times as described by the indicator variable I\_{c,t}.
      2. Copy new paragraph here. Low price inflation defined as a dummy variable, which is equal to one for periods when countries experienced lagged price inflation below the threshold of 2% and above -2% (Ii,t = 1 if − 2% < πp i,t−1 < 2%) and equal to 0 when countries experienced high price inflation (Ii,t = 0 if 2% ≤ πp i,t−1 < 40%). Results are robust to different thresholds. The choice of the 2% was motivated by the recent inflation targets of many central banks in the sample and the lower and upper bounds were defined to exclude episodes of extreme inflation or deflation based on the percentile 1 and 99 of my sample.
   5. **How can you argue for anchoring of inflation in some subperiods (as you do in the second paragraph of page 8), when you have such large heterogeneity in inflation rates? Do you have a specific country in mind? It's hard to argue that a country with 11% inflation has any strong anchoring. And anchoring to what?**
      1. A closer look at the Gold Standard period allows me to identify
      2. Anchored to the price of gold during Gold Standard (add citations).
      3. Add a new appendix with a Table rationalizing the anchoring sentence.
   6. **Given this wide range, it would be interesting to report median values as well.**
      1. I added median values to Table 1.
   7. **How are the mean stats calculated? Do you weight by countries' GDP (or other variables)? Is there a country that drives statistics?**
      1. The mean stats are not weighted. I now make that even clearer in the Table notes.
      2. I add a robustness exercise where I produce a weighted version of Table 1. You can find it in Table ??
      3. I also did some scatter plots to investigate whether a specific country is driving the statistics, despite the natural level differences
   8. **The one issue that arises with such a long-time sample is that you include in the same regression several different monetary policy frameworks. Monetary policy during Bretton Woods was quite different from ore recently, for example. This is true for the US and for the remaining of the sample too. Moreover, the 2000s came with significant changes to monetary policy for the bulk of the countries in your sample, with the introduction of the euro. How do these structural breaks affect your findings and how are they controlled for?**
      1. I agree that Monetary policy was different during Bretton Woods compared to the recent years. Moreover, in the first years of the sample some countries did not even had an established and independent central bank. Nevertheless, I am only using countries in a fixed exchange rate regime. I believe that this already mitigates the potential heterogeneous effects of different monetary policies being conducted in different countries because the monetary policy shocks that I am using are not coming from national authorities. Notwithstanding, I acknowledge that these structural breaks present in the base countries like the U.S. and even Germany might affect the response of other countries to these “imported” monetary policy shocks.
      2. Thus, I ran a robustness check in which I included monetary policy regime dummies based on the new Table ?? in Appendix that allocates countries to specific monetary regimes identified based on previous research pieces. Results are robust and presented in Figure ??.
   9. **2020 seems to be included in the sample. This was a year in which we saw massive swings in unemployment, inflation, and other statistics. It may be best to just stop your sample in 2019.**
      1. I am now stopping my sample in 2019, but I did extend all series until 2022 and will make them available
2. **Figure 3 shows results for the Philips multiplier and IRFs for the unemployment rate and wage inflation. How can monetary shocks have effects 10 years from impact?**
   1. The results display cumulative IRFs. I am plotting the typical IRF in graph ??
3. **Go back to point 1e and ask what makes you define one period as "anchored" versus an "unanchored"? Is this preestablished with some statistical test? Is it unanchored in the (recent) common meaning of unanchored inflation expectations? From reading your paper it seems that you simply look at the charts, pick where things correlate one way or another and run a regression to confirm that they correlate one way or another. Instead, your sample split should be based on exogeneous drivers.**
   1. Connect to 1.e).
4. **Finally, and more importantly, how much of the flattening is about low inflation (luck) versus (good) monetary policy?**
   1. I honestly do not think that the econometric tools employed in my paper are suitable to answer this question. Given the nature of the trilemma instrument, I am taking advantage of the fact that economies with fixed exchange rates under perfect capital mobility are unable to implement independent monetary policies.
   2. Thus, a monetary policy surprise in a country with a fixed exchange rate regime is directly link to the base country’s monetary policy conduct that is responding to their own business cycle fluctuations.
   3. So, it is unlikely that countries in a fixed exchange rate regime can conduct “good” or “bad” monetary policy because by definition, they abdicate their monetary policy independence when entering such a regime.

## **Main Comments**

**Referee 1**

1. I don't understand exactly where the identification comes from. You state (p10) that the IV can only be computed when a country's exchange rate is fixed with respect to a baseline country. Does it mean you estimates only pertain to these countries? This would also mean that the variation that you exploit changes as more or less countries join the pool with a fixed exchange rate? This could matter for your results over different samples or conditioning on the average inflation rate. It would be helpful to report (like table 1), the list of countries that you use t different points in time. (Relatedly I didn't find table 1 very useful, but it could be more useful by displaying the information you exploit for identification)
   1. Yes, the IV estimates in the paper only pertain to countries in a fixed exchange rate regime. In my sample, the baseline countries are usually the U.S. or Germany. The reason for the former is that the proposed instrument only works when this is the case. I have now revised the exposition of the instrument’s construction in equation 2 (page ??) to make it clearer.
   2. I understand the concern that the variation changes as more or less countries join the pool with a fixed exchange rate, and this might matter for my results over different subsamples. In order to address this, I start by following the first suggestion and present a new figure where I explicitly show which countries are being used at which point in time when estimating the Panel LP-IV from Equations 4 and 5. Then, I perform a robustness check where I keep the sample composition fixed using only the same set of countries (A,B,C,??) and re-do the main results. These robustness exercises can be found in Figures ??? – and show that ??
   3. I feel that the goal of Table 1 is to present the characteristics of the newly assembled data and hint to the level and stability of inflation measures in different historical periods that is then explored in Section 4 with a more rigorous econometric approach. Nevertheless, I am sympathetic with the comment in brackets and if both referees agree that the value added of Table 1 is minimal, in a further revision, I would be comfortable to put the Table in appendix.
2. Regarding the state dependent effects, how can you tell that the results for state dependence in inflation are not driven by different sample periods? Different sample periods could capture different regimes (hence different Phillips multipliers) and also different average inflation. More generally, I would be a bit more careful. You are only documenting a correlation between the pi-ur trade-off and the average level of inflation. Relatedly, I would get rid of the last sentence in the conclusion. We don't know if this is true.
   1. Revised my conclusion accordingly.

## **Other Comments**

1. Why do you leave out data from the inter-war period? How does that affect the result? What is the rational for removing it?
   1. Merge with question 4 O.C.?
   2. In the previous version, I was only leaving the inter-war period out in one exercise (corresponding to Figure 4). The reason is: ???.
2. I'm not sure I understand the lines in Figures 4 and 5. Is the plain line the linear estimate? Does the error-band correspond to uncertainty around the linear estimate? Is it the weak-IV robust error-band?
3. More importantly, we would need the error-bands for the state-dependent estimates. Preferably weak-IV robust. And then a formal test that the estimates are statistically significantly different at some horizon. It's fine if it's not significant, but this should be reported.
   1. The requested results were reported in Appendix: Tables A.6 and A.7. I acknowledge that the exposition about this was poor, so I heavily rewrote the section to make it clearer that a formal test is being performed. I am now also including the error bands for the state-dependent estimates and removed it from the baseline estimate which is now the dotted line. In my view, this has substantially improved the exposition of the results.
4. Why is there a huge price puzzle in Fig 4 but not in Fig3?
   1. Because these Figures are using different samples. While Figure 3 and 5 use the fully available sample for countries in a fixed exchange rate – now described in Table ??. Figure 4 contrasts the years where inflation was more credibly anchored - the last 20 years (2000-2020) and the Gold Standard epoch (1870-1913) - and compare them against the post-war period (1946-1999) leaving the between-war period (1920-1938) out of this analysis.
5. Maybe you want to normalize the IR of unemployment to peak at one, instead of the IR of the policy rate. This could be clearer to the reader so your focus is on the ratio of the IRs of pi and ur anyway.

## **Expositional Comments**

1. The introduction could be better focused: start directly with what you do: the method LP to get the Phillips multiplier using IV early on and describe the IV and the intuition for the identification. (right now it comes too late and it's not explained. The identification is the core of the paper.
   1. I rewrote the introduction and tried to emphasize the method and the identification strategy used already in the second paragraph, with a focus on the intuition where I explore the trilemma of international finance.
2. The data contribution is minor, just compiling two datasets built by others. No need to overemphasize it.
   1. I rewrote the data contribution part and revised the online data appendix with all the different data sources used in the data compilation.
3. The paper states "this is the first paper to bring a historical perspective on the … inflation—unemployment trade-off". This is a bit exaggerated, the first paper on the topic by Phillips precisely used historical data…
   1. I am sympathetic with this comment. I thus rewrote my contribution to make it seem less presumptuous and squared it with literature addressing the topic in a similar fashion. For example, you can now read “” on page ??
4. Figure 2. I don't understand the series for inflation. Why is there no big increase in inflation in the 70s, early 80s?
   1. There are two explanations for that. First, Figure 2 plots an average across 18 countries that are equally being equally weighted so for example, the high inflation felt in the U.S. is being smoothed by the relatively lower price inflation felt by Switzerland during the same period. Secondly, it is a 20-year rolling window of cpi inflation meaning that the high inflation that some countries felt in the 70s and early 80s are being countered by the relatively lower and less volatile inflation felt in the 50s and 60s.
   2. To show that the correlation is not being driven by this equal weights average, I run an additional exercise where I plot the same cpi inflation 20’year moving average weighting countries according to their GDP size. Results are displayed in Figure ??

## **Minor Comments**

1. **The author uses a number of adjectives that are either not precisely defined, or not needed and distracting. Eg., p1 "carefully" identified shocks, "clean" identification strategy**
   1. (Add this to the previous response on all the rewriting exercise.) I implemented all suggestions from minor comments including a, b, and c.
2. **P12, "central banks have sufficient ability". Don't say sufficient. We don't know what sufficient would be here. Just say "substantial"**
3. **On p3, I suggest writing "according to a standard instead of "according to the NK model".**
   1. Regarding the use of a standard New Keynesian Model, I included in appendix a derivation for a panel framework on a microfounded New Keynesian wage Phillips curve following very closely the work from Galí (2011) and Erceg et al. (2000) on staggered wage contracts.
   2. I also include a calibration exercise emphasizing the relation between the wage stickiness parameter and the slope of the New Keynesian wage Phillips curve.
   3. I understand that the contribution from this exercise is minimal but it is one way to give more structure to my hypothesis and explain what I actually mean when referring to a “standard New Keynesian model”. Nevertheless, if the referees think that this is unnecessary, I can remove this section out of the appendix.

Finally, in an effort to be the change I want to see in the profession, I want to state that I created an open repository in GitHub with the entire workflow from this project including all the codes, data, outputs, documentation, and tex files. This is available to everyone. So, feel free to check it out. The GitHub repository can be found here: <https://github.com/RicardoGabriel/Monetary-Policy-and-the-Wage-Inflation-Unemployment-Tradeoff>

Letter to Editor

* Regarding your first comment: I collected quarterly data for all countries excluding Switzerland (due to data availability on the wage series) from FRED on wages, unemployment, consumer price index as the key variables to estimate the Phillips Multiplier, and also interest rates, GDP, Investment, and Consumption to compute the instrument at the quarterly frequency. Given the lack of available data, I assumed that the capital openness index did not change within the year, so I use the same value for each quarter within the same year from the extended series of the Quinn et al. paper. Only then, I was able to estimate the Phillips multiplier for the sample 1995-2019 and test for eventual state dependencies using the same rule as for the baseline analysis. Results are displayed in Appendix ???
* I am not entirely sure all robustness exercises need to be in the printed version paper, for example comments A, B, C do not display significant changes. So, in case this revision goes through, and you agree with my interpretation of some of the robustness checks, I would suggest either created an online appendix with a set of non-essential robustness checks, or I can just add some footnotes regarding these potential challenges and making the results available upon request.