Econometric issues with Laubach and Williams’ estimates of the natural rate of interest

Daniel Buncic
Stockholm Business School
Stockholm University

First Version: November 22, 2019
This Version (nr.5h): July 28, 2021

Abstract

Holston, Laubach and Williams’ (2017) estimates of the natural rate of interest are driven by the downward trending behaviour of ‘other factor’ $z_t$. I show that their implementation of Stock and Watson’s (1998) Median Unbiased Estimation (MUE) to determine the size of the signal-to-noise parameter $\lambda_z$ which controls the severity of the downward trend in $z_t$ is unsound. It cannot recover the ratio of interest $\lambda_z = a_r \sigma_z / \sigma_{\tilde y}$ from MUE because of a misspecification in Holston et al.’s (2017) Stage 2 model. Moreover, their implementation of MUE on this misspecified Stage 2 model spuriously amplifies the point estimate of $\lambda_z$. Using a simulation experiment, I show that their procedure leads to excessively large estimates of $\lambda_z$ when applied to data generated from a model where the true $\lambda_z$ is zero. Correcting the misspecification in their Stage 2 model and the implementation of MUE results in a substantially smaller (and highly insignificant) $\lambda_z$ point estimate, and thereby a more subdued downward trend in ‘other factor’ $z_t$ and the natural rate. The paper also outlines various other issues with Holston et al.’s (2017) model of the natural rate that make it unsuitable for policy analysis.

Keywords: Natural rate of interest, Median Unbiased Estimation, Kalman Filter, spurious relations, misspecified econometric models.

JEL Classification: C32, E43, E52, O40.

Without implications, I am grateful to Jim Stock, Adrian Pagan, Neil Ericsson, Paolo Giordani, Michael Kiley, Refet Gürkaynak (EEA Session Chair), Fabrizio Venditti, Kirstin Hubrich, Benoit Mojon, Gabriel Perez Quiros, Glenn Rudebusch, Luc Bauwens, Francesco Ravazollo, Simon van Norden, Fabio Canova, Eric Leeper, Lawrence Summers, Georgi Krustev, Bernd Schwaab, Alessandro Galesi, Claus Brand, Wolfgang Lemke, Eric Renault, Robin Lumsdaine, Jean-Paul Renne, Robert Carl Michael Beyer, Volker Wieland, John Paul Broussard, Christopher Gibss, Andrea Papetti, Per Krusell, Ingvar Strid, Xin Zhang, Emanuel Moench, Elmar Mertens, Stéphane Dupraz, Mikkel Plagborg-Meller, as well as seminar and conference participants at Stockholm University, Sveriges Riksbank, the ECB, Deutsche Bundesbank, Norges Bank, Banque de France, NIER/KI, Banca d’Italia, the 2020 EEA Congress, the 2020 SNDE Symposium, eMAF 2020, IAAE 2021, and SED 2021 for comments that helped to improve the paper. Parts of this paper were written while I was visiting the Research Division of the Monetary Policy Department at the Riksbank. I thank Jesper Lindé and my former colleagues at FOE for their hospitality and many stimulating discussions. Lorand Abos provided excellent research assistance.

Corresponding author: Stockholm Business School, Stockholm University, SE-103 37, Stockholm, Sweden. Email: daniel.buncic@sbs.su.se. Web: http://www.danielbuncic.com.

Electronic copy available at: https://ssrn.com/abstract=3897838
1. Introduction

Since the global financial crisis, nominal interest rates have declined substantially to levels last witnessed following the Great Depression. The academic as well as policy literature has attributed this decline in nominal interest rates to a decline in the natural rate of interest; namely, the rate of interest consistent with employment at full capacity and inflation at its target. In this literature, Holston, Laubach and Williams’ (2017) estimates of the natural rate have become particularly influential and are widely regarded as a benchmark. The Federal Reserve Bank of New York (FRBNY) maintains an entire webpage dedicated to providing updates to Holston et al.’s (2017) estimates of the natural rate, not only for the United States (U.S.), but also for the Euro Area, Canada and the United Kingdom (U.K.) (see https://www.newyorkfed.org/research/policy/rstar).

In Holston et al.’s (2017) model, the natural rate of interest is defined as the sum of trend growth of output $g_t$ and ‘other factor’ $z_t$. This ‘other factor’ $z_t$ is meant to capture various underlying structural factors such as savings/investment imbalances, demographic changes, and fiscal imbalances that influence the natural rate, but which are not captured by trend growth $g_t$. In Figure 1 below, I show filtered (as well as smoothed) estimates of Holston et al.’s (2017) ‘other factor’ $z_t$.

![Figure 1: Filtered and smoothed estimates of Holston et al.’s (2017) ‘other factor’ $z_t$.](image)

The dashed lines in Figure 1 show estimates obtained with data ending in 2017:Q1, while the solid lines are estimates based on data extended to 2019:Q2. The strong and persistent downward trending behaviour of ‘other factor’ $z_t$ is striking from Figure 1, particularly from 2012:Q1 onwards. The two (black) dashed vertical lines mark the periods 2012:Q1 and 2015:Q4. In 2015:Q4, the Federal Reserve started the tightening cycle and raised nominal interest rates by 25 basis points. In 2012:Q1, real rates began to rise due to a (mild) deterioration in inflation expectations.² Both led to an increase in the real rate. Yet, Holston

---

¹Holston et al. (2017) do not show a plot of ‘other factor’ $z_t$ on the FRBNY website (as of 22nd of June, 2020).
²See panel (a) of Figure 2, which shows plots of the federal funds rate, the real interest rate, as well as inflation and inflation expectations. This may be model specific, as inflation expectations here are simply modelled as an equally weighted MA(4) of current and past inflation.
et al.’s (2017) estimates of ‘other factor’ $z_t$ declined by about 50 basis points from 2012:Q1 to 2015:Q4, and then another 50 basis points from 2015:Q4 to 2019:Q2, reaching a value of $-1.58$ in 2019:Q2. Because $z_t$ evolves as a driftless random walk in the model, the only parameter that affects the ‘trending behaviour’ of $z_t$ is the ‘signal-to-noise ratio’ parameter $\lambda_z$. The size of $\lambda_z$ thereby has a direct impact on the natural rate estimate.

In this paper, I show that Holston et al.’s (2017) implementation of Stock and Watson’s (1998) Median Unbiased Estimation (MUE) of the ‘signal-to-noise ratio’ parameter is unsound. It cannot recover the ratios of interest $\lambda_g = \sigma_g / \sigma_y$ and $\lambda_z = a_r \sigma_z / \sigma_\tilde{y}$ in Stages 1 and 2 of their three stage procedure. Their implementation of MUE of $\lambda_z$ in Stage 2 is particularly problematic, as Holston et al.’s (2017) procedure is based on a misspecified Stage 2 model. This misspecified Stage 2 model not only fails to identify the ratio of interest $\lambda_z = a_r \sigma_z / \sigma_\tilde{y}$, but further, due to the way Holston et al. (2017) implement the structural break tests in Stage 2 MUE, leads to spuriously large estimates of $\lambda_z$. Since the magnitude of $\lambda_z$ determines the severity of the trend in ‘other factor’ $z_t$, this misspecification is consequential. Correcting the Stage 2 model specification and the structural break test implementation in MUE leads to a substantial quantitative reduction in the point estimate of $\lambda_z$. The resulting filtered (and smoothed) estimates of ‘other factor’ $z_t$ are markedly lower, very close to zero, and highly statistically insignificant. The $p$–values corresponding to the structural break statistics from which $\lambda_z$ is estimated are around 0.5. The strong and persistent downward trend in Holston et al.’s (2017) natural rate estimate driven by ‘other factor’ $z_t$ is thus spurious.

In Section 4.2, I outline in detail the Stage 2 model and the MUE procedure that Holston et al. (2017) implement to estimate $\lambda_z$. I show that their Stage 2 model is misspecified and that their MUE of $\lambda_z$ cannot identify the ratio of interest $a_r \sigma_z / \sigma_\tilde{y}$. Instead, it recovers $\lambda_z = a_r \sigma_z / (\sigma_\tilde{y} + 0.5 a_g \sigma_g)$ if $(a_g + 4a_r) = 0$. If $(a_g + 4a_r) \neq 0$, then additional parameters enter the denominator of $\lambda_z$, making it more intricate to back out $\sigma_z$ from $\lambda_z$, as it will be necessary to make additional assumptions about the time series properties of the nominal interest rate which is not explicitly modelled. The terms $a_r$ and $a_g$ are the parameters on the lagged real interest rate and lagged trend growth in the Stage 2 model of the output gap equation (see Section 4.2 for more details). In the full model, these are restricted so that $a_g = -4a_r$. In their specification of the Stage 2 model, Holston et al. (2017) do not impose this restriction. Moreover, they include only one lag of trend growth $g_t$ in the output gap equation, and further add an intercept term not present in the full structural model (see equation (36c)). Since Stock and Watson’s (1998) MUE relies upon Chow (1960) type structural break tests to estimate $\lambda_g$, these differences in the output gap specification lead to substantially larger $F$ statistics on the break dummy (see Figure 5 for a visual presentation) and therefore estimates of $\lambda_z$. To demonstrate that their misspecified Stage 2 model — when combined with their MUE procedure — leads to spuriously large estimates of $\lambda_z$ when the true value is zero, I implement a simulation experiment in Section 4.2. This simulation experiment shows that the mean estimate of $\lambda_z$ can be as high as 0.028842, with a 45.7% probability (relative frequency) of observing a value larger than estimated from the empirical data, when
computed from artificial data which were simulated from a model where the true $\lambda_z = 0$. These simulation results are concerning, as they suggest that it is Holston et al.’s (2017) MUE procedure that generates the excessively large estimates of $\lambda_z$.

Although Section 4.2 describes the core problem with Holston et al.’s (2017) estimation procedure, there are other issues with the model and how it is estimated. Some of these are outlined in Section 5. For instance, Holston et al.’s (2017) estimates of the natural rate, trend growth, ‘other factor’ $z_t$ and the output gap are extremely sensitive to the starting date of the sample used to estimate the model. Estimating the model with data beginning in 1972:Q1 (or 1967:Q1) leads to negative estimates of the natural rate towards the end of the sample period. These negative estimates are again driven purely by the exaggerated and spurious downward trend in ‘other factor’ $z_t$. The fact that it is possible to generate such negative estimates also for the U.S. by simply adjusting the start of the estimation period highlights the extreme sensitivity of Holston et al.’s (2017) MUE procedure with this model.

In Section 5 I also raise the broader issue of ‘circularity’ between the natural rate estimate as defined in their model and the exogenously determined policy interest rate. Because Kalman Filtered estimates of the natural rate are moving averages of all observed variables that enter the state-space model, any central bank induced change in the policy rate is mechanically transferred to the natural rate via the Kalman Filter recursions of the state vector. This makes it impossible to address ‘causal’ questions regarding the relationship between natural rates and policy rates. Moreover, since ‘other factor’ $z_t$ automatically adjusts to ensure a zero mean output gap equation, Holston et al.’s (2017) natural rate estimate is by default constructed so as to match the movements in the observed policy rate. That is, any persistent deviations will be ‘filled’ with the ‘plug-variable’ $z_t$ to ensure a zero mean in the real rate gap in the output gap equation. In this model, all that needs to be done for a higher natural rate to be realized is to raise the exogenously determined policy rate.

In Holston et al. (2017), ‘other factor’ $z_t$ is defined as a driftless random walk. One reason why the estimate of $\lambda_z$ shrinks towards zero (computed either from MLE or from MUE based on the correct Stage 2 model) could be due to incompatibility of such a specification with the observed data. To understand this, I show in Section 5 that the difference between GDP growth and the real interest rate under Holston et al.’s (2017) model specification is equal to the sum of a stationary ARMA process and ‘other factor’ $z_t$, which is an $I(1)$ process in the model. The ‘integratedness’ of this GDP growth minus real interest rate series is thus dominated by ‘other factor’ $z_t$ and should as such show up as a unit-root in the data. However, the empirical GDP growth minus real interest rate series is only weakly correlated (autocorrelation coefficient < 0.6 and quickly decays towards zero). Formal statistical tests strongly reject the null hypothesis of a unit-root in the series. Shrinking the estimate of $\sigma_z$...
towards zero may thus be the only way to make the \( I(1) \) specification of \( z_t \) in Holston et al.’s (2017) model compatible with the time series properties of the empirical data.

Median Unbiased Estimation is neither well known nor widely used at policy institutions. To give some background on the methodology, and to help understand why Holston et al.’s (2017) implementation of MUE in Stage 2 is unsound, I provide a concise review of the methodology in Section 3. This section will be beneficial to readers unfamiliar with the estimator. It reviews and summarises the conditions when it is likely to encounter ‘pile-up’ at zero problems with Maximum Likelihood Estimation (MLE) of such models. Namely, MLE is likely to generate higher ‘pile-up’ at zero frequencies than MUE when the initial conditions of the state vector are unknown and need to be estimated, and when the true ‘signal-to-noise ratio’ is very small (close to zero). Since Holston et al. (2017) do not estimate the initial conditions of the state vector, but instead use tightly specified prior values, and because their MUEs of the ‘signal-to-noise ratio’ are anything but ‘very small’ in the context of MUE, it seems highly unlikely that MLE should generate higher ‘pile-up’ at zero probabilities than MUE.

For reasons of completeness, I provide a comprehensive description of Holston et al.’s (2017) Stage 1 model and their first stage MUE implementation in Section 4.1. As in the Stage 2 model, I show algebraically that their MUE procedure cannot recover the ratio \( \sigma_g / \sigma_{y^*} \) from \( \lambda_g \) because the error term in the first difference of the constructed trend variable \( y_t^* \) in the first stage model depends on the real interest rate, as well as ‘other factor’ \( z_t \) and trend growth \( g_t \). This means that when the long-run standard deviation from the MUE procedure is constructed, it will not only equal \( \sigma_{y^*} \) as required, but also depend on \( \sigma_z, \sigma_g \), as well as the long-run standard deviation of the real rate. Rewriting a simpler version of the Stage 1 model in local level model form also fails to identify the ratio of interest \( \sigma_g / \sigma_{y^*} \) from the MUE of \( \lambda_g \). This section further illustrates that it is empirically entirely unnecessary to use MUE to estimate \( \sigma_g \) in the first stage model as its MLE does not ‘pile-up’ at zero; not in the local level model, nor in the local linear trend (or unobserved component) model. Estimating \( \sigma_g \) directly by MLE in the second and third stages confirms this result, yielding in fact larger point estimates than implied by the first stage MUE of \( \lambda_g \) obtained from Holston et al.’s (2017) procedure (ie., from \( \sigma_g = \lambda_g^{\text{MUE}} \sigma_{y^*} \)). Readers not interested in the computational intricacies and nuances of the Stage 1 model may skip this section entirely, and only refer back to it as needed for clarification of later results.

MUE of \( \lambda_z \) based on the correctly specified Stage 2 model suggests that there is no role for ‘other factor’ \( z_t \) in this model, given the data.\(^4\) This brings the focus to the estimates of trend growth in this model. Holston et al.’s (2017) estimates give the impression that trend growth has markedly slowed since the global financial crisis, particularly in the immediate aftermath of the crisis. In panels (b) and (c) of Figure 2, I show plots of Holston et al.’s

\(^4\)This result is inline with the MLE based estimates of \( \sigma_z \). Furthermore, these results also carry over to the Euro Area, Canadian and U.K. estimates of \( z_t \) which are not reported here to conserve space, but are documented in detail in Buncic (2020) and are also made available on the author’s webpage at: http://www.danielbuncic.com/data/correct.HLW.factors.zip.
(2017) estimates of $g_t$ together with a few simple and intuitive alternative ones (annualized GDP growth is superimposed in panel (b)). Trend growth is severely underestimated from 2009:Q3 onwards. Robust (median) estimates of average GDP growth over the various business cycle expansion periods shown in Table 1 indicate that trend growth was only approximately 25 basis points lower at 2.25% in the expansion from 2009:Q3 until 2017:Q1 (or 2019:Q2) than over the pre financial crisis expansion from 2002:Q1 to 2007:Q4. \(^5\) Survey based 10 year-ahead expectations of annualized real GDP growth plotted in Figure A.8 and Figure A.9 also suggest that trend growth remained stable (these plots are discussed further in Section 5). The key point to take away from this discussion is that Holston et al.’s (2017) (one sided) Kalman Filter based estimate of $g_t$ is excessively ‘pulled down’ by the large decline in GDP during the financial crisis, and this strongly effects the estimate of trend growth for many periods after the crisis.

The rest of the paper is organised as follows. In Section 2, Holston et al.’s (2017) structural model of the natural rate of interest is described. Section 3 gives a concise background to Stock and Watson’s (1998) Median Unbiased Estimation. In Section 4, I provide a detailed description of the Stage 1 and Stage 2 models, and report the results of the full Stage 3 model estimates. Some additional issues with the model are discussed in Section 5. Section 6 concludes the study. The Appendix provides supplementary information with regards to Holston et al.’s (2017) model, the matrix expansions and derivations of the models in all three stages, and a few selected estimation results from an extended sample period.

2. Holston, Laubach and Williams’ (2017) Model

Holston, Laubach and Williams (2017) use the following ‘structural’ model to estimate the natural rate of interest:\(^6\)

\[
\begin{align*}
\text{Output:} & \quad y_t = y_t^* + \tilde{y}_t \\
\text{Inflation:} & \quad \pi_t = b_\pi \pi_{t-1} + (1 - b_\pi) \pi_{t-2,4} + b_y \tilde{y}_{t-1} + \varepsilon_t^\pi \\
\text{Output gap:} & \quad \tilde{y}_t = a_{y,1} \tilde{y}_{t-1} + a_{y,2} \tilde{y}_{t-2} + \frac{\alpha}{2} \left[ (r_{t-1} - r_{t-1}^*) + (r_{t-2} - r_{t-2}^*) \right] + \varepsilon_t^g \\
\text{Output trend:} & \quad y_t^* = y_{t-1}^* + g_{t-1} + \varepsilon_t^{y^*} \\
\text{Trend growth:} & \quad g_t = g_{t-1} + \varepsilon_t^g \\
\text{Other factor:} & \quad z_t = z_{t-1} + \varepsilon_t^z,
\end{align*}
\]

where $y_t$ is 100 times the (natural) log of real GDP, $y_t^*$ is the permanent or trend component

\(^5\)GDP growth is close to being serially uncorrelated over the last two expansion periods. Also, the volatility of output growth has declined, with the period following 2015:Q4 exhibiting particular low volatility.

\(^6\)In what follows, I use the same notation as in Holston et al. (2017) (see equations 3 to 9 on pages S61 to S63) to facilitate a direct comparison. Also note that this model builds on an earlier specification of Laubach and Williams (2003), where trend growth $g_t$ is scaled by another parameter $c$, and where also a stationary AR(2) process for ‘other factor’ $z_t$ was considered in addition to the $I(1)$ specification in (1f). The estimation procedure with the $I(1)$ specification of $z_t$ in Laubach and Williams (2003) is identical to the one I describe here.
of GDP, \( \tilde{y}_t \) is its cyclical component, \( \pi_t \) is annualized quarter-on-quarter PCE inflation, and 
\[ \pi_{t-2,4} = (\pi_{t-2} + \pi_{t-3} + \pi_{t-4})/3. \] The real interest rate \( r_t \) is computed as:
\[ r_t = i_t - \pi_t^e, \] (2)
where expected inflation is constructed as:
\[ \pi_t^e = (\pi_t + \pi_{t-1} + \pi_{t-2} + \pi_{t-3})/4 \] (3)
and \( i_t \) is the \textit{exogenously} determined nominal interest rate, the federal funds rate.

The natural rate of interest \( r^*_t \) is computed as the sum of trend growth \( g_t \) and \textit{‘other factor’} \( z_t \), both of which are \( I(1) \) processes. The real interest rate gap is defined as \( \tilde{r}_t = (r_t - r^*_t) \).

The error terms \( \varepsilon^\ell_t, \forall \ell = \{\pi, \tilde{y}, y^*, g, z\} \) are assumed to be \textit{i.i.d} normal distributed, mutually uncorrelated, and with time-invariant variances denoted by \( \sigma^2_\ell \). Notice from (1b) that inflation is restricted to follow an integrated AR(4) process. From the description of the data, we can see that the nominal interest rate \( i_t \) as well as inflation \( \pi_t \) are defined in annual or annualized terms, while output, and hence the output gap, trend and trend growth in output are defined at a quarterly rate. Due to this measurement mismatch, Holston \textit{et al.} (2017) adjust the calculation of the natural rate in their code so that trend growth \( g_t \) is scaled by 4 whenever it enters equations that relate it to annualized variables. The natural rate is thus factually computed as \( r^*_t = 4g_t + z_t \).

In the descriptions that follow, I will use the annualized \( 4g_t \) trend growth rate whenever it is important to highlight a result or in some of the algebraic derivations, and will leave the equations in (1) as in Holston \textit{et al.} (2017) otherwise for ease of comparability.

Holston \textit{et al.} (2017) argue that due to \textit{‘pile-up’} at zero problems with Maximum Likelihood (ML) estimation of the variances of the innovation terms \( \varepsilon^\pi_t \) and \( \varepsilon^z_t \) in (1), estimates of \( \sigma^2_g \) and \( \sigma^2_z \) are \“likely to be biased towards zero\“ (page S64). To avoid such \textit{‘pile-up’} at zero problems, they employ Median Unbiased Estimation (MUE) of Stock and Watson (1998) in two preliminary steps — Stage 1 and Stage 2 — to get estimates of what they refer to as \textit{‘signal-to-noise ratios’} defined as \( \lambda^g = \sigma^g / \sigma_y \) and \( \lambda^z = a_* \sigma_z / \sigma_{\tilde{y}} \). In Stage 3, the remaining parameters of the full model in (1) are estimated, conditional on the median unbiased estimates \( \hat{\lambda}^g \) and \( \hat{\lambda}^z \) obtained in Stages 1 and 2, respectively.

Before the three stage procedure of Holston \textit{et al.} (2017) is described, I outline in detail how Stock and Watson’s (1998) median unbiased estimator is implemented, what normalization assumptions it imposes, and how look-up tables for the construction of the estimator are computed. I also include a replication of Stock and Watson’s (1998) empirical estimation of trend growth of U.S. real GDP per capita. Although the section that follows below may seem excessively detailed, long, and perhaps unnecessary, the intention here is to provide

\footnote{This generates some confusion when working with the model, as it is not clear whether the estimated \( z_t \) factor is to be interpreted at an annual or quarterly rate.}
the reader with an overview of how median unbiased estimation is implemented, what it is intended for, and when one can expect to encounter ‘pile-up’ at zero problems to materialize.

3. Stock and Watson’s (1998) Median Unbiased Estimation

Stock and Watson (1998) proposed Median Unbiased Estimation (MUE) in the general setting of Time Varying Parameter (TVP) models. TVP models are commonly specified in a way that allows their parameters to change gradually or smoothly over time. This is achieved by defining the parameters to evolve as driftless random walks (RWs), with the variances of the innovation terms in the RW equations assumed to be small. One issue with Kalman Filter based ML estimation of such models is that estimates of these variances can frequently ‘pile-up’ at zero when the true error variances are ‘very’ small, but nevertheless, non-zero.8

Stock and Watson (1998) show simulation evidence of ‘pile-up’ at zero problems with Kalman Filter based ML estimation in Table 1 on page 353 of their paper. In their simulation set-up, they consider the following data generating process for the series $GY_t$:9

$$GY_t = \beta_t + \varepsilon_t$$

(4a)

$$\beta_t = \beta_{t-1} + (\lambda/T)\eta_t,$$

(4b)

where $\varepsilon_t$ and $\eta_t$ are drawn from i.i.d. standard normal distributions, $\beta_{00}$ is initialized at 0, and the sample size is held fixed at $T = 500$ observations, using 5000 replications. The $\lambda$ values that determine the size of the variance of $\Delta \beta_t$ are generated over a grid from 0 to 30, with unit increments.10 Four median unbiased estimators relying on four different structural break test statistics are compared to two ML estimators. The first ML estimator, referred to as the maximum profile likelihood estimator (MPLE), treats the initial state vector as an unknown parameter to be estimated. The second estimator, the maximum marginal likelihood estimator (MMLE), treats the initial state vector as a Gaussian random variable with a given mean and variance. When the variance of the integrated part of the initial state vector goes to infinity, MMLE produces a likelihood with a diffuse prior.

How one treats the initial condition in the Kalman Filter recursions matters substantially for the ‘pile-up’ at zero problem with MLE. This fact has been known, at least, since the work of Shephard and Harvey (1990).11 The simulation results reported in Table 1 on page 353 in

---

8See the discussion in Section 1 of Stock and Watson (1998) for additional motivation and explanations. As the title of Stock and Watson’s (1998) paper suggests, MUE was introduced for “coefficient variance estimation in TVP models” when this variance is expected to be small.

9See their GAUSS files TESTCDF.GSS and ESTLAM.GSS for details on the data generating process, which are available from Mark Watson’s homepage at http://www.princeton.edu/~mwatson/ddisk/tvpci.zip.

10To be precise, in their GAUSS code, Stock and Watson (1998) use a range from 0 to 80 for $\lambda$, with finer step sizes for lower $\lambda$ values (see, for instance, the file TESTCDF.GSS). That is, $\lambda$ is a sequence between 0 to 30 with increments of 0.25, then 0.5 unit increments from 30 to 60, and unit increments from 60 to 80. In Tables 1 to 3 of their paper, results are reported for $\lambda$ values from 0 up to 30 only, with unit increments.

11On page 340, Shephard and Harvey (1990) write to this: “…we show that the results for the fixed and known
Stock and Watson (1998) show that ‘pile-up’ at zero frequencies are considerably lower when MMLE with a diffuse prior is used than for MPLE, which estimates the initial state vector. For instance, for the smallest considered non-zero population value of \( \lambda = 1 \), which implies a standard deviation of \( \Delta \beta_i \) (\( \sigma_{\Delta \beta} \) henceforth) of \( \lambda/T = 1/500 = 0.002 \), MMLE produces an at most 14 percentage points higher ‘pile-up’ at zero frequency than MUE (ie., 0.60 or 60% for MMLE versus 0.46 or 46% for MUE based on the Quandt (1960) Likelihood Ratio, henceforth QLR, structural break test statistic).\(^{12}\) For MPLE, this frequency is 45 percentage points higher at 0.91 (91%). At \( \lambda = 5 \) (\( \sigma_{\Delta \beta} = 0.01 \) and \( \lambda = 10 \) (\( \sigma_{\Delta \beta} = 0.02 \)), these differences in the ‘pile-up’ at zero frequencies reduce to 11 and 4 percentage points, respectively, for MMLE, but remain still sizeable for MPLE. At \( \lambda = 20 \) (\( \sigma_{\Delta \beta} = 0.04 \)), the ‘pile-up’ at zero problem disappears nearly entirely for MMLE and MUE, with ‘pile-up’ frequencies dropping to 2 and 1 percentage points, respectively, for these two estimators, staying somewhat higher at 7 percentage points for MPLE.

Using MUE instead of MLE to mitigate ‘pile-up’ at zero problems comes, nevertheless, at a cost; that is, a loss in estimator efficiency whenever \( \lambda \) (or \( \sigma_{\Delta \beta} \)) is not very small. From Table 2 on page 353 in Stock and Watson (1998), which shows the asymptotic relative efficiency of MUE (and MPLE) relative to MMLE, it is evident that for true \( \lambda \) values of 10 or greater (\( \sigma_{\Delta \beta} \geq 0.02 \)), the 4 different MUEs yield asymptotic relative efficiencies (AREs) as low as 0.65 (see the results under the \( L \) and MW columns in Table 2). This means that MMLE only needs 65% of MUE’s sample size to achieve the same probability of falling into a given null set. Only for very small values of \( \lambda \leq 4 \) (\( \sigma_{\Delta \beta} \leq 0.008 \)) are the AREs of MUE and MMLE of a similar magnitude, ie., close to 1, suggesting that both estimators achieve approximately the same precision.

Three important points are to be taken away from this review of the simulation results reported in Stock and Watson (1998). First, with MLE, ‘pile-up’ at zero frequencies are substantially smaller when the initial state vector is treated as a known fixed quantity or when a diffuse prior is used. Second, ‘pile-up’ at zero frequencies of MMLE are at most 4 percentage points higher than those of MUE once \( \lambda \geq 10 \) (\( \sigma_{\Delta \beta} = 0.02 \)). Third, MUE can be considerably less efficient than MMLE, in particular for ‘larger’ values of \( \lambda \geq 10 \) (\( \sigma_{\Delta \beta} = 0.02 \)). This suggests that MLE with a diffuse prior should be preferred whenever MUE based estimates of \( \lambda \) (or \( \sigma_{\Delta \beta} \)) are ‘large’ enough to indicate that ‘pile-up’ at zero problems are unlikely to materialize.

To provide the reader with an illustration of how MUE is implemented, and how its estimates compare to the two maximum likelihood based procedures (MPLE and MMLE), I replicate the empirical example in Section 4 of Stock and Watson (1998) which provides esti-

---

\(^{12}\)The four different MUEs based on the different structural break tests appear to perform equally well.
mates of trend growth of U.S. real GDP per capita over the period from 1947:Q2 to 1995:Q4. Note that trend growth in GDP is one of the two components that make up the real natural rate $r^*_t$ in Holston et al. (2017). It is thus beneficial to illustrate the implementation of MUE in this specific context, rather than in the more general setting of time varying parameter models.

### 3.1. Median unbiased estimation of U.S. trend growth

Stock and Watson (1998) use the following specification to model the evolution of annualized trend growth in real per capita GDP for the U.S., denoted by $GY_t$ below:\(^\text{13}\)

\begin{align}
GY_t &= \beta_t + u_t \\
\Delta \beta_t &= (\lambda/T) \eta_t \\
a(L)u_t &= \varepsilon_t,
\end{align}

where $a(L)$ is a (stationary) lag polynomial with all roots outside the unit circle, $\lambda$ is the parameter of interest, $T$ is the sample size, and $\eta_t$ and $\varepsilon_t$ are two uncorrelated disturbance terms, with variances $\sigma_{\eta}^2$ and $\sigma_{\varepsilon}^2$, respectively. The growth rate of per capita GDP is thus composed of a stationary component $u_t$ and a random walk component $\beta_t$ for trend growth. Stock and Watson (1998) set $a(L)$ to a 4th order lag-polynomial, so that $u_t$ follows an AR(4) process. The model in (5) can be recognized as the local level model of Muth (1960), albeit with the generalisation that $u_t$ follows an AR(4) process, rather than white noise. Being in the class of local level models means that the estimate of trend growth will be an exponentially weighted moving average (EWMA) of $GY_t$.\(^\text{14}\)

It is important to highlight here that Stock and Watson’s (1998) discussion of the theoretical results of the estimator in Sections 2.2 – 2.3 of their paper emphasizes that MUE of $\lambda$ in the model in (5) is only possible with the “normalisation $D = 1$”. They write at the top of page 351 (right column): “Henceforth, when $k = 1$, we thus set $D = 1$. When $X_t = 1$, under this normalization, $\lambda$ is $T$ times the ratio of the long-run standard deviation of $\Delta \beta_t$ to the long run standard deviation of $u_t$.\(^\text{15}\)” Denoting the long-run standard deviation of a stochastic process by $\tilde{\sigma}(\cdot)$, this means that

$$\lambda = T \frac{\tilde{\sigma}(\Delta \beta_t)}{\tilde{\sigma}(u_t)} = T \frac{\sigma_{\Delta \beta}}{\sigma_{\varepsilon}/a(1)},$$

\(^\text{13}\)That is, $GY_t = 400\Delta \ln(\text{real per capita GDP})$, where $\Delta$ is the first difference operator (see Section 4 on page 354 in Stock and Watson (1998)). I again follow their notation as closely as possible for comparability reasons.

\(^\text{14}\)Stock and Watson (1998) offer a discussion of the rationale behind the random walk specification of trend growth in $GY_t$ in the second paragraph on the left of page 355. Without wanting to get into a technical discussion, one might want to view the random walk specification of trend growth $\beta_t$ as a purely statistical tool to allow for a slowly changing mean, rather than interpreting trend growth as an $I(1)$ process.

\(^\text{15}\)The parameter $k$ here refers to the column dimension of regressor vector $X_t$. When $k = 1$, then only a model with an intercept is fitted, i.e., $X_t$ contains only a unit constant and no other regressors.
or alternatively, expressed in signal-to-noise ratio form as used by Holston et al. (2017):

\[
\frac{\lambda}{T} = \frac{\hat{\sigma}(\Delta \beta_t)}{\sigma(u_t)} = \frac{\sigma_{\Delta \beta}}{\sigma(\epsilon) a(1)},
\]

(7)

where \(\hat{\sigma}(u_t) = \sigma_{\epsilon}/a(1)\) since \(u_t\) follows a stationary AR(4) process, \(a(1) = (1 - \sum_{i=1}^{4} a_i)\), and \(\hat{\sigma}(\Delta \beta_t) = \sigma_{\Delta \beta}\) due to \(\eta_t\) being i.i.d., yielding further the relation \(\sigma_{\Delta \beta} = (\lambda/T)\sigma_{\eta}\). As a result of the identifying “normalization \(D = 1\)” of MUE, (7) implies that \(\sigma_{\eta} = \sigma_{\epsilon}/a(1)\). That is, the long-run standard deviation of the stationary component \(u_t\) is equal to the standard deviation of the trend growth innovations \(\eta_t\).

Stock and Watson (1998) write on page 354: “Table 3 is a lookup table that permits computing median unbiased estimates, given a value of the test statistic. The normalization used in Table 3 is that \(D = 1\), and users of this lookup table must be sure to impose this normalization when using the resulting estimator of \(\lambda\).” Moreover, the numerical results that are reported in Section 3, which is appropriately labelled “Numerical Results for the univariate Local-Level Model”, are obtained from simulations that employ the local level model of (4) as the data generating process (see Stock and Watson’s (1998) GAUSS programs ESTLAM.GSS, TESTCDF.GSS, and LOOKUP.GSS in the tvpci.zip file that accompanies their paper). These numerical results do not only include the simulations regarding ‘pile-up’ at zero frequencies reported in Table 1, asymptotic power functions plotted in Figure 1, or the AREs provided in Table 2 of Stock and Watson (1998), but also the look-up tables for the construction of the median unbiased estimator of \(\lambda\) in Table 3. It must therefore be kept in mind that these look-up table values are valid only for the univariate local level model, or for models that can be (re-)written in “local level form”.

Table 2 below reports the replication results of Tables 4 and 5 in Stock and Watson (1998).16 Columns one and two in the top half of Table 2 show test statistics and \(p\)-values of the four structural break tests that are considered: i) Nyblom’s (1989) \(L\) test, ii) Andrews and Ploberger’s (1994) mean Wald (MW) test, iii) Andrews and Ploberger’s (1994) exponential Wald (EW) test, and iv) Quandt’s (1960) Likelihood ratio (QLR) test, together with corresponding \(p\)-values.

16All computations are implemented in Matlab, using their GDP growth data provided in the file DYPCE.ASC. Note that I also obtained look-up table values based on a finer grid of \(\lambda\) values from their original GAUSS file LOOKUP.GSS (commenting out the lines if (lamdat[1,1] .<= 30) .and (lamdat[1,1]-floor(lamdat[1,1]) .== 0); in LOOKUP.GSS to list look-up values for the entire grid of \(\lambda\’s\) considered), rather than those listed in Table 3 on page 354 of their paper, where the grid is based on unit increments in \(\lambda\) from 0 to 30. I further changed the settings in the tolerance on the gradient in their maximum likelihood (maxlik) library routine to maxGradTol = 1e-08 and used the printing option format /rd 14,14 for a more precise printing of all results up to 14 decimal points. Lastly, there is a small error in the construction of the lag matrix in the estimation of the AR(4) model in file TST_GDP1.GSS (see lines 40 to 47). The first column in the \(w\) matrix is the first lag of the demeaned per capita trend growth series, while columns 2 to 4 are the second to fourth lags of the raw, that is, not demeaned per capita trend growth series. Correcting this leads to mildly higher, yet still insignificant, point estimates of all \(\sigma_{\Delta \beta}\). For instance, the point estimate of \(\sigma_{\Delta \beta}\) based on Nyblom’s (1989) \(L\) statistic yields 0.1501, rather than 0.1303, but remains still statistically insignificant, with the lower value of the confidence interval being 0. To exactly replicate the results in Stock and Watson’s (1998), I compute the lag matrix as they do.
As a reminder, the MW, EW and QLR tests are Chow (1960) type structural break tests, which test for a structural break in the unconditional mean of a series at a given or known point in time. Chow (1960) break tests require a partitioning of the data into two sub-periods. When the break date is unknown, these tests are implemented by rolling through the sample. To be more concrete, denote by \( Y_t \) the series to be tested for a structural break in the unconditional mean. Let the dummy variable \( D_t(\tau) = 1 \) if \( t > \tau \), and 0 otherwise, where \( \tau = \{ \tau_0, \tau_0 + 1, \tau_0 + 2, \ldots, \tau_1 \} \) is an index (or sequence) of grid points between endpoints \( \tau_0 \) and \( \tau_1 \). As is common in this literature, Stock and Watson (1998) set these endpoints at the 15\(^{th}\) and 85\(^{th}\) percentiles of the sample size \( T \), that is, \( \tau_0 = 0.15T \) and \( \tau_1 = 0.85T \).\(^{17}\) For each \( \tau \in [\tau_0, \tau_1] \), the following regression of \( Y_t \) on a constant and dummy \( D_t(\tau) \) is estimated:

\[
Y_t = \zeta_0 + \zeta_1 D_t(\tau) + \epsilon_t,
\]

and the \( F \) statistic (the square of the \( t \)-statistic) on the point estimate \( \hat{\zeta}_1 \) is constructed. The sequence \( \{ F(\tau) \} \big|_{\tau=\tau_0}^{\tau_1} \) of \( F \) statistics is then used to compute the MW, EW and QLR structural break test statistics needed in the implementation of MUE. These are calculated as:

\[
\text{MW} = \frac{1}{N_\tau} \sum_{\tau=\tau_0}^{\tau_1} F(\tau) \tag{9a}
\]

\[
\text{EW} = \ln \left( \frac{1}{N_\tau} \sum_{\tau=\tau_0}^{\tau_1} \exp \left\{ \frac{1}{2} F(\tau) \right\} \right) \tag{9b}
\]

\[
\text{QLR} = \max_{\tau \in [\tau_0, \tau_1]} \{ F(\tau) \} \big|_{\tau=\tau_0}^{\tau_1}, \tag{9c}
\]

where \( N_\tau \) denotes the number of grid points in \( \tau \). Nyblom’s (1989) \( L \) test statistic is computed without sequentially partitioning the data via the sum of squared cumulative sums of \( Y_t \). More specifically, let \( \hat{\mu}_Y \) denote the sample mean of \( \hat{Y}_t \), \( \hat{\sigma}_Y^2 \) the sample variance of \( Y_t \), and \( \tilde{Y}_t = Y_t - \hat{\mu}_Y \) the demeaned \( Y_t \) process. Nyblom’s (1989) \( L \) statistic is then constructed as:

\[
L = T^{-1} \sum_{t=1}^{T} \frac{\vartheta_t^2}{\hat{\sigma}_Y^2}, \tag{10}
\]

where \( \vartheta_t \) is the scaled cumulative sum of \( \tilde{Y}_t \), i.e., \( \vartheta_t = T^{-1/2} \sum_{s=1}^{t} \tilde{Y}_s \).

Median unbiased estimates of \( \lambda \) based on Stock and Watson’s (1998) look-up tables are

\(^{17}\)To be precise, \( \tau_0 \) is computed as \( 100 \times 0.15 \times T \) and \( \tau_1 \) as \( T - \tau_0 \). Also, it is standard practice in the structural break literature to trim out some upper/lower percentiles of the search variable to avoid having too few observations at the beginning or at the end of the sample in the 0 and 1 dummy regimes created by \( D_t(\tau) \). In fact, the large sample approximation of the distribution of the QLR test statistic depends on \( \tau_0 \) and \( \tau_1 \). Stock and Watson (2011) write to this on page 558: “For the large-sample approximation to the distribution of the QLR statistic to be a good one, the sub-sample endpoints, \( \tau_0 \) and \( \tau_1 \), cannot be too close to the beginning or the end of the sample.” Employing endpoints other than the 15\(^{th}\) upper/lower percentile values used by Stock and Watson (1998) in the simulation of the look-up table for \( \lambda \) is thus likely to affect the values provided in Table 3 of Stock and Watson (1998), due to the endpoints’ influence on the distribution of the structural break test statistics.
reported in column 3 of Table 2, followed by respective 90% confidence intervals (CIs) in square brackets. The last two columns show estimates of $\sigma_{\Delta \beta}$ computed as $\hat{\sigma}_{\Delta \beta} = \hat{\lambda}/T \times \hat{\sigma}_{\varepsilon}/\hat{\sigma}(1)$, with 90% CIs also in square brackets. In the bottom half of Table 2, MLE and MUE based parameter estimates of the model in (5) are reported. The columns under the MPLE and MMLE headings show, respectively, MLE based results when the initial state vector is estimated and when a diffuse prior is used. The diffuse prior for the $I(1)$ element of the state vector is centered at 0 with a variance of $10^6$. The next two columns under the headings MUE(0.13) and MUE(0.62) report parameter estimates of the model in (5) with $\sigma_{\Delta \beta}$ held fixed at its MUE point estimate of 0.13 and upper 90% CI value of 0.62, respectively. The last column under the heading SW.GAUSS lists the corresponding MUE(0.13) estimates obtained from running Stock and Watson’s (1998) GAUSS code as reference values.  

As can be seen from the results in Table 2, consistent with the ‘pile-up’ at zero problem documented in the simulations in Stock and Watson (1998) (and also Shephard and Harvey (1990)), the MPLE estimate of $\sigma_{\Delta \beta}$ goes numerically to zero (up to 11 decimal points), while MMLE produces a ‘sizeable’ point estimate of $\sigma_{\Delta \beta}$ of 0.044. Although Stock and Watson (1998) (and also I) do not report a standard error for $\hat{\sigma}_{\Delta \beta}$ in the tables containing the estimation results, the estimate of stderr($\hat{\sigma}_{\Delta \beta}$) is 0.1520, suggesting that $\hat{\sigma}_{\Delta \beta}$ is very imprecisely estimated. From the MUE results reported in the first column of the top half of Table 2 it is evident that all 4 structural break tests yield confidence intervals for $\lambda$ and hence also $\sigma_{\Delta \beta}$ that include zero. Thus, even when using MUE as the ‘preferred’ estimator, one would conclude that $\hat{\lambda}$ and $\hat{\sigma}_{\Delta \beta}$ are not statistically different from zero.  

An evident practical problem with the use of Stock and Watson’s (1998) MUE is that the 4 different structural break tests can produce vastly different point estimates of $\lambda$. This is clearly visible from Table 2, where the 4 tests yield $\lambda$ estimates with an implied $\hat{\sigma}_{\Delta \beta}$ range between 0.0250 (for QLR) and 0.1303 (for $L$). From the simulation results in Stock and Watson (1998) we know that all 4 tests seem to behave equally well in the ‘pile-up’ at zero frequency simulations (see Table 1 in Stock and Watson (1998)). However, the QLR test performed ‘best’ in the efficiency results, producing the largest (closer to 1) asymptotic relative efficiencies in Table 2 of Stock and Watson (1998). Analysing these results in the context of the empirical estimation of trend growth, the most accurate MUE estimator based on the QLR structural break test produces an estimate of $\sigma_{\Delta \beta}$ that is 5 times smaller than the largest one based on the $L$ structural break test, with the MMLE estimate of $\sigma_{\Delta \beta}$ being approximately double the size of the QLR estimate.

---

18 See the results reported in Table 5 on page 354 in Stock and Watson (1998), where nevertheless only two decimal points are reported. MPLE and MMLE are also replicated accurately to 6 decimal points.

19 Stock and Watson (1998) compute standard errors for the remaining MMLE parameters (see column three in the upper part of Table 5 on page 354 in their paper. They write in the notes to Table 5: “Because of the nonnormal distribution of the MLE of $\lambda$, the standard error for $\sigma_{\Delta \beta}$ is not reported.” Evidently, ‘testing’ the null hypothesis of $\sigma_{\Delta \beta} = 0$ using a standard $t$—ratio does not make any sense statistically. Nevertheless, $\sigma_{\Delta \beta}$ is very imprecisely estimated, and highly likely to be ‘very’ close to zero. The MMLE log-likelihood function with the restriction $\sigma_{\Delta \beta} = 0$ is $-547.5781$, while the (unrestricted) MMLE is $-547.4805$, with the difference between the two being very small of about 0.10.
To provide a visual feel of how different the MLE and MUE based estimates of U.S. trend growth are, I show plots of the smoothed estimates in Figure 3 (these correspond to Figures 4 and 3 in Stock and Watson (1998)). The top panel displays the MPLE, MMLE, MUE(0.13), and MUE(0.62) estimates together with a 90% CI of the MMLE estimate (shaded area), as well as a dashed yellow line that shows Stock and Watson’s (1998) GAUSS code based MUE(0.13) estimate for reference. The plot in the bottom panel of Figure 3 superimposes the actual $GY_t$ series to portray the variability in the trend growth estimates relative to the variation in the data from which these were extracted.\(^{20}\) The $y-$axis range is set as in Figures 4 and 3 in Stock and Watson (1998). As can be seen from Figure 3, there is only little variability in the MLE based trend growth estimates, with somewhat more variation from MUE(0.13). Nonetheless, all three trend growth point estimates stay within the 90% error bands of MMLE. Moreover, the plots in Figure 3 confirm the lack of precision of MUE. Trend growth could be anywhere between a constant value of about 1.8% ($\hat{\beta}_0$ from MPLE), which is a flat line graphically when $\sigma_{\Delta \beta}$ is held fixed at its lower 90% CI value of 0, and a rather volatile series which produces a range between nearly 4.5% in 1950 and less than 0.5% in 1980 when $\sigma_{\Delta \beta}$ is set at its upper 90% CI value of 0.62.

Given the previous results and discussion, one could argue that the statistical evidence in support of any important changes to trend growth in real U.S. GDP per capita is rather weak in this data set. As a robustness check and in the context of a broader replication of the time varying trend growth estimates of Stock and Watson (1998), I re-estimate the model using a more recent vintage of real GDP per capita data from the Federal Reserve Economic Data (FRED2) database. These results are reported in Table 3, which is arranged in the same way as Table 2 (only the last column with heading SW.GAUSS is removed). The sample period is again from 1947:Q2 to 1995:Q4, using an AR(4) model to approximate $u_t$ in (5a).\(^{21}\) From Table 3 it is clear that not only do the two MLE based estimates of $\sigma_{\Delta \beta}$ yield point estimates that are numerically equal to zero, but so do all 4 MUEs. Hence, trend growth may well be constant. More importantly, it demonstrates that MUE can also lead to (exactly) zero estimates of $\sigma_{\Delta \beta}$.\(^{22}\)

Before I proceed to describe how the three stage procedure of Holston et al. (2017) is implemented, a brief procedural description of Stock and Watson’s (1998) MUE that lists the main steps needed to replicate the results reported in Table 2 and Figure 3 is given below.

\footnotesize{(i) Fit an AR(4) model to $GY_{t}$, construct $\hat{\lambda}(L)$ from the estimated AR(4) coefficients $\{\hat{\alpha}_j\}_{j=1}^4$.}

\footnotesize{\(^{20}\)Notice from the top panel of Table 2 that there are four different estimates of $\lambda$, and thus four $\hat{\sigma}_{\Delta \beta}$. Rather then showing smoothed trend estimates for all four of these, I follow Stock and Watson (1998) and only show estimates based on Nyblom’s (1989) $L$ statistic, which has the largest $\lambda$ estimate, and hence also $\hat{\sigma}_{\Delta \beta}$.\(^{21}\)The results using an ARMA(2, 2) model for $u_t$ instead are qualitatively the same.\(^{22}\)I show later that the Stage 2 MUE procedure of Holston et al. (2017) is incorrectly implemented and based on a misspecified Stage 2 model. Once this is corrected, the Stage 2 $\lambda_t$ that one obtains is very close to zero, resulting in the full model MLE and MUE estimates of the natural rate being very similar. Although I do not show this here to avoid a repetition of the results, the similarity between the MLE and correct Stage 2 MUE based estimates holds also for the remaining three data sets analysed by Holston et al. (2017); namely, for the Euro Area, the U.K. and Canada. These results are documented in Buncic (2020).}
and filter the series to remove the AR(4) serial dependence. Let \( \tilde{Y}_t = \hat{a}(L)G_Y t \) denote the AR(4) filtered series. Use the residuals \( \hat{\varepsilon}_t \) from the fitted AR(4) model for \( G_Y t \) to compute an estimate of the standard deviation of \( \varepsilon_t \) and denote it by \( \hat{\sigma}_\varepsilon \). Also, let \( \hat{a}(1) = (1 - \sum_{j=1}^{4} \hat{a}_j) \).

(ii) Test for a structural break in the unconditional mean of the AR(4) filtered series \( \tilde{G}_Y t \) using the four structural break tests described above. That is, replace \( Y_t \) in (8) with \( \tilde{G}_Y t \), run the dummy variable regression in (8), and compute the structural break statistics as defined in (9) and (10).

(iii) Given these structural break test statistics, use the look-up values provided in Table 3 on page 354 in Stock and Watson (1998) to find the corresponding \( \lambda \) value by interpolation. Once an estimate of \( \lambda \) is available, compute \( \hat{\sigma}_{\Delta \beta} = T^{-1} \hat{\lambda} \hat{\sigma}_\varepsilon / \hat{a}(1) \), where \( \hat{\sigma}_\varepsilon \) and \( \hat{a}(1) \) are obtained from Step (i).

(iv) With \( \sigma_{\Delta \beta} \) held fixed at its median unbiased estimate obtained in Step (iii), estimate the remaining parameters of the model in (5) using the Kalman Filter and MLE, namely, MPLE, where the initial value is estimated as well.

Finally, using the estimates of the full set of parameters of the model in (5), apply the Kalman Smoother to extract an estimate of annualized trend growth of U.S. real per capita GDP.

4. Three stage estimation procedure of Holston et al. (2017)

Holston et al. (2017) employ MUE in two preliminary stages that are based on restricted versions of the full model in (1) to obtain estimates of the ‘signal-to-noise ratios’ \( \lambda_g = \sigma_g / \sigma_{\gamma} \) and \( \lambda_z = a_r \sigma_z / \sigma_{\gamma} \). These ratios are then held fixed in Stage 3 of their procedure, which produces estimates of the remaining parameters of the model in (1). In order to conserve space in the main text, I provide all algebraic details needed for the replication of the three individual stages in the Appendix, which includes also some additional discussion as well as R-Code extracts to show the exact computations. In the results that are reported in this section, I have used their R-Code from the file HLW_Code.zip made available on Williams’ website at the New York Fed to (numerically) accurately reproduce their results. The sample period

\(^{23}\)This is the generalized least squares (GLS) step in the original TVP model description on page 350 in Stock and Watson (1998).

\(^{24}\)Note here that Stock and Watson (1998) fix the value of \( \sigma_{\Delta \beta} \) at \( T^{-1} \hat{\lambda} \hat{\sigma}_\varepsilon / \hat{a}(1) \), where the hats denote that these are estimates from Steps (i) to (iii), and do not re-estimate \( \sigma_\varepsilon \) or the AR(4) \( \{a_j\}_{j=1}^{4} \) parameters in the final estimation of the full model in (1). That is, in the GAUSS files LMAIRC.PRC and LMAIR1.PRC, the SERW variable in the q matrix is kept at the OLS based estimates, i.e., not making the SERW variable a function of \( \sigma_\varepsilon \) and \( \{a_j\}_{j=1}^{4} \) (see line 19, with corresponding code: \( q[1,1]=\text{serw}^2 \)). In Holston et al. (2017), only \( \hat{\lambda} / T \) is held fixed in the full Stage 3 model, with the other parameters that make up the long-run standard deviation re-estimated. One could also implement this in Stock and Watson’s (1998) model, however, the goal here is purely to illustrate the computations in an empirical application, as opposed to providing new or different estimates.

\(^{25}\)Williams’ website at the Federal Reserve Bank of New York is at: https://www.newyorkfed.org/research/economists/williams/pub. Their code is available from the website: https://www.newyorkfed.org/media-
that I cover ends in 2017:Q1. The beginning of the sample is the same as in Holston et al. (2017). That is, it starts in 1960:Q1, where the first 4 quarters are used for initialisation of the state vector, while the estimation period starts in 1961:Q1.

Holston et al. (2017) adopt the general state-space model (SSM) notation of Hamilton (1994) in their three stage procedure. The SSM is formulated as follows:

\[
y_t = Ax_t + H\xi_t + \nu_t \quad \text{where} \quad \begin{bmatrix} \nu_t \\ \xi_t \end{bmatrix} \sim \text{MNorm} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & W \end{bmatrix} \right),
\]

where I define \( \epsilon_t = SE_t \), so that \( \text{Var}(\epsilon_t) = \text{Var}(SE_t) = SWS' = Q \) to make it consistent with the notation used in Holston et al. (2017). The (observed) measurement vector is denoted by \( y_t \) in (11), \( x_t \) is a vector of exogenous variables, \( A, H \) and \( F \) are conformable system matrices, \( \xi_t \) is the latent state vector, \( S \) is a selection matrix, and the notation \( \text{MNorm}(\mu, \Sigma) \)denotes a multivariate normal random variable with mean vector \( \mu \) and covariance matrix \( \Sigma \). The disturbance terms \( \nu_t \) and \( \epsilon_t \) are serially uncorrelated, and the (individual) covariance matrices \( R \) and \( W \) are assumed to be diagonal matrices, implying zero correlation between the elements of the measurement and state vector disturbance terms. The measurement vector \( y_t \) in (11) is the same for all three stages and is defined as \( y_t = [y_t, \pi_t]' \), where \( y_t \) and \( \pi_t \) are the log of real GDP and annualized PCE inflation, respectively, as defined in Section 2. The exact form of the remaining components of the SSM in (11) changes with the estimation stage that is considered, and is described in detail either in the text below or in the Appendix.

As I have emphasized in the description of MUE in Section 3, the simulation results of Stock and Watson (1998) show that ‘pile-up’ at zero frequencies for MLE are not only a function of the size of the variance of \( \Delta\beta_t = (\lambda/T)\eta_t \) (or alternatively \( \lambda \)), but also depend critically on whether the initial condition of the state vector is estimated or not. Now Holston et al. (2017) do not estimate the initial condition of the state vector in any of the three stages that are implemented. Instead, they apply the HP filter to log GDP data with the smoothing parameter set to 36000 to get a preliminary estimate of \( y_t^* \) and trend growth \( g_t \) (computed as the first difference of the HP filter estimate of \( y_t^* \)) using data from 1960:Q1 onwards. ‘Other factor’ \( z_t \) is initialized at 0. This means that \( \xi_{00} \) has known and fixed quantities in all three

\[ y_{00} = \begin{bmatrix} y_{01}^*, y_{01}^*, g_{-1}, g_{-2}, z_{-1}, z_{-2} \end{bmatrix}' \]
stages. Given the simulation evidence provided in Table 1 on page 353 in Stock and Watson (1998), one may thus expect a priori ‘pile-up’ at zero frequencies of MLE (without estimation of the initial conditions) to be only marginally higher than those of MUE, especially for ‘larger’ values of $\lambda$.

Also, Holston et al. (2017) determined the covariance matrix of the initial state vector in an unorthodox way. Even though every element of the state vector $\xi_t$ in all three estimation stages is an $I(1)$ variable, they do not use a diffuse prior on the state vector. Instead, the covariance matrix is determined with a call to the function `calculate.covariance.R` (see the code snippet in R-Code 2 for details on this function, and also lines 66, 84, and 88, respectively, in their R-files `rstar.stage1.R`, `rstar.stage2.R`, and `rstar.stage3.R`, with line 88 in `rstar.stage3.R` also shown on the second page of the code snippet in R-Code 1). To summarize what this function does, consider the Stage 1 model, which is estimated with a call to `rstar.stage1.R`. The function `calculate.covariance.R` first sets the initial covariance matrix to $0.2 \times I_3$ and the estimate $\hat{\sigma}_{\gamma}^2$ from the initial run of the Stage 1 model. What this procedure effectively does is to set $P_{00}$ to the first time period’s predicted state covariance matrix, given an initial state covariance of $0.2 \times I_3$ and the estimate $\hat{\sigma}_{\gamma}^2$ was obtained by MLE and the Kalman Filter using $0.2 \times I_3$ as the initial state covariance. This way of initialising $P_{00}$ is rather circular, as it fundamentally presets $P_{00}$ at $0.2 \times I_3$.

$$P_{00} = F \text{diag}([0.2, 0.2, 0.2]) F' + \hat{Q}$$

(12a)

$$= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}' + \begin{bmatrix} \hat{\sigma}_{\gamma}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(12b)

$$= \begin{bmatrix} 0.4711 & 0.2 & 0.0 \\ 0.2 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{bmatrix}$$

(12c)

with $\hat{Q}$ a $(3 \times 3)$ dimensional zero matrix with element $(1, 1)$ set to $\hat{\sigma}_{\gamma}^2 = 0.27113455739$ from the initial run of the Stage 1 model. What this procedure effectively does is to set $P_{00}$ to the first time period’s predicted state covariance matrix, given an initial state covariance matrix of $0.2 \times I_3$ and the estimate $\hat{\sigma}_{\gamma}^2$, where $\hat{\sigma}_{\gamma}^2$ was obtained by MLE and the Kalman Filter using $0.2 \times I_3$ as the initial state covariance. This way of initialising $P_{00}$ is rather circular, as it fundamentally presets $P_{00}$ at $0.2 \times I_3$.\(^{28}\)

When the state vector contains $I(1)$ variables, it is not only standard practice to use a
diffuse prior, but it is highly recommended. For instance, Harvey (1989) writes to this on the bottom of page 121: “When the transition equation is non-stationary, the unconditional distribution of the state vector is not defined. Unless genuine prior information is available, therefore, the initial distribution of $\alpha_0$ must be specified in terms of a diffuse or non-informative prior.” (emphasis added, $\alpha_t$ is the state-vector in Harvey’s notation). It is not clear why Holston et al. (2017) do not use a diffuse prior. However, one may conjecture that it could be due to their preference for reporting Kalman Filtered (one-sided) rather than the more efficient Kalman Smoothed (two-sided) estimates of the latent state vector $\xi_t$ which includes trend growth $g_t$ and ‘other factor’ $z_t$ needed to construct $r^*_t$.  

As a final point in relation to the probability of ‘pile-up’ at zero problems arising due to small variances of the state innovations, and hence the rationale for employing MUE rather than MLE in the first place, one can observe from the size of the $\sigma_g$ and $\sigma_z$ estimates for the U.S. reported in Table 1 on page S60 in Holston et al. (2017) that these are rather ‘large’ at 0.122 and 0.150, respectively. The simulation results in Table 1 in Stock and Watson (1998) show that ‘pile-up’ at zero frequencies drop to 0.01 for both, MMLE and MUE, when the true population value of $\lambda$ is 30 ($\sigma_{\Delta \beta} = 0.06$). Given the fact that Holston et al. (2017) do not estimate the initial value of the state vector, and that their median unbiased estimates are about two times larger than 0.06, it seems highly implausible that ‘pile-up’ at zero problems should materialize with a higher frequency for MLE than for MUE.

### 4.1. Stage 1 model

Holston et al.’s (2017) first stage model takes the following restricted form of the full model presented in equation (1):  

\begin{equation}
  y_t = y^*_t + \hat{y}_t \tag{13a}
\end{equation}

\begin{equation}
  \pi_t = b_{\pi} \pi_{t-1} + (1 - b_{\pi}) \pi_{t-2,4} + b_y \hat{y}_{t-1} + \epsilon_t \pi_t \tag{13b}
\end{equation}

In an earlier paper using a similar model for the NAIRU, Laubach (2001) discusses the use of diffuse priors. Laubach (2001) writes on page 222: “The most commonly used approach in the presence of a nonstationary state is to integrate the initial value out of the likelihood by specifying an (approximately) diffuse prior.” He then proceeds to describe an alternative procedure that can be implemented by using: “a few initial observations to estimate the initial state by GLS, and use the covariance matrix of the estimator as initial value for the conditional covariance matrix of the state.” The discussion is then closed with the statement: “This is the first approach considered here. Because this estimate of the initial state and its covariance matrix are functions of the model parameters, under certain parameter choices the covariance matrix may be ill conditioned. The routines then choose the diffuse prior described above as default.” Thus even here, the diffuse prior is the “safe” default option. Note that their current procedure does not use: “a few initial observations to estimate the initial state”, but the same sample of data that are used in the final model, i.e., with data beginning in 1961:Q1.  

Note that Filtered estimates of $g_t$, $z_t$ and thus also $r^*_t$ are very volatile at the beginning of the sample period (until about 1970) when $P_{00}$ is initialized with a diffuse prior.  

See Section A.1 in the Appendix for the exact matrix expressions and expansions of the first stage SSM. Note that one key difference of Holston et al.’s (2017) SSM specification described in equations (A.3) and (A.4) in the Appendix is that the expansion of the system matrices for the Stage 1 model does not include the drift term $g$ in the trend specification in (13d), so that $y^*_t$ follows a random walk without drift. Evidently, such a specification cannot match the upward trend in the GDP data. To resolve this mismatch, Holston et al. (2017) ‘detrend’ output $y_t$ in the estimation (see Section A.1 in the Appendix which describes how this is done and also shows snippets of their R-Code).
\[ y_t = a_{y,1}y_{t-1} + a_{y,2}y_{t-2} + \varepsilon_t^y \]  
\[ y_t^* = g + y_{t-1}^* + \tilde{\varepsilon}_t^y, \]  
(13c)  
(13d)

where the vector of Stage 1 parameters to be estimated is:

\[ \Theta_1 = [a_{y,1}, a_{y,2}, b_{\pi}, b_y, g, \sigma_g, \sigma_{\pi}, \sigma_{y^*}]'. \]  
(14)

To be able to distinguish the disturbance terms of the full model in (1) from the ones in the restricted Stage 1 model in (13) above, I have placed a ring (˚) symbol on the error terms in (13c) and (13d). These two disturbance terms from the restricted model are defined as:

\[ \tilde{\varepsilon}_t^y = g_{t-1} - g + \varepsilon_t^y \]  
(15)

and

\[ \tilde{\varepsilon}_t^y = \frac{a_r}{2} [ (r_{t-1} - 4g_{t-1} - z_{t-1}) + (r_{t-2} - 4g_{t-2} - z_{t-2}) ] + \varepsilon_t^y. \]  
(16)

From (15) and (16) it is clear that, due to the restrictions in the Stage 1 model, the error terms \( \tilde{\varepsilon}_t^y \) and \( \tilde{\varepsilon}_t^y^* \) in (13) will not be uncorrelated anymore, since \( \text{Cov}(\tilde{\varepsilon}_t^y, \tilde{\varepsilon}_t^y^*) = -\frac{a_r}{2} 4\sigma_g^2 \) given the assumptions of the full model in (1). The separation of trend and cycle shocks in this formulation of the Stage 1 model is thus more intricate, as both shocks will respond to one common factor; the missing \( g_{t-1} \).

In the implementation of the Stage 1 model, Holston et al. (2017) make two important modelling choices that have a substantial impact on the \( \Theta_1 \) parameter estimates, and thus also the estimate of the ‘signal-to-noise ratio’ \( \lambda_g \) used in the later stages. The first is the tight specification of the prior variance of the initial state vector \( P_{00} \) discussed in the introduction of this section. The second is a lower bound restriction on \( b_y \) in the inflation equation in (13b) \( b_y \geq 0.025 \) in the estimation). The effect of these two choices on the estimates of the Stage 1 model parameters are shown in Table 4 below. The left block of the estimates in Table 4 (under the heading ‘HLW Prior’) reports four sets of results where the state vector was initialized using their values for \( \xi_{00} \) and \( P_{00} \). The first column of this block (HLW.R-File) reports estimates from running Holston et al.’s (2017) R-Code for the first stage model. These are reported as reference values. The second column \( (b_y \geq 0.025) \) shows my replication of Holston et al.’s (2017) results using the same initial values for parameter vector \( \Theta_1 \) in the optimisation routine and also the same lower bound constraint on \( b_y \). The third column \( (\text{Alt.Init.Vals}) \) displays the results I obtain when a different initial value for \( b_y \) is used, with the lower bound restriction \( b_y \geq 0.025 \) still in place. The fourth column \( (b_y \text{ Free}) \) reports results when the lower bound constraint on \( b_y \) is removed.32 The right block in Table 4

32 To find the initial values for \( \Theta_1 \), Holston et al. (2017) apply the HP filter to GDP to obtain an initial estimate of the cycle and trend components of GDP. These estimates are then used to find initial values for (some of) the components of parameter vector \( \Theta_1 \) by running OLS regressions of the HP cycle estimate on two of its own lags (an AR(2) essentially), and by running regressions of inflation on its own lags and one lag of the HP cycle. Interestingly, although readily available, rather than taking the coefficient on the lagged value of the HP cycle...
shows parameter estimates when a diffuse prior for $\xi_t$ is used, where $P_{00}$ is set to $10^6$ times a three dimensional identity matrix, with the left and right columns showing, respectively, the estimates with and without the lower bound restriction on $b_y$ imposed.

Notice initially from the first two columns in the left block of Table 4 that their numerical results are accurately replicated up to 6 decimal points. From these results we also see that the lower bound restriction on $b_y$ is binding. Holston et al. (2017) set the initial value for $b_y$ at 0.025, and there is no movement away from this value in the numerical routine. Specifying an alternative initial value for $b_y$, which is determined in the same way as for the remaining parameters in $\theta_1$, leads to markedly different estimates, while removing the lower bound restriction on $b_y$ all together results in the ML estimate of $b_y$ to converge to zero. Evidently, these three scenarios yield also noticeably different values for $\hat{\sigma}_{y^*}$, that is, values between 0.4190 and 0.6177. The diffuse prior based results (with and without the lower bound restriction) in the right block of Table 4 show somewhat less variability in $\hat{\sigma}_{y^*}$, but affect the persistence of the cycle variable $\tilde{y}_t$ in the model, with the smallest AR(2) lag polynomial root being 1.1190 when $b_y \geq 0.025$ is imposed, while it is only 1.0251 and thus closer to the unit circle when $b_y$ is left unrestricted.

There is only little variation in the likelihoods of the different estimates that are reported in the respective left and right blocks of Table 4. For instance, the largest difference in log-likelihoods is obtained from the diffuse prior results shown in the right block of Table 4. If we treat the lower bound as a restriction, a Likelihood Ratio (LR) test of the null hypothesis of the difference in these likelihoods being zero yields $-2(-536.9803 - (-535.9596)) = 2.0414$, which, with one degree of freedom has a $p$-value of 0.1531 and cannot be rejected at conventional significance levels. Hence, there is only limited information in the data to compute a precise estimate of $b_y$. This empirical fact is known in the literature as a ‘flat Phillips curve’.

Given the Stage 1 estimate $\hat{\theta}_1$, Holston et al. (2017) use the following steps to implement median unbiased estimation of their ‘signal-to-noise ratio’ $\lambda_g = \sigma_g / \sigma_{y^*}$.

(a) Use the Stage 1 model to extract an estimate of $y^*_t$ from the Kalman Smoother and construct annualised trend growth as $\Delta \hat{y}^*_{t|T} = 400(\hat{y}^*_t|T - \hat{y}^*_t|T-1|T)$, where $\hat{y}^*_t|T$ here denotes the Kalman Smoothed estimate of $y^*_t$.

(b) Apply the three structural break tests described in (9) to the $\Delta \hat{y}^*_{t|T}$ series. Specifically, replace $Y_t$ in (8) with the constructed $\Delta \hat{y}^*_{t|T}$ series, run the dummy variable regression in the initialization of $b_y$, which yields a value of 0.0921, Holston et al. (2017) use the lower bound value of 0.025 for $b_y$ as the initial value. In the optimisation, this has the effect that the estimate for $b_y$ is effectively stuck at 0.025, although it is not the global optimum in the restricted model, which is at $b_y = 0.097185$ (see also the values of the log-likelihood function reported in the last row of Table 4).

That the output gap is nearly uninformative for inflation (forecasting) once structural break information is conditioned upon — regardless of what measure of the output gap is used or whether it is combined as an ensemble from multiple measures — is shown in Buncic and Müller (2017) for the U.S. and for Switzerland.

Note that, although the series is annualised (scaled by 400), this does not have an impact on the magnitude of the structural break tests. The numerical values that one obtains for $\lambda_g$ are identical if scaled by 100 instead.
in (8), and compute the structural break statistics as defined in (9) and (10). Note that Holston et al. (2017) specify the endpoint values of the search-grid over \( \tau \) at \( \tau_0 = 4 \) and \( \tau_1 = T - 4 \).35

(c) Given the structural break test statistics computed in Step (b), find the corresponding \( \lambda \) values in the look-up table of Stock and Watson (1998). Return the ratio \( \lambda / T = \sigma_g / \sigma_y^* \) which Holston et al. (2017) denoted by \( \lambda_g \), where their preferred estimate of \( \lambda \) is based on the EW statistic of Andrews and Ploberger (1994).

Table 5 shows the range of \( \lambda_g \) estimates computed from the five sets of \( \hat{\theta}_1 \) values reported in Table 4, using all four structural break tests of Stock and Watson (1998). Table 5 is arranged in the same format as Table 4, again showing Holston et al.’s (2017) estimates of \( \lambda_g \) obtained from running their R-Code in the first column of the left block for reference. As can be seen from Table 5, the range of \( \hat{\lambda}_g \) values one obtains from Holston et al.’s (2017) MUE procedure is between 0 to 0.08945 (if only the three structural break tests implemented by Holston et al. (2017) are considered, and up to 0.09419 if the \( L \) statistic is computed as well. Note that this range is not due to statistical uncertainty, but simply due to the choice of structural break test, which prior for \( P_{00} \) is used, and whether the lower bound constraint on \( b_y \) is imposed. Since these estimates determine the relative variation in trend growth through the magnitude of \( \sigma_y^* \), they have a direct impact not only on the variation in the permanent component of GDP, but also on the natural rate of interest through the ratio \( \lambda_g = \sigma_g / \sigma_y^* \) utilized in the later stages of the three step procedure of Holston et al. (2017).

4.1.1. Holston et al.’s (2017) rational for MUE in Stage 1

Comparing the MUE procedure that Holston et al. (2017) implement to the one by Stock and Watson (1998), it is evident that they are fundamentally different. Instead of rewriting the true model of interest in local level form to make it compatible with Stock and Watson’s (1998) look-up tables, Holston et al. (2017) instead formulate a restricted Stage 1 model that not only sets \( a_r \) in the output gap equation to zero, but also makes the awkward assumption that trend growth is constant when computing the ‘preliminary’ estimate of \( y_t^* \).

The rationale behind Holston et al.’s (2017) implementation of MUE is as follows. Suppose we observe trend \( y_t^* \). Then, a local level model for \( \Delta y_t^* \) can be formulated as:

\[
\Delta y_t^* = g_t + \xi_t^{y^*} \tag{17a}
\]
\[
\Delta g_t = \xi_t^g \tag{17b}
\]

where \( \Delta y_t^* \), \( g_t \) and \( \xi_t^{y^*} \) are the analogues to \( GY_t \), \( \beta_t \) and \( u_t \), respectively, in Stock and Wat-
son’s (1998) MUE in (5), with \( \varepsilon^y_t \) in (17a), nonetheless, assumed to be i.i.d. rather than an autocorrelated AR(4) process as \( u_t \) in (5a). Under Stock and Watson’s (1998) assumptions, MUE of the local level model in (17) yields \( \lambda_g = \lambda / T \) defined as:

\[
\frac{\lambda}{T} = \frac{\bar{\sigma}(\varepsilon_t^g)}{\bar{\sigma}(\varepsilon_t^y)} = \frac{\sigma_g}{\sigma_y},
\]

where \( \bar{\sigma}(\cdot) \) denotes again the long-run standard deviation, and the last equality in (18) follows due to \( \varepsilon_t^y \) and \( \varepsilon_t^g \) assumed to be uncorrelated white noise processes.

Since \( \Delta y_t^* \) is not observed, Holston et al. (2017) replace it with the Kalman Smoother based estimate \( \Delta \tilde{y}_t \) obtained from the restricted Stage 1 model in (13). To illustrate what impact this has on their MUE procedure, let \( a_y(L) = (1 - a_{y1}L - a_{y2}L^2) \) and \( a_r(L) = \frac{a_r}{T}(L + L^2) \) denote two lag polynomials that capture the dynamics in the output gap \( \tilde{y}_t \) and the real rate cycle \( \tilde{r}_t = (r_t - r^*) = (r_t - 4g_t - z_t) \), respectively. Also, define \( \psi(L) = a_y(L)^{-1}a_r(L) \) and \( \psi(1) = a_r/(1 - a_{y1} - a_{y2}) \). The output gap equation of the true (full) model in (1) can then be written compactly as:

\[
a_y(L)\tilde{y}_t = a_r(L)\tilde{r}_t + \varepsilon_t^g,
\]

or in differenced form and solved for \( \Delta \tilde{y}_t \) as:

\[
\Delta \tilde{y}_t = a_y(L)^{-1} \left[ a_r(L)\Delta \tilde{r}_t + \Delta \varepsilon_t^g \right].
\]

Observed output, and trend and cycle are related by the identity

\[
y_t = y_t^* + \tilde{y}_t \quad \therefore \Delta y_t = \Delta y_t^* + \Delta \tilde{y}_t.
\]

This relation, together with (17a) and (20), can be written as:

\[
\Delta y_t - \Delta \tilde{y}_t = \Delta y_t^* - a_y(L)^{-1} \left[ a_r(L)\Delta \tilde{r}_t + \Delta \varepsilon_t^g \right] = g_t + \varepsilon^{y^*}_t.
\]

Because the data \( \Delta y_t \) are fixed, any restriction imposed on the \( \Delta \tilde{y}_t \) process translates directly into a misspecification of the right hand side of (22); the \( \Delta y_t^* \) term. In the Stage 1 model, \( a_r \) is restricted to zero. For the relation in (22) to balance, \( \Delta y_t^* \) effectively becomes:\textsuperscript{36}

\[
\Delta y_t^* = g_t + \nu_t^{y^*}
\]

\textsuperscript{36}Note that we need to formulate a local level model for trend growth as in (17) to be able to apply the MUE framework of Stock and Watson (1998). To arrive at (23a), add \( [a_y(L)^{-1}a_r(L)\Delta \tilde{r}_t] \) to both sides of (22). The ring ( ) symbol on \( \nu_t^{y^*} \) highlights again that it is obtained from the restricted model.
\[ \Delta g_t = \epsilon_t^g, \]  

(23b)

where

\[ \hat{\psi}^y_t = \epsilon_t^y + \psi(L) \Delta r_t. \]  

(24)

Holston et al.’s (2017) implementation of MUE relies on the (constructed) local level model relations from the restricted Stage 1 model in (23) and requires us to evaluate the ratio of the long-run standard deviations of \( \epsilon_t^g \) and \( \hat{\psi}^y_t \):

\[ \frac{\bar{\sigma}(\epsilon_t^g)}{\bar{\sigma}(\hat{\psi}^y_t)}. \]  

(25)

Evidently, \( \epsilon_t^g \) in (23b) has not changed, so the numerator of the ‘signal-to-noise ratio’ in (25) is still \( \bar{\sigma}(\epsilon_t^g) = \sigma_g \), due to \( \epsilon_t^g \) being an i.i.d. process. However, the term \( \hat{\psi}^y_t \) in (23a) is not uncorrelated white noise anymore. Moreover, the long-run standard deviation \( \bar{\sigma}(\hat{\psi}^y_t) \) in the denominator of (25) now also depends on the (long-run) standard deviation of \( \psi(L) \Delta r_t \), and will be equal to \( \sigma_y^* \) if and only if \( a_r = 0 \) in the empirical data.

To see what the long-run standard deviation of \( \hat{\psi}^y_t \) looks like, assume for simplicity that \( \epsilon_t^y \) and \( \Delta r_t \) are uncorrelated, so that the long-run standard deviation calculation of \( \hat{\psi}^y_t \) can be broken up into a part involving \( \epsilon_t^y \) and another part involving \( \psi(L) \Delta r_t \), where the latter decomposes as:

\[ \psi(L) \Delta r_t = \psi(L)[\Delta r_t - 4\Delta g_t - \Delta z_t] \]

\[ = \psi(L)[\Delta r_t - 4\epsilon_t^g - \epsilon_t^\pi]. \]

Assuming that the shocks \( \{\epsilon_t^g, \epsilon_t^\pi\} \) are uncorrelated with the (change in the) real rate \( \Delta r_t \), the long-run standard deviation of \( \psi(L) \Delta r_t \) can be evaluated as:

\[ \bar{\sigma}(\psi(L) \Delta r_t) = \bar{\sigma}(\psi(L) \Delta r_t) + \bar{\sigma}(\psi(L)4\epsilon_t^g) + \bar{\sigma}(\psi(L)\epsilon_t^\pi) \]

\[ = \bar{\sigma}(\psi(L) \Delta r_t) + \psi(1)[4\sigma_g + \sigma_z], \]  

(26)

since \( \epsilon_t^g \) and \( \epsilon_t^\pi \) are uncorrelated in the model. Because the nominal rate \( i_t \) is exogenous, it will not be possible to say more about the first term on the right hand side of (26) unless we assume some time series process for \( \Delta r_t \). Suppose that \( r_t \) follows a random walk, so that

\[ \Delta r_t = \epsilon_t^r, \text{ with } \text{Var}(\epsilon_t^r) = \sigma_r^2. \]  

Then \( \bar{\sigma}(\psi(L) \Delta r_t) = a_r/(1 - a_{y,1} - a_{y,2}) \left[ \sigma_r + 4\sigma_g + \sigma_z \right] \), and we obtain \( \bar{\sigma}(\hat{\psi}^y_t) = \sigma_y^* + a_r/(1 - a_{y,1} - a_{y,2}) \left[ \sigma_r + 4\sigma_g + \sigma_z \right] \). The MUE ratio in (25) based

\[ \text{on the empirical data.} \]

\[ \text{37If monetary policy is believed to be effective in cyclical aggregate demand management, then } a_r \text{ cannot be 0 and one would not have formulated the main model of interest assuming that } a_r \text{ is different from zero (viz, negative). Also, this restriction cannot be enforced in the data.} \]
on the restricted Stage 1 model yields:

\[
\frac{\bar{\sigma}(\varepsilon_i^g)}{\bar{\sigma}(\psi_i^{\nu^*})} = \frac{\sigma_g}{\sigma_{y^*} + a_r / (1 - a_{y,1} - a_{y,2}) \left[ \sigma_r + 4\sigma_g + \sigma_z \right]} \neq \frac{\sigma_g}{\sigma_{y^*}}.
\] (27)

Thus, Holston et al.’s (2017) implementation of MUE in Stage 1 cannot recover the ‘signal-to-noise ratio’ of interest \(\frac{\sigma_g}{\sigma_{y^*}}\) from \(\lambda_g\).

Note here that the autocorrelation pattern in \(\nu^{\ast}_y\) is also reflected in the \(\Delta \hat{y}^{\ast}_t\) series which is used as the observable counterpart to \(\Delta y_t^*\) in (23a). That is, \(\Delta \hat{y}^{\ast}_t\) has a significant and sizeable AR(1) coefficient of \(-0.2320\) (standard error \(\approx 0.0649\)). Inline with Step (i) of Stock and Watson’s (1998) implementation of MUE (the GLS step), one would thus need to AR(1) filter the constructed \(\Delta \hat{y}^{\ast}_t\) series used in the local level model before implementing the structural break tests. Accounting for this autocorrelation pattern in \(\Delta \hat{y}^{\ast}_t\) leads to very different \(\lambda_g\) point estimates (see Table 6, which is arranged in the same way as the top half of Table 2, with the last column showing \(\lambda_g = \lambda / T\) rather than \(\sigma_g\) to be able to compare these to column one of Table 5).

4.1.2. Rewriting the Stage 1 model in local level model form

One nuisance with the Stage 1 model formulation of Holston et al. (2017) in (13) is that trend growth is initially assumed to be constant to compute a first estimate of \(y^*_t\). This estimate is then used to construct the empirical counterpart of \(\Delta y^*_t\) to which MUE is applied.

A more coherent way to implement MUE in the context of the Stage 1 model is to rewrite the local linear trend model in local level form. To see how this could be done, we can simplify the Stage 1 model by excluding the inflation equation (13b) and replacing the constant trend growth equation in (13d) with the original trend and trend growth equations in (1d) and (1e). Since the specification of the full model in (1) assumes that the error terms \(\varepsilon_t^{\ell}, \forall \ell = \{\pi, \bar{y}, y^*, g, z\}\) are i.i.d. Normal and mutually uncorrelated, and \(\bar{b}_y \approx 0\) in the unrestricted Stage 1 model (see the results under the heading ‘by Free’ in Table 4), this simplification is unlikely to induce any additional misspecification into the model.

The modified Stage 1 model we can work with thus takes the following form:

\[
y_t = y^*_t + \bar{y}_t \tag{28a}
\]

\[
a_y(L) \bar{y}_t = \bar{\varepsilon}^{\bar{y}}_t \tag{28b}
\]

\[
y^*_t = y^*_{t-1} + \varepsilon^y_t + \varepsilon^y_t \tag{28c}
\]

\[
g_t = g_{t-1} + \bar{\varepsilon}^g_t, \tag{28d}
\]

where \(\varepsilon^y_t = a_r(L) \bar{\varepsilon} + \varepsilon^y_t\) again due to the restriction of the output gap equation of the full model in (1).\(^{38}\) The local linear trend model in (28) can now be rewritten in local level model

\(^{38}\)If the disturbance term \(\varepsilon^y_t\) is i.i.d., then the model in (28) can be recognized as Clark’s (1987) Unobserved

Electronic copy available at: https://ssrn.com/abstract=3897838
form by differencing (28a) and (28b), and bringing \( y_{t-1}^* \) to the left side of (28c) to yield:

\[\Delta y_t = \Delta y_t^* + \Delta \hat{y}_t \quad (30a)\]
\[a_y(L) \Delta \hat{y}_t = \Delta \epsilon_t^\theta \quad (30b)\]
\[\Delta y_t^* = g_{t-1} + \epsilon_t^\gamma \quad (30c)\]
\[g_t = g_{t-1} + \epsilon_t^\delta.\]

Substituting (30b) and (30c) into (30a) gives the local level model:

\[\Delta y_t = g_{t-1} + u_t \quad (31)\]
\[\Delta g_t = \epsilon_t^g, \quad (32)\]

where \( u_t \) is defined as:

\[u_t = \epsilon_t^\gamma + a_y(L)^{-1} \Delta \epsilon_t^\theta \]
\[a_y(L)u_t = a_y(L)\epsilon_t^\gamma + \Delta \epsilon_t^\theta \quad \text{(AR(2))} \]
\[a_y(L)u_t = b(L)\epsilon_t, \quad \text{(MA(2))} \]
\[a_y(L)u_t = b(L)\epsilon_t, \quad (33)\]

with \( b(L)\epsilon_t = a_y(L)\epsilon_t^\gamma + \Delta \epsilon_t^\theta \) on the right hand side of (33) denoting a general MA process. The \( u_t \) term in (33) thus follows a higher order ARMA model. If \( a_r = 0 \), then \( \Delta \epsilon_t^\theta = \epsilon_t^\theta \) in (29) and \( \Delta \epsilon_t^\gamma = \Delta \epsilon_t^\theta \), which is an integrated MA(1) process, so that the right hand side would be the sum of an MA(2) and an MA(1), yielding an overall MA(2) for \( b(L)\epsilon_t \). With \( a_y(L) \) being an AR(2) lag polynomial for the cycle component, we would then get an ARMA(2, 2) for \( u_t \) in (33). If \( a_r \neq 0 \), then \( \Delta \epsilon_t^\gamma \) follows a higher order ARMA process. In the empirical implementation of MUE, I follow Stock and Watson (1998), and use an AR(4) as an approximating model for \( u_t \).

Component (UC) model. However, \( \epsilon_t^\gamma \) is not i.i.d. and instead follows a general ARMA process with non-zero autocovariances, which are functions of \( \sigma_x^2, \sigma_z^2 \), the autocovariances of inflation \( \pi_t \), as well as the exogenously specified interest rate \( i_t \). To see this, recall from Section 2 that the real interest rate gap \( \bar{r}_t \) is defined as \( \bar{r}_t = [i_t - \delta(L)\pi_t - 4g_t - z_t] \), where expected inflation \( \pi_t^e \) is \( \delta(L)\pi_t \) and \( \delta(L) = \frac{1}{4}(1 + L + L^2 + L^3) \), so that we can re-express \( \epsilon_t^\gamma \) as:

\[\epsilon_t^\gamma = a_r(L) [i_t - \delta(L)\pi_t - 4g_t - z_t] + \epsilon_t^\eta. \quad (29)\]

The product of the two lag polynomials \( a_r(L)\delta(L) \) in (29) yields a 5th order lag polynomial for inflation. If \( i_t \) and \( \pi_t \) were uncorrelated white noise processes (which they are clearly not), then we would obtain an MA(5) process for \( \epsilon_t^\gamma \) when \( a_r \) is non-zero. Since \( \pi_t \) is modelled as an integrated AR(4), the implied process for \( \epsilon_t^\gamma \) is a higher order ARMA process, the exact order of which depends on the assumptions one places on the exogenously specified interest rate \( i_t \). To determine this process exactly is of no material interest here. However, the important point to take away from this is that \( \epsilon_t^\gamma \) is autocorrelated and follows a higher order ARMA process. Moreover, if \( i_t, \pi_t, g_t \) and \( z_t \) do not co-integrate, then \( \epsilon_t^\gamma \) will be an \( L(1) \) process.

They also considered an ARMA(2, 3) model (see page 355 in their paper). It is well known that higher order ARMA models can be difficult to estimate numerically due to potential root cancellations in the AR and MA lag polynomials. Inspection of the autocorrelation and partial autocorrelation functions of \( \Delta y_t \) indicate that
The relations in (31) to (33) are now in local level model form to which MUE can be applied to as outlined in equations (5) to (7) in Section 3.1.\textsuperscript{40} To examine if we can recover the ‘signal-to-noise ratio’ of interest $\sigma_g / \sigma_y^*$ from this MUE procedure, we need to evaluate

$$\frac{\bar{\sigma}(\epsilon^g_t)}{\bar{\sigma}(u_t)}. \quad (34)$$

In the numerator of (34), the term $\bar{\sigma}(\epsilon^g_t) = \sigma_g$ as before. Nevertheless, the denominator term $\bar{\sigma}(u_t) = \bar{\sigma}(\epsilon^y_t + a_y(L)^{-1}\Delta \hat{\epsilon}^y_t) \neq \sigma_y^*$. With $\hat{\epsilon}^y_t = a_r(L)\tilde{r}_t + \epsilon^g_t$, we have:

$$\bar{\sigma}(u_t) = \bar{\sigma}(\epsilon^y_t + a_y(L)^{-1}\Delta \hat{\epsilon}^y_t)$$

$$= \bar{\sigma}(\epsilon^y_t + \psi(L)\Delta \tilde{r}_t + a_y(L)^{-1}\Delta \epsilon^g_t), \quad (35)$$

where the middle part in (35) (i.e., $\psi(L)\Delta \tilde{r}_t$) will again be as before in (26) and therefore depend on $\Delta r_t$, $\epsilon^g_t$ and $\epsilon^\tilde{r}_t$. Notice here also that even if we knew $a_r = 0$, so that the middle part in (35) is 0, there is no mechanism to enforce a zero correlation between $\epsilon^y_t$ and $\epsilon^g_t$ in the data, because $u_t$ appears in reduced form in the local level model. We would thus need the empirical correlation between $\epsilon^y_t$ and $\epsilon^g_t$ to be zero for the long-run standard deviation $\bar{\sigma}(u_t)$ to equal $\sigma_y^*$ even when the true $a_r = 0$. Estimates from the existing business cycle literature suggest that trend and cycle shocks are negatively correlated (see for instance Table 3 in Morley et al. (2003), who estimate this correlation to be $-0.9062$, or Table 1 in the more recent study by Grant and Chan (2017) whose estimate is $-0.87$). I obtain an estimate of $-0.9426$ (see Table 8 below).

For completeness, parameter estimates of MUE applied to the local level transformed Stage 1 model defined in (28) are reported in Table 7. This table is arranged in the same way as Table 2, with all computations performed in exactly the same way as before. The MUE results in the last two columns of the bottom part of the table are based on the exponential Wald (EW) structural break test as used in Holston et al. (2017). Overall, these estimates are very similar to Stock and Watson’s (1998) estimates, despite different time periods and GDP data being used. The $\lambda$ (and also $\sigma_g$) estimates are not statistically different from 0, and the MMLE $\hat{\sigma}_g$ of 0.1062 is rather sizeable and quite close to the one implied by MUE.

\subsection*{4.1.3. Estimating the local linear trend version of the Stage 1 model}

So far, ‘pile-up’ at zero problems were examined in the local level model form which is compatible with MUE. As a last exercise, I estimate the modified Stage 1 model in (28) in local linear trend model form. Two different specifications of the model are estimated. The first assumes all error terms to be uncorrelated. This version is referred to as Clark’s (1987) UC0

\footnote{An AR(4) model is more than adequate to capture the time series dynamics of $\Delta y_t$. I have also estimated an ARMA(2, 2) model for $\Delta y_t$, with the overall qualitative conclusions being the same and the quantitative results very similar.}

\footnote{I am grateful to Jim Stock for his email correspondence on this point.}

Electronic copy available at: https://ssrn.com/abstract=3897838
model. The second allows for a non-zero correlation between $\varepsilon_t^{y^*}$ and $\tilde{\varepsilon}_t^{\tilde{y}}$. This version is labelled Clark’s (1987) UC model. The aim here is to not only examine empirically how valid the zero correlation assumption is and to quantify its magnitude, but also to investigate whether ‘pile-up’ at zero problems materialize more generally in UC models. In Table 8, the parameter estimates of the two UC models are reported, together with standard errors of the parameter estimates (these are listed under the columns with the heading Std.error).

As can be seen from the estimates in Table 8, there exists no evidence of ‘pile-up’ at zero problems with MLE in either of these two UC models.\(^{41}\) The estimates of $\sigma_g$ from the two UC models are 0.0463 and 0.0322, respectively, and are based on quarterly data. Expressed at an annualized rate, they amount to approximately 0.1852 and 0.1288, and hence are similar in magnitude to the corresponding MUE based estimates obtained from the transformed model in Table 7. Notice also that the correlation between $\tilde{\varepsilon}_t^{\tilde{y}}$ and $\varepsilon_t^{y^*}$ (denoted by Corr($\tilde{\varepsilon}_t^{\tilde{y}}, \varepsilon_t^{y^*}$) in Table 8) is estimated to be $-0.9426$ ($t-$statistic is approximately $-10$). The magnitude of the $\tilde{\sigma}_{y^*}$ and $\tilde{\sigma}_{\tilde{y}}$ coefficients nearly doubles when an allowance for a non-zero correlation between $\tilde{\varepsilon}_t^{\tilde{y}}$ and $\varepsilon_t^{y^*}$ is made.\(^{42}\)

Figure 4 shows plots of the various trend growth estimates from the modified Stage 1 models reported in Table 7 and Table 8. The plots are presented in the same way as in Figure 3 earlier, with the (annualized) trend growth estimates from the two UC models superimposed. Analogous to the results in Stock and Watson (1998), the variation in the MUE based estimates is once again large. Trend growth can be a flat line when the lower 90% CI of MUE is considered or rather variable when the upper CI bound is used. Interestingly, the MMLE, Clark UC model (with non-zero Corr($\tilde{\varepsilon}_t^{\tilde{y}}, \varepsilon_t^{y^*}$)) and MUE($\hat{\lambda}_{EW}$) trend growth estimates are very similar visually. More importantly, the effect of restricting Corr($\tilde{\varepsilon}_t^{\tilde{y}}, \varepsilon_t^{y^*}$) to zero on the trend growth estimate can be directly seen in Figure 4. The UC0 model produces a noticeably more variable trend growth estimate than the UC model.

Two conclusions can be drawn from this section. Firstly, Holston et al.’s (2017) implementation of MUE in Stage 1 and the resulting $\lambda_g$ estimate cannot recover the ‘signal-to-noise ratio’ of interest $\sigma_g / \sigma_{y^*}$. Secondly, there is no evidence of ‘pile-up’ at zero problems materializing when estimating $\sigma_g$ directly by MLE. Therefore, replacing $\sigma_g$ in $Q$ by $\hat{\lambda}_g \sigma_{y^*}$ in the Stage 2 and Stage 3 model’s log-likelihood functions (see (A.17) and (A.31)) where $\hat{\lambda}_g$ was obtained from MUE applied to the Stage 1 model is not only unsound, but empirically entirely unnecessary.

\(^{41}\)I use a diffuse prior on the initial state vector in the estimation of both UC models, and do not estimate the initial value. This is analogous to MMLE in Stock and Watson (1998). Input data are 100× the log of real GDP.

\(^{42}\)As is common with UC models, the improvement in the log-likelihood due to the addition of the extra correlation parameter is rather small. Although it is important to empirically capture the correlation between $\tilde{\varepsilon}_t^{\tilde{y}}$ and $\varepsilon_t^{y^*}$ as it affects the trend growth estimate (see Figure 4), the overall level of information contained in the data appears to be limited and therefore makes it difficult to decisively distinguish one model over the other statistically. Also, one other aspect of the empirical GDP data that both models fail to capture is the global financial crisis. The level of GDP dropped substantially and in an unprecedented manner. Simply ‘smoothing’ the data to extract a trend as the UC models implicitly do may thus not adequately capture this drop in the level of the series.
4.2. Stage 2 Model

The second stage model of Holston et al. (2017) consists of the following system of equations, which are again a restricted version of the full model in (1):

\[ y_t = y^*_t + \tilde{y}_t \quad (36a) \]
\[ \pi_t = b_{\pi} \pi_{t-1} + (1 - b_{\pi}) \pi_{t-4} + b_y \tilde{y}_{t-1} + \epsilon^\pi_t \quad (36b) \]
\[ a_y(L) \tilde{y}_t = a_0 + \frac{a_r}{2} (r_{t-1} + r_{t-2}) + a_g g_{t-1} + \epsilon^y_t \quad (36c) \]
\[ y^*_t = y^*_{t-1} + g_{t-2} + \epsilon^y_{t-1} \quad (36d) \]
\[ g_{t-1} = g_{t-2} + \epsilon^g_{t-1} \quad (36e) \]

Given the estimate of \( \lambda_g \) from Stage 1, the vector of Stage 2 parameters to be estimated by MLE is:

\[ \Theta_2 = [a_{y,1}, a_{y,2}, a_r, a_0, a_g, b_{\pi}, b_y, \sigma_y, \sigma_{\pi}, \sigma_{y^*}]'. \quad (37) \]

As in the first stage model in (13), I again use the ring symbol (‘) on the disturbance terms in (36c) and (36d) to distinguish them from the i.i.d. error terms of the full model in (1).

Examining the formulation of the Stage 2 model in (36) and comparing it to the full model in (1), it is evident that Holston et al. (2017) make two ‘misspecification’ choices that are important to highlight. First, they include \( g_{t-2} \) instead of \( g_{t-1} \) in the trend equation in (36d), so that the \( \epsilon^y_{t-1} \) error term is in fact:

\[ \epsilon^y_{t} = \epsilon^y_{t-1} + \epsilon^g_{t-1} + \epsilon^g_{t-2} = \epsilon^y_{t-1} + \epsilon^g_{t-1}. \quad (38) \]

As a result of this, \( \epsilon^y_{t-1} \) in (38) follows an MA(1) process, instead of white noise as \( \epsilon^y_{t} \) in (1d). Moreover, due to the \( \epsilon^g_{t-1} \) term in (38), the covariance between the two error terms in (36d) and (36e) is no longer zero, but rather \( \sigma^2_g \). Thus, treating \( W \) in (11) as a diagonal variance-covariance matrix in the estimation of the second stage model is incorrect.

Second, Holston et al. (2017) do not only add an intercept term \( a_0 \) to the output gap equation in (36c), but they also account for only one lag in trend growth \( g_t \), and further fail to impose the \( a_g = -4a_r \) restriction in the estimation of \( a_g \). Due to this, the error term \( \epsilon^y_{t} \) in

---

43See Section A.2 in the Appendix for the exact matrix expressions and expansions of the SSM of Stage 2. In the Q matrix, \( \sigma_g \) is replaced by \( \hat{\lambda}_g \sigma_g \), where \( \hat{\lambda}_g \) is the estimate from the first stage model (see (A.17)). The state vector \( \xi_t \) is initialized using the same procedure as outlined in (12a) and Footnote 27, with the numerical values of \( \xi_{00} \) and \( P_{00} \) given in (A.22) and (A.23).

44Holston et al. (2017) only report the Q matrix in their documentation, which is a diagonal matrix and takes the form given in (A.17). In Section A.2 of the Appendix, I show how this matrix is obtained. In Section A.2.1, the correct Stage 2 model state-space form is provided, applying the same ‘trick’ as used in the Stage 3 state-space model specification. The two Q matrices are listed in (A.17) and (A.25).
(36c) can now be seen to consist of the following two components:

\[\begin{align*}
\hat{\varepsilon}_i^\theta &= -a_r(L)4g_t - a_r(L)z_t + \hat{\varepsilon}_i^\theta - (a_0 + a_g g_{t-1}) \\
&= -a_r(L)z_t + \hat{\varepsilon}_i^\theta - (a_0 + a_g g_{t-1} + a_r(L)4g_t),
\end{align*}\]  

(39)

where the ‘desired terms’ on the right-hand side of (39) are needed for Holston et al.’s (2017) implementation of MUE in the second stage, whose logic I will explain momentarily, while the ‘unnecessary terms’ are purely due to the ad hoc addition of an intercept term, changing lag structure on \(g_t\) and failure to impose the \(a_g = -4a_r\) restriction.

To be consistent with the full model specification in (1), the relations in (36c) and (36d) should have been formulated as:

\[\begin{align*}
a_y(L)\tilde{y}_t &= a_r(L)[r_t - 4g_t] + \hat{\varepsilon}_i^\theta \\
y_t^* &= y_{t-1}^* + g_{t-1} + \varepsilon^y_t
\end{align*}\]  

(40a)

so that only the two missing lags of \(z_t\) from (40a) appear in the error term \(\hat{\varepsilon}_i^\theta\), specifically:

\[
\hat{\varepsilon}_i^\theta = -a_r(L)z_t + \hat{\varepsilon}_i^\theta.
\]  

(41)

Such a specification could have been easily obtained from the full Stage 3 state-space model form (see Section A.3 in the Appendix) by simply removing the last two row entries of the state vector \(\xi_t\) in (A.28), and adjusting the \(H, F, \text{and } S\) matrices in the state and measurement equations to be conformable with this state vector. This is illustrated in Section A.2.1 in the Appendix. The ‘correctly specified’ Stage 2 model should have been formulated as:

\[\begin{align*}
y_t &= y_t^* + \tilde{y}_t \\
\pi_t &= b_\pi\pi_{t-1} + (1 - b_\pi)\pi_{t-2,4} + b_y\tilde{y}_{t-1} + \varepsilon^\pi_t \\
a_y(L)\tilde{y}_t &= a_r(L)[r_t - 4g_t] + \hat{\varepsilon}_i^\theta \\
y_t^* &= y_{t-1}^* + g_{t-1} + \varepsilon^y_t \\
g_{t-1} &= g_{t-2} + \varepsilon^g_{t-1}.
\end{align*}\]  

(42a-e)

To see why this matters, let us examine how one would implement MUE in the Stage 2 model, following again Holston et al.’s (2017) logic as applied in Stage 1. That is, one would first need to define a local level model involving \(z_t\) to be in the same format as in (5). If we assume for the moment that the true state variables \(\tilde{y}_t\) and \(g_t\), as well as parameters \(a_{y,1}, a_{y,2}\), and \(a_r\) are known, and we ignore the econometric issues that arise when these are replaced by estimates, then the following local level model from the ‘correctly specified’ Stage 2 model...
in (42c) can be formed:

\[
\begin{align*}
\text{analogue to GY in (5a)} & \quad \alpha_y(L)\tilde{y}_t - \alpha_r(L)[r_t - 4g_t] = -\alpha_r(L)z_t + \varepsilon^y_t \\
\text{analogue to } \beta_t \text{ in (5a)} & \quad \varepsilon^y_t \text{ in (41)} \\
\text{analogue to } \Delta \beta_t \text{ in (5b)} & \quad \Delta \varepsilon^y_t = -\alpha_r(L)\varepsilon^z_t, \\
\text{analogue to } \lambda/\eta \text{ in (5b)} & \quad \lambda/\eta \text{ in (41)}
\end{align*}
\]

(43a)

(43b)

where \( \alpha_y(L)\tilde{y}_t - \alpha_r(L)[r_t - 4g_t] \) and \( -\alpha_r(L)z_t \) in (43a) are the analogues to GY and \( \beta_t \) in (5a), \( \varepsilon^y_t \) corresponds to \( u_t \) (but is i.i.d. from the full model assumptions in (1) rather than an autocorrelated time series process as \( u_t \) in (5a)), and \( -\alpha_r(L)\Delta z_t \) and \( -\alpha_r(L)\varepsilon^z_t \) are the counterparts to \( \Delta \beta_t \) and \( \lambda/\eta \) in the state equation in (5b).

The equations in (43) are now in local level model form suitable for MUE. The Stage 2 MUE procedure implemented on this constructed GY = \( \alpha_y(L)\tilde{y}_t - \alpha_r(L)[r_t - 4g_t] \) series produces the \( \lambda_z = \lambda/T \) ratio corresponding to (7), that is:

\[
\frac{\lambda}{T} = \frac{\sigma(\Delta \beta_t)}{\sigma(\varepsilon^y_t)} = \frac{\sigma(-\alpha_r(L)\Delta z_t)}{\sigma(y)} = \frac{\alpha_r(1)\sigma_z}{\sigma_y} = \frac{\alpha_r \sigma_z}{\sigma_y}. \tag{44}
\]

The last two steps in (44) follow due to \( \alpha_r(1) = \frac{\alpha_r}{2}(1 + 1^2) = \alpha_r \) and \( \sigma(\varepsilon^y_t) = \sigma_y \), with \( \sigma(\cdot) \) denoting again the long-run standard deviation. The final term in (44) gives Holston et al.’s (2017) ratio \( \lambda_z = \frac{\alpha_r \sigma_z}{\sigma_y} \). This is the logic behind Holston et al.’s (2017) implementation of MUE in Stage 2.

However, because Holston et al. (2017) define the Stage 2 model in ‘misspecified’ form in (36), \( \varepsilon^y_t \) is no longer simply equal to \( -\alpha_r(L)z_t + \varepsilon^y_t \) as needed for the right-hand side of (43a), but now also includes the ‘unnecessary terms’ \[ a_0 + a_g g_{t-1} + a_r(L)4g_t \] (see the decomposition in (39)). What effect this has on the Stage 2 MUE procedure can be seen by first rewriting \( a_g g_{t-1} \) as:

\[
\begin{align*}
a_g g_{t-1} & = \frac{a_g}{2}(g_{t-1} + g_{t-1}) \\
& = \frac{a_g}{2}(g_{t-1} + g_{t-2} + \varepsilon^g_{t-1}) \quad \text{(from (36e))} \\
& = a_g(L)g_t + \frac{a_g}{2} \varepsilon^g_{t-1}, \tag{45}
\end{align*}
\]

where \( a_g(L) = \frac{a_g}{2}(L + L^2) \). The additional ‘unnecessary terms’ on the right-hand side of (39)

\[46To make this clear, MUE returns an estimate of \( \lambda \) by using the look-up table on page 354 in Stock and Watson (1998) to find the closest matching value of one of the four structural break test statistics defined in (9) and (10) which test for a structural break in the unconditional mean of the constructed GY series by running a dummy variable regression of the form defined in (8).

\[46In Laubach and Williams (2003), \( \lambda_z \) is curiously defined as the ratio \( \alpha_r \sigma_z/\sigma_y \sqrt{2} \) (see page 1064, second paragraph on the right). It is not clear where the extra \( \sqrt{2} \) term comes from.
become:

\[ - \left[ a_0 + a_g \tilde{g}_{t-1} + a_r(L)4\tilde{g}_t \right] = - \left[ a_0 + a_g(L)\tilde{g}_1 + \frac{a_g}{2} \epsilon_{t-1}^g + a_r(L)4\tilde{g}_t \right] = - \left[ a_0 + \frac{(a_g + 4a_r)}{2} (\tilde{g}_{t-1} + \tilde{g}_{t-2}) + \frac{a_g}{2} \epsilon_{t-1}^g \right]. \]

(46)

In Holston et al.'s (2017) Stage 2 model in (36), the constructed local level model takes then the form:

\[
\begin{align*}
\tilde{a}_y(L)\tilde{y}_t - a_0 &- a_r(L)r_t - a_g \tilde{g}_{t-1} = - a_r(L)z_t + \nu_t^\beta, \\
-a_r(L)\Delta z_t &= - a_r(L)\tilde{e}_t^\beta, \\
\end{align*}
\]

(47a)

(47b)

where \( \nu_t^\beta \) in (47a) is the misspecified analogue to \( \epsilon_t^\beta \) in (43a) and is defined as:

\[ \nu_t^\beta = \epsilon_t^\beta - \left[ a_0 + \frac{(a_g + 4a_r)}{2} (\tilde{g}_{t-1} + \tilde{g}_{t-2}) + \frac{a_g}{2} \epsilon_{t-1}^g \right]. \]

(48)

As can be seen, the error term \( \nu_t^\beta \) in (48) will not be white noise. Moreover, forming the MUE \( \lambda/T \) ratio from the model in (47) in the same way as in (44) leads to:

\[ \frac{\lambda}{T} = \frac{\sigma(-a_r(L)\Delta z_t)}{\tilde{\sigma}(\nu_t^\beta)} = \frac{a_r(1)\sigma_z}{\tilde{\sigma}(\nu_t^\beta)} = \frac{a_r\sigma_z}{\tilde{\sigma}(\nu_t^\beta)}. \]

(49)

and now requires the evaluation of the long-run standard deviation of \( \nu_t^\beta \) in the denominator. This quantity will not be equal to \( \sigma_{\tilde{y}} \) as from the ‘correctly’ specified Stage 2 model defined in (42). Note here that even if \( (a_g + 4a_r) = 0 \) in the data, the long-run standard deviation of \( \nu_t^\beta \) will also depend on \( \frac{a_g}{2} \sigma_{\tilde{y}} \) because of the extra \( \frac{a_g}{2} \epsilon_{t-1}^g \) term in \( \nu_t^\beta \), so that:

\[ \lambda_z = \frac{\lambda}{T} = \frac{a_r\sigma_z}{\sigma_{\tilde{y}} + \frac{1}{2} a_g \sigma_{\tilde{y}}}. \]

(50)

Thus, MUE applied to Holston et al.'s (2017) ‘misspecified’ Stage 2 model defined in (36) cannot recover the ratio of interest \( \lambda_z = a_r\sigma_z/\sigma_{\tilde{y}} \) as claimed. For the reader’s convenience, the correctly and misspecified Stage 2 model equations are listed side by side in (51) below:

| Correctly specified | HLW misspecified |
|---------------------|------------------|
| \( y_t = y_t^* + \tilde{y}_t \) | \( y_t = y_t^* + \tilde{y}_t \) |
| \( b_\pi(L)\pi_t = b_y \tilde{y}_{t-1} + \epsilon_t^\pi \) | \( b_\pi(L)\pi_t = b_y \tilde{y}_{t-1} + \epsilon_t^\pi \) |
| \( a_y(L)\tilde{y}_t = a_r(L) [r_t - 4\tilde{g}_1] + \tilde{e}_t^g \) | \( a_y(L)\tilde{y}_t = a_0 + a_r(L) r_t + a_g \tilde{g}_{t-1} + \tilde{e}_t^g \) |
| \( y_t^* = y_{t-1}^* + \tilde{g}_{t-1} + \epsilon_{t-1}^g \) | \( y_t^* = y_{t-1}^* + \tilde{g}_{t-2} + \epsilon_{t-1}^g \) |
| \( \tilde{g}_{t-1} = \tilde{g}_{t-2} + \epsilon_{t-1}^g \) | \( \tilde{g}_{t-1} = \tilde{g}_{t-2} + \epsilon_{t-1}^g \) |

(51a)

(51b)

(51c)

(51d)

(51e)
Parameter estimates corresponding to the various Stage 2 models are report in Table 9. The first and second columns show replicated results which are based on Holston et al.’s (2017) R-Code as well as my own implementation and serve as reference values. In the third column under the heading ‘MLE(σₙₐ)’, σₙₐ is estimated directly by MLE together with the other parameters of the model without using ˆλₙₐ from Stage 1. The last column under the heading ‘MLE(σₙₐ).M₀’ reports estimates obtained from the ‘correctly specified’ Stage 2 model defined in (42), where σₙₐ is once again estimated directly by MLE.

The results in Table 9 can be summarized as follows. First, there exists no evidence of ‘pile-up’ problems when estimating σₙₐ directly by MLE; not in the ‘misspecified’ Stage 2 model, nor in the ‘correctly specified’ one. This finding is consistent with the earlier results from the first stage. In fact, the Stage 2 MLE of σₙₐ is nearly 50% larger than the estimate implied by ˆλₙₐ from MUE in Stage 1. The Stage 1 model is redundant. Second, the point estimate of aₙ is about eight times the magnitude of −aₙ, so that (aₙ + 4aₙ) ≈ 0.3132 ̸= 0. And third, despite the different Stage 2 model specifications, the resulting parameter estimates as well as the log-likelihood values across the three different models in columns two to four of Table 9 are rather similar. This suggests that, overall, the data are uninformative about the model parameters.

Note here that, although the parameter estimates from the various Stage 2 models are similar, it is the estimate of λ₂ from Holston et al.’s (2017) implementation of MUE on the ‘misspecified’ Stage 2 model that are spuriously large and very different to the ones obtained from the ‘correctly specified’ Stage 2 model. This is what is shown in the next section.

4.2.1. Holston et al.’s (2017) implementation of MUE in Stage 2

Conceptually, MUE in Stage 2 needs to be implemented following the same logic as in Stage 1 before. First, one needs to construct an observable counterpart to GYₜ as given in (43a) from the Stage 2 model estimates. Then, the four structural break tests described in Section 3.1 are applied to test for a break in the unconditional mean of (the AR filtered) GYₜ series. This corresponds to Step (ii) in Stock and Watson’s (1998) procedural description. Constructing a local level model of the form described in (43) enables us to implement MUE to yield the ratio λ/T = σ(∆βₜ)/σ(εₜ) as defined in (44).

Holston et al.’s (2017) implementation of MUE in Stage 2, nonetheless, departs from this description in two important ways. First, instead of using the ‘correctly specified’ Stage 2 model defined in (42), they work with the ‘misspecified’ model given in (36). Second, rather than leaving the aᵧ₁, aᵧ₂, aᵣ, aₙ and aₒ parameters fixed at their Stage 2 estimates and constructing the observable counterpart to GYₜ in (47a) only once outside the dummy variable

\[ \frac{\lambda}{T} = \frac{\sigma(\Delta \beta_t)}{\sigma(\varepsilon_t^0)} \]

\[ \lambda_g = \frac{\sigma_g}{\sigma_y} \]

\[ \text{This is also the case for Euro Area, the U.K., and Canadian data (see Buncic, 2020).} \]

\[ \text{I use the same initial values for the parameter and the state vector (mean and variance) as in the exact replication of Holston et al. (2017). Using a diffuse prior instead leads to only minor differences in the numerical values. The implied} \]

\[ \lambda_g \text{ and } \sigma_g \text{ estimates are shown in brackets and were computed from the 'signal-to-noise ratio' relation } \lambda_g = \frac{\sigma_g}{\sigma_y}. \]
regression loop, Holston et al. (2017) essentially ‘re-estimate’ these parameters by including the vector $X_t$ defined in (53) below as a extra regressor in the structural break regression in (54). This is not only entirely unnecessary, as the $a_{r,1}$, $a_{r,2}$, $a_{r}$, $a_{g}$ and $a_{0}$ parameters are already known from the Stage 2 model estimates. Moreover, it deviates from the simulation set-up used by Stock and Watson (1998) to generate the look-up values in Table 3 of their paper. In the ‘misspecified’ Stage 2 model, the additional regressors substantially increase the size and variability of not only the dummy variable coefficients $\hat{\zeta}_1$ in (54), but also the corresponding $F$ statistics used in the computation of the MW, EW, and QLR structural break test statistics. Recall that these break statistics are then ‘looked-up’ in Table 3 of Stock and Watson (1998) to obtain the MUE of $\lambda_z$.

It will be usefully here to list the main steps that Holston et al. (2017) follow in their implementation of MUE in the second stage to arrive at an estimate of $\lambda_z$.

(I) Given the Stage 2 estimate $\hat{\theta}_2$ from the model in (36), use the Kalman Smoother to obtain (smoothed) estimates of the latent state vector $\xi_t = [y^*_t, y^*_{t-1}, y^*_t, g_{t-1}]'$. Then form estimates of the cycle variable and its lags as $\hat{y}_{t-i|T} = (y_{t-i} - \hat{y}^*_{t-i|T})$, $\forall i = 0, 1, 2$.

(II) Construct

$$Y_t = \hat{y}_{t|T}$$

and the $(1 \times 5)$ vector

$$X_t = [\hat{y}_{t-1|T}, \hat{y}_{t-2|T}, (r_{t-1} + r_{t-2})/2, \hat{g}_{t-1|T}, 1]$$

where $r_t$ is the real interest rate, $\hat{g}_{t-1|T}$ is the Kalman Smoothed estimate of $g_{t-1}$ and 1 is a scalar to capture the constant $a_0$ (intercept term).

(III) For each $\tau \in [\tau_0, \tau_1]$, run the following dummy variable regression analogous to (8):

$$Y_t = X_t \phi + \zeta_1 D_t(\tau) + \epsilon_t$$

where $X_t$ is as defined in (53) and $\phi$ is a $(5 \times 1)$ parameter vector. The structural break dummy variable $D_t(\tau)$ takes the value 1 if $t > \tau$ and 0 otherwise, and $\tau = \{\tau_0, \ldots, \tau_1\}$ is an index of grid points between $\tau_0 = 4$ and $\tau_1 = T - 4$. Use the sequence of $F$ statistics $\{F(\tau)\}^\tau_{\tau_0}$ on the dummy variable coefficients to compute the MW, EW, and QLR structural break test statistics needed for MUE.

(IV) Given the structural break statistics computed in Step (III), find the corresponding $\lambda$ values in look-up Table 3 of Stock and Watson (1998) and return the ratio $\lambda/T = \lambda_z$, where their preferred estimate of $\lambda$ is again based on the EW break statistic defined in (9b) as in the Stage 1 MUE.

In the top and bottom panels of Figure 5 I show plots of the sequences of $F$ statistics $\{F(\tau)\}^\tau_{\tau_0}$ computed from Holston et al.’s (2017) ‘misspecified’ Stage 2 model and the ‘cor-
rectly specified’ Stage 2 model defined in (42), respectively. Two sets of sequences are drawn in each panel. The first sequence, which I refer to as ‘time varying φ’ (drawn as a red line in Figure 5) is constructed by following Holston et al.’s (2017) implementation outlined in Steps (I) to (III) above. I call this the ‘time varying φ’ sequence because the \( a_{y,1}, a_{y,2}, a_r, a_\gamma \) and \( a_0 \) parameters needed to ‘construct’ the observable counterpart to \( GY_t \) in (47a) are effectively ‘re-estimated’ for each \( \tau \in [\tau_0, \tau_1] \) in the dummy variable regression loop due to the inclusion of the extra \( \boldsymbol{X}_t \phi \) term in (54). For the ‘correctly specified’ Stage 2 model in (42), \( \boldsymbol{X}_t \) in (53) is replaced by the \((1 \times 3)\) vector \([\hat{y}_{t-1|T}, \hat{y}_{t-2|T}, (r_{t-1} + r_{t-2} - 4(\hat{s}_{t-1|T} + \hat{s}_{t-2|T}))/2]\).

In the second sequence, labelled ‘constant φ’ in Figure 5 and drawn as a blue line, the observable counterpart to \( GY_t \) is computed only once outside the structural break regression loop, with the dummy variable regression performed without the extra \( \boldsymbol{X}_t \phi \) term in (54), i.e., it is computed in its ‘original’ form as given in (8). More specifically, for the ‘misspecified’ and ‘correctly specified’ Stage 2 models, the observable counterparts to the \( GY_t \) series are constructed as:

\[
GY_t = \hat{y}_{t|T} - \hat{a}_{y,1}\hat{y}_{t-1|T} - \hat{a}_{y,2}\hat{y}_{t-2|T} - \hat{a}_r(r_{t-1} + r_{t-2})/2 - \hat{a}_\gamma\hat{s}_{t-1|T} - \hat{a}_0, \tag{55}
\]

and

\[
GY_t = \hat{y}_{t|T} - \hat{a}_{y,1}\hat{y}_{t-1|T} - \hat{a}_{y,2}\hat{y}_{t-2|T} - \hat{a}_r(r_{t-1} + r_{t-2} - 4(\hat{s}_{t-1|T} + \hat{s}_{t-2|T}))/2, \tag{56}
\]

respectively. The \( \hat{a}_{y,1}, \hat{a}_{y,2}, \hat{a}_r, \hat{a}_\gamma, \) and \( \hat{a}_0 \) coefficients are the (full sample) estimates reported in columns 2 and 4 of Table 9 under the headings ‘Replicated’ and ‘MLE(\( \sigma_\gamma \)).M0’, with the corresponding latent state estimates from the respective models.

As can be seen from Figure 5, the \( \{F(\tau)\}_{\tau = \tau_0}^{\tau_1} \) sequences from the ‘correctly specified’ Stage 2 models shown in the bottom panel are not only smaller overall, but they are nearly unaffected by Holston et al.’s (2017) approach to ‘re-estimate’ the parameters in the structural break loop. Both, the ‘constant φ’ and the ‘time varying φ’ versions generate \( \{F(\tau)\}_{\tau = \tau_0}^{\tau_1} \) sequences that are overall very similar, with their maximum values being around 4.5. For the ‘misspecified’ Stage 2 model shown in the top panel, this is not the case. The variation as well as the magnitude of \( \{F(\tau)\}_{\tau = \tau_0}^{\tau_1} \) from the ‘time varying φ’ and ‘constant φ’ implementations are vastly different, with the former having a much higher mean and maximum value.

These large differences in the \( \{F(\tau)\}_{\tau = \tau_0}^{\tau_1} \) sequences from the ‘misspecified’ Stage 2 models also lead to very different estimates of \( \lambda_\gamma \). This can be seen from Table 10, which shows the

---

49 The same sequence computed from data updated to 2019:Q2 is shown in Figure A.4 in the Appendix.

50 Note that Stock and Watson’s (1998) MUE look-up table values for \( \lambda \) were constructed by simulation with the structural break test testing the unconditional mean of the \( GY_t \) series for a break, without any other variables being included in the regression. This form of the structural break regression is thus compatible with Stock and Watson’s (1998) look-up table values.

51 For instance, \( \hat{s}_{t-1|T} \) in (55) is the Kalman Smoothed estimate of trend growth from Holston et al.’s (2017) ‘misspecified’ Stage 2 model, while trend growth \( \hat{s}_{t-1|T} \) in (56) is the corresponding estimate from the ‘correctly specified’ Stage 2 model.
resulting $\lambda_z$ estimates in the top part with the corresponding $L$, MW, EW, and QLR structural break test statistics in the bottom part. Table 10 is arranged further into a left and a right column block, referring to the ‘time varying $\phi$’ and the ‘constant $\phi$’ MUE implementations for the three different models reported in (9). ‘Replicated’ refers to the baseline replicated results, ‘$\text{MLE}(\sigma_g)$’ corresponds to the ‘misspecified’ Stage 2 model but with $\sigma_g$ estimated by MLE, and ‘$\text{MLE}(\sigma_g).M_0$’ is from the ‘correctly specified’ Stage 2 model with $\sigma_g$ again estimated directly by MLE. The ‘HLW.R-File’ column lists the results from Holston et al.’s (2017) R-Code. Note that Holston et al. (2017) do not report estimates based on Nyblom’s (1989) $L$ statistic. The entries in the $L$ rows in Table 10 under ‘HLW.R-File’ thus simply list ‘—’. 90% confidence intervals for $\lambda_z$ and $p-$values for the structural break tests are reported in square and round brackets, respectively.\footnote{As in the replication of Stock and Watson’s (1998) results reported in (2), these were again obtained from their GAUSS files.}

Consistent with the visual findings from Figure 5, the structural break statistics from the ‘misspecified’ Stage 2 model shown under the ‘Replicated’ heading for the ‘time varying $\phi$’ and ‘constant $\phi$’ settings are very different. The MW, EW, and QLR statistics are approximately 4 to 5 times larger under the ‘time varying $\phi$’ setting than under the ‘constant $\phi$’ scenario. Because Nyblom’s (1989) $L$ statistic is constructed as the scaled cumulative sum of the demeaned ‘GYi’ series and thus does not require the partitioning of data, creation of dummy variables, or looping over break dates $\tau$, it is not affected by this choice, yielding the same test statistic of about 0.05 under both settings.

Under the ‘time varying $\phi$’ setting, the MW, EW, and QLR statistics and Nyblom’s (1989) $L$ statistic generate vastly different $\lambda_z$ estimates. Nyblom’s (1989) $L$ statistic is highly insignificant with a $p-$value of 0.87, resulting in a $\lambda_z$ estimate of exactly 0 (Nyblom’s (1989) $L$ statistic is less than 0.118, the smallest value in Stock and Watson’s (1998) look-up Table 3 which corresponds to $\lambda = 0$). The MW, EW, and QLR structural break statistics on the other hand are either weakly significant or marginally insignificant, with $p-$values between 0.045 and 0.13. These borderline significant structural break statistics generate sizable $\lambda_z$ point estimates between 0.025 and 0.034. The resulting 90% confidence intervals for $\lambda_z$ are, nonetheless, rather wide with 0 as the lower bound, suggesting that these point estimates are not significantly different from zero.\footnote{Given the earlier discussion in Section 3.1 and the ARE results in Table 2 of Stock and Watson (1998), we know that MUE can be a very inefficient estimator.} Under the ‘constant $\phi$’ setting, the four structural break statistics and the resulting $\lambda_z$ estimates tell a consistent story (see the ‘Replicated’ heading in the right column block). All structural break statistics are highly insignificant, with their respective $\lambda_z$ point estimates being equal to zero.

For the ‘correctly specified’ Stage 2 models shown under the headings ‘$\text{MLE}(\sigma_g).M_0$’ in Table 10, the ‘time varying $\phi$’ and the ‘constant $\phi$’ estimates of $\lambda_z$ reflect the visual similarity of the $\{F(\tau)\}_{\tau=\tau_0}^{\tau_1}$ sequences shown in the bottom panel of Figure 5. The $\lambda_z$ point estimates are of the same order of magnitude, very close to zero (they are exactly equal to zero for

52 As in the replication of Stock and Watson’s (1998) results reported in (2), these were again obtained from their GAUSS files.
53 Given the earlier discussion in Section 3.1 and the ARE results in Table 2 of Stock and Watson (1998), we know that MUE can be a very inefficient estimator.
Nyblom’s (1989) L statistic and MW under the ‘constant φ’ setting), and most importantly, substantially smaller than those constructed from Holston et al.’s (2017) ‘misspecified’ Stage 2 model.54

What is causing this large difference in the \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) sequences between the ‘misspecified’ and ‘correctly specified’ Stage 2 models in the ‘time varying φ’ setting? There are two components. First, the Kalman Smoothed estimates of the output gap (cycle) \( \hat{\gamma}_{t|T} \) and of (annualized) trend growth \( \hat{g}_{t|T} \) can be quite different from these two models, despite the parameter estimates and values of the log-likelihoods being very similar. This difference is more pronounced for the cycle estimate \( \hat{y}_{t|T} \), particulary towards the end of the sample period than for the trend growth estimate \( \hat{g}_{t|T} \) (see Figure A.2 in the Appendix which shows a comparison of \( \hat{y}_{t|T} \) and \( \hat{g}_{t|T} \) from the ‘misspecified’ and ‘correctly specified’ Stage 2 models).

Second, the parameter restriction \((a_g + 4a_r)\) on the relationship between the real rate and trend growth matters. More specifically, when conditioning on \( X_t \) in (54), it is the restriction \((r_t - 1 - 4\hat{g}_{t-1|T})\) in \( X_t \) that makes the largest difference to the \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) sequence. To see this, I show plots of the \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) sequences from various \( X_t \) constructs corresponding to the different Stage 2 model specifications in Figure A.3 in the Appendix. I use the ‘correctly specified’ Stage 2 model’s \( \{\hat{g}_{t-1|T}\}_{i=1}^{2} \) and \( \hat{g}_{t-1|T} \) estimates to form three sets of \( X_t \) vectors for the dummy variable regressions in (54). These are:

\[
X_t = [\hat{g}_{t-1|T}, \hat{y}_{t-2|T}, (r_t - 1 + r_t - 2)/2, \hat{g}_{t-1|T}, 1] \quad (57a)
\]
\[
X_t = [\hat{y}_{t-1|T}, \hat{y}_{t-2|T}, r_t - 1, \hat{g}_{t-1|T}, 1] \quad (57b)
\]
\[
X_t = [\hat{y}_{t-1|T}, \hat{y}_{t-2|T}, (r_t - 1 - 4\hat{g}_{t-1|T})] \quad (57c)
\]

and are labelled accordingly in Figure A.3 (the preceding ‘MLE(\(\sigma_g\)).M_0’ signifies that these were constructed using the \( \{\hat{g}_{t-1|T}\}_{i=1}^{2} \) and \( \hat{g}_{t-1|T} \) estimates from the ‘correctly specified’ Stage 2 model). The corresponding \( Y_t \) dependent variable for these structural break regressions also uses the ‘correctly specified’ Stage 2 model’s output gap estimate \( \hat{y}_{t|T} \). The \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) sequences from Holston et al.’s (2017) ‘misspecified’ and the ‘correctly specified’ Stage 2 models are superimposed as reference values and are denoted by ‘HLW’ and ‘MLE(\(\sigma_g\)).M_0’.

The plot corresponding to (57a) (orange dashed line in Figure A.3) shows a rather small difference relative to the ‘HLW’ benchmark (blue solid line). Thus, exchanging \( \{\hat{g}_{t-1|T}\}_{i=1}^{2} \) and \( \hat{g}_{t-1|T} \) from Holston et al.’s (2017) ‘misspecified’ Stage 2 model for those from the ‘correctly specified’ one only has a small impact on the \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) sequence and is most visible over the 1994 to 2000 period. Dropping the second lag in \( r_t \) from \( X_t \) in (57b) (see the cyan dotted line in Figure A.3) also has only a small impact on the \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) sequence. The biggest effect on \( \{F(\tau)\}_{\tau=\tau_0}^{\tau_1} \) has the restriction \((r_t - 1 - 4\hat{g}_{t-1|T})\) as imposed in (57c) (green dashed-dotted line Figure A.3). This is evident from the near overlap with the red solid line corresponding

---

54 In Table A.2 in the Appendix, I present these Stage 2 MUE results for data that was updated to 2019:Q2. The conclusion is the same.
to the ‘correctly specified’ Stage 2 model’s \( \{F(\tau)\}_{T=1}^{T=10} \) sequence. Recall that the only difference between these two is that an extra lag of \((r_{t-1} - 4\hat{g}_{t-1} | T)\) is added to \(X_t\), and that these enter as an average, viz, 
\[
X_t = [\hat{y}_{t-1}|T, \hat{y}_{t-2}|T, (r_{t-1} + r_{t-2} - 4\{\hat{g}_{t-1} | T + \hat{g}_{t-2} | T\})]/2.
\]

4.2.2. What does Holston et al.’s (2017) Stage 2 MUE procedure recover?

Holston et al.’s (2017) Stage 2 MUE procedure implemented on the ‘misspecified’ Stage 2 model leads to spuriously large estimates of \(\lambda_z\) when the true value is zero. To illustrate this, I implement two simple simulation experiments.

In the first experiment, I simulate data from the full structural model in (1) using the Stage 3 parameter estimates of Holston et al. (2017) reported in column one of Table 12 as the true values that generate the data, but with ‘other factor’ \(z_t\) set to zero for all \(t\). The natural rate \(r^*_t\) in the output gap equation in (1c) is thus solely determined by (annualized) trend growth, that is, \(r^*_t = 4g_t\), which implies that \(\lambda_z\) is zero in the simulated data.\(^{55}\) I then implement Holston et al.’s (2017) Stage 2 MUE procedure on the simulated data following steps (I) to (IV) outlined in Section 4.2.1 above to yield a sequence of \(S = 1000\) estimates of \(\lambda_z\) \(\{\{\hat{\lambda}^s_z\}_{s=1}^{S}\}\).

I use two different scenarios for \(\theta_2\) in the Kalman Smoother recursions described in Step (I) to extract the latent cycle as well as trend growth series needed for the construction of \(Y_t\) and \(X_t\) in the dummy variable regression in (54). The first scenario simply takes Holston et al.’s (2017) empirical Stage 2 estimate \(\hat{\theta}_2\) as reported in column one of Table 9, and keeps these values fixed for all 1000 generated data sequences when applying the Kalman Smoother. In the second scenario, I re-estimate the Stage 2 parameters for each simulated sequence to obtain new estimates \(\hat{\theta}^s_2\), \(\forall s = 1, \ldots, S\). I then apply the Kalman Smoother using these estimates to generate the \(Y_t\) and \(X_t\) sequences for the regression in (54).

Finally, I repeat the above computations on data that were generated from the full model in (1) with the natural rate of interest determined by both factors, namely, \(r^*_t = 4g_t + z_t\), where \(z_t\) was simulated as a pure random walk, with its standard deviation set at the Stage 3 estimate of \(\sigma_z\) and \(a_r\), i.e., at \(\sigma_z = \lambda_z\sigma_g/a_r \approx 0.15\) (see row \(\sigma_z\) (implied) of column one in Table 12). The goal here is to provide a comparison of the range of \(\lambda_z\) estimates that can be obtained when implementing Holston et al.’s (2017) Stage 2 MUE procedure on data that

\(^{55}\)To implement the simulations from the full Stage 3 model, I need to define a process for the exogenously determined interest rate in Holston et al.’s (2017) model. For simplicity, I estimate a parsimonious, but well fitting, ARMA(2, 1) model for the real interest rate series, and then use the ARMA(2, 1) coefficients to generate a sequence of 229 simulated observations for \(r_t\). Recall that Holston et al. (2017) use data from 1960:Q1, where the first 4 quarters are used for initialisation of the state vector, so that in total \(4 + 225 = T\) observations are available. The remaining series are simulated from the Stage 3 model given in (1). To get a realistic simulation path from the Stage 3 model, I initialize the first four data points for the simulated inflation series at their observed empirical values. For the \(y^*_t\) series, the HP-filter based trend estimates of GDP (also utilized in the initialisation of the State vector in Stage 1) are used to set the first four observations. The cycle variable \(\tilde{y}_t\) initializes at zero, while trend growth \(g_t\) is initialized at 0.75, which corresponds to an annualized rate of 3 percent. In the analysis that requires a simulated path of ‘other factor’ \(z_t\), i.e., when the natural rate is generated from \(r^*_t = 4g_t + z_t\), the first four entries in \(z_t\) are initialized at zero. A total of \(S = 1000\) sequences are simulated with a total sample size of 229 observations, where the first four entries are discarded in later analysis.
were generate with and without ‘other factor’ $z_t$ in the data generating processes (DGPs) of the natural rate.

In Table 11, summary statistics of $\hat{\lambda}_z^s$ from the two different DGPs are reported. The left column block shows results for the two different DGPs when the Stage 2 parameter vector $\theta_2$ is held fixed at the estimates reported in column one of Table 9. The right column block shows corresponding results when $\theta_2$ is re-estimated for each simulated data series. The summary statistics are the minimum, maximum, standard deviation, mean, and median of $\hat{\lambda}_z^s$, as well as the relative frequency of obtaining a value larger than the empirical point estimate of Holston et al. (2017). This point estimate and the corresponding relative frequency are denoted by $\hat{\lambda}_z^\text{HLW}$ and $\Pr(\hat{\lambda}_z^s > \hat{\lambda}_z^\text{HLW})$, respectively. To complement the summary statistics in Table 11, histograms of $\hat{\lambda}_z^s$ are shown in Figure 6 to provide visual information about its sampling distribution.

From the summary statistics in Table 11 as well as the histograms in Figure 6 it can be seen how similar the $\hat{\lambda}_z^s$ coefficients from these two different DGPs are. For instance, when the data were simulated without ‘other factor’ $z_t$ (ie., $\lambda_z = 0$), the sample mean of $\hat{\lambda}_z^s$ is 0.028842. When the data were generated from the full model with $r_t^* = 4g_t + z_t$, the sample mean of $\hat{\lambda}_z^s$ is only 6.53% higher at 0.030726. Similarly, the relative frequencies $\Pr(\hat{\lambda}_z^s > \hat{\lambda}_z^\text{HLW})$ for these two DGPs are 45.70% and 49%, respectively. The inclusion of ‘other factor’ $z_t$ in the DGP of the natural rate thus results in only a 3.3 percentage points higher $\Pr(\hat{\lambda}_z^s > \hat{\lambda}_z^\text{HLW})$. The histograms in Figure 6 paint the same overall picture. The Stage 2 MUE implementation has difficulties to discriminate between these two DGPs.

In a second experiment I simulate DGPs from entirely unrelated univariate ARMA processes of the individual components of the $Y_t$ and $X_t$ series needed for the regressions in (54). To match the time series properties of the $Y_t$ and $X_t$ elements given in (52) and (53), I fit simple low-order ARMA models to $\hat{g}_{t|T}$, $r_t$ and $\hat{g}_{t|T}$, and then use these ARMA estimates to simulate artificial data. Finally, I apply Holston et al.’s (2017) Stage 2 MUE procedure to the simulated data as before, nevertheless starting from Step (II), and thereby skipping the Kalman Smoother step. The full results from the second experiment are reported in Table A.1 and Figure A.1 in the Appendix. These yield magnitudes of $\hat{\lambda}_z^s$ that are similar to those from the first simulation experiment, with mean estimates being between 0.026117 and 0.031798, and relative frequencies corresponding to $\Pr(\hat{\lambda}_z^s > \hat{\lambda}_z^\text{HLW})$ being between 38.40% and 49.80%.

From these two experiments, it seems that it is Holston et al.’s (2017) procedure itself that leads to the spuriously large estimates of $\lambda_z$, regardless of the properties of the data.

---

56 When the Stage 2 parameter vector $\theta_2$ is re-estimated for each simulated sequence shown in the right column block in Table 11, the sample means as well as the relative frequency $\Pr(\hat{\lambda}_z^s > \hat{\lambda}_z^\text{HLW})$ are somewhat lower at 0.025103 and 0.027462, and 33.90% and 39.30%, respectively.

57 I use 4 different time series processes for $\hat{g}_{t|T}$ in these simulations. Complete details of the simulation design are given in Section A.4 of the Appendix.
4.3. Stage 3 Model

The analysis so far has demonstrated that the ratios of interest \( \lambda_g = \sigma_g / \sigma_y \) and \( \lambda_z = a_r \sigma_z / \sigma_y \) required for the estimation of the full structural model in (1) cannot be recovered from Holston et al.’s (2017) MUE procedure implemented in Stages 1 and 2. Moreover, since their procedure is based on the ‘misspecified’ Stage 2 model in (36), it results in a substantially larger estimate of \( \lambda_z \) than when implemented on the ‘correctly specified’ Stage 2 model in (42). This larger estimate of \( \lambda_z \) in turn leads to a strong downward trend in ‘other factor’ \( z_t \). To show the impact of this on Holston et al.’s (2017) estimate of the natural rate of interest, I initially report parameter estimates of the full Stage 3 model in Table 12, followed by plots of filtered estimates of the natural rate \( r^*_t \), trend growth \( g_t \), ‘other factor’ \( z_t \), and the output gap (cycle) variable \( \tilde{y}_t \) in Figure 7.58

Given estimates of the ratios \( \lambda_g = \sigma_g / \sigma_y \) and \( \lambda_z = a_r \sigma_z / \sigma_y \) from the previous two stages, the vector of Stage 3 parameters to be computed by MLE is:

\[
\theta_3 = \left[ a_{y,1}, a_{y,2}, a_r, b_\pi, b_y, \sigma_g, \sigma_\pi, \sigma_y \right]'.
\] (58)

In Table 12, estimates of \( \theta_3 \) are presented following the same format as in Table 4 and Table 9 previously. Since I also estimate \( \sigma_g \) and \( \sigma_z \) directly together with the other parameters by MLE without using the Stage 1 and Stage 2 estimates of \( \lambda_g \) and \( \lambda_z \), additional rows are inserted, with the values in brackets denoting implied estimates. The first two columns in Table 12 show estimates of \( \theta_3 \) obtained from running Holston et al.’s (2017) R-Code and my replication. The third and fourth columns (under headings ‘MLE(\( \sigma_g | \lambda_z^{HLW} \))’ and ‘MLE(\( \sigma_g | \lambda_z^{M0} \))’, respectively) report estimates when \( \sigma_g \) is estimated freely by MLE, while \( \lambda_z \) is held fixed at either \( \lambda_z^{HLW} = 0.030217 \) obtained from Holston et al.’s (2017) ‘misspecified’ Stage 2 model under their ‘time varying \( \phi \)’ approach, or at \( \lambda_z^{M0} = 0.000754 \) computed from the ‘correctly specified’ Stage 2 model in (42) with ‘constant \( \phi \)’. The last column of Table 12 under heading ‘MLE(\( \sigma_g, \sigma_z \))’ lists the estimates of \( \theta_3 \) when \( \sigma_g \) and \( \sigma_z \) are computed directly by MLE, with the implied values of \( \lambda_g \) and \( \lambda_z \) reported in brackets.

The Stage 3 results in Table 12 can be summarized as follows. The MLE of \( \sigma_g \) does not ‘pile-up’ at zero and is again approximately 50% larger than the estimate implied by the Stage 1 MUE of \( \lambda_g \). That is, \( \hat{\sigma}_g \approx 0.045 \) in the last three columns of Table 12, and thus very similar in size to the Stage 2 estimates of 0.044 and 0.045 shown in the last two columns of Table 9. The MLE of \( \sigma_z \) shrinks numerically to zero, while the estimates of the other parameters remain largely unchanged. Notice again that the log-likelihood values of the last three models in Table 12 are very similar, ie., between \(-514.8307\) and \(-514.2899\). Yet, the corresponding estimates of \( \sigma_z \) are either very small at 0 or comparatively large at 0.1371

\[\text{58Smoothed estimates are shown in Figure 8. In Section A.3 in the Appendix, the expansion of the system matrices are reported as for the earlier Stage 1 and Stage 2 models. These are in line with the full model reported in (1). As before, the state vector \( \xi_t \) is initialized using the same procedure as outlined in (12a) and Footnote 27, with the numerical values of \( \xi_{00} \) and \( P_{00} \) given in (A.35) and (A.36).}\]
when implied from the ‘misspecified’ Stage 2 model’s $\hat{\lambda}_2^{HLW}$ estimate. The $\delta_z$ coefficient from the ‘correctly specified’ Stage 2 model is 0.0037 and thereby nearly 40 times smaller than from the ‘misspecified’ Stage 2 model.

The findings from Table 12 are mirrored in the filtered estimates of $r_t^*, g_t, z_t$ and $\tilde{y}_t$ plotted in Figure 7. The ‘MLE($\sigma_g | \lambda_2^{M0}$)’ and ‘MLE($\sigma_g, \sigma_z$)’ estimates are visually indistinguishable. Unsurprisingly, out of the four estimates, ‘other factor’ $z_t$ is overall most strongly affected by the two different $\lambda_z$ values that are conditioned upon, showing either vary large variability and a pronounced downward trend in $z_t$, or being close to zero with very little variation (see panel (c) in Figure 7). The effect on the estimate of the natural rate is largest in the immediate aftermath of the global financial crisis, namely, from 2010 onwards. Interestingly, the output gap estimates shown in panel (d) of Figure 7 are quite similar, with the largest divergence occurring after 2012. The three trend growth estimates in panel (b) of Figure 7 where $\sigma_g$ is estimated directly by MLE are visually indistinguishable. Trend growth estimated from Holston et al.’s (2017) Stage 1 MUE of $\lambda_g$ is larger from 2009 to 2014. In comparison to the plots shown in panel (c) of Figure 2, the drop in all four trend growth estimates following the financial crisis seems exaggerated. The pure backward looking nature of the Kalman Filtered $g_t$ series exacerbates the effect of the decline in GDP during the financial crisis on the trend growth estimates after the crisis.59

5. Other issues

There are other issues with Holston et al.’s (2017) structural model in (1) that make it unsuitable for policy analysis. For instance, the policy rate $i_t$ is included as an exogenous variable. With $r_t^* = 4g_t + z_t$, and ‘other factor’ $z_t$ the ‘free’ variable due to $g_t$ being driven by GDP, $z_t$ effectively matches the ‘leftover’ movements in the interest rate to make it compatible with trend growth in the model. Since the central bank has full control over the policy rate, it can set $i_t$ to any desired level and the model will produce a natural rate through ‘other factor’ $z_t$ that will match it. Also, there is nothing in the structural model of (1) that makes the system stable. For the output gap relation in (1c) to be stationary, the real rate cycle $\tilde{r}_t = r_t - r_t^* = (i_t - \pi_e^t) - (4g_t + z_t)$ must be $I(0)$, yet there is no co-integrating relation imposed anywhere in the system to ensure that this holds in the model.60

A broader concern in the context of policy analysis is the fact that the filtered estimates of the state vector $\xi_t$ will be (weighted combinations of the) one-sided moving averages of the three observed variables that enter the state-space model; namely, $i_t, y_t$, and $\pi_t$.61 This Estimation results using updated data up to 2019:Q2 together with corresponding plots of filtered (and smoothed) estimates are reported in Table A.3, Figure A.5 and Figure A.6 in Section A.3 of the Appendix.

60 This insight is not new and has been discussed in, for instance, Pagan and Wickens (2019) (see pages 21 – 23). When trying to simulate from such a model, with $\pi_t$ being integrated of order 1, the simulated paths of the real rate $r_t = i_t - \pi_e^t$ can frequently diverge to very large values, even with samples of size $T = 229$ observations (the empirical sample size).

61 Smoothed estimates will be (weighted combinations of the) two-sided moving averages of the observables.
can be seen by writing out the Kalman Filtered estimate of the state vector as:

\[ \hat{\xi}_t|t = \hat{\xi}_t|t-1 + P_{t|t-1}H'(HP'_{t|t-1}H' + R)^{-1}(y_t - Ax_t - H\hat{\xi}_t|t-1) \]

\[ = \hat{\xi}_t|t-1 + G_t(y_t - Ax_t - H\hat{\xi}_t|t-1) \]

\[ = (I - G_tH)\hat{\xi}_t|t-1 + G_t(y_t - Ax_t) \]

\[ = (I - G_tH)\hat{\xi}_t|t-1 + G_t\tilde{y}_t \]

which is a (linear) recursion in \( \hat{\xi}_t|t \) and can thus be rewritten as:

\[ = \Psi_t\xi_0|0 + \sum_{i=0}^{t-1} \Psi_tG_{t-i}\tilde{y}_{t-i} \]

\[ = \Psi_t\xi_0|0 + \sum_{i=0}^{t-1} \omega_{ti}\tilde{y}_{t-i} \]

(59)

where \( \Psi_t = \prod_{n=0}^{t-1} \Phi_t, \forall i = 1, 2, \ldots, \Psi_0 = I, I \) is the identity matrix, \( \hat{\xi}_t|t-1 = \Phi\hat{\xi}_{t-1|t-1} \) is the predicted state vector, \( \xi_0|0 \) is the prior mean, \( P_{t|t-1} = FP_{t-1|t-1}F + Q \) is the predicted state variance, \( \omega_{ti} = \Psi_tG_{t-i} \) is a time varying weight matrix, and \( \tilde{y}_t \) is a \((2 \times 1)\) vector containing the observed variables \( y_t, \pi_t \), and \( i_t \).

This creates the following two issues. First, since the nominal interest rate \( i_t \) is directly controlled by the central bank and the natural rate \( r^*_t \) is constructed as a moving average of \( i_t \) (and the other observed variables), a circular relationship can be seen to evolve. Any central bank induced change in the policy rate \( i_t \) is mechanically transferred to the natural rate \( r^*_t \) via the Kalman Filtered estimate of the state vector \( \hat{\xi}_t|t \) in (59). This confounds the relationship between \( r^*_t \) and \( i_t \), making it impossible to address ‘causal’ policy questions of interest such as: “Is the natural rate low because \( i_t \) is low, or is \( i_t \) low because the natural rate is low?” as one

\[ \text{See also Durbin and Koopman (2012), who write to this on page 104: “It follows that these conditional means are weighted sums of past (filtering), of past and present (contemporaneous filtering) and of all (smoothing) observations. It is of interest to study these weights to gain a better understanding of the properties of the estimators as is argued in Koopman and Harvey (2003). ... In effect, the weights can be regarded as what are known as kernel functions in the field of nonparametric regression; ...”.} \]

\[ \text{I again use the notation of Hamilton (1994), see pages 394-395, with the matrices } A \text{ and } H \text{ however not transposed to be consistent with the earlier notation.} \]

\[ \text{To understand what is driving the downward trend in ‘other factor’ } z_t \text{ since the early 2000s in the model, one could examine the weight matrix } \omega_{ti} \text{ in (59) more closely to see how it interacts with the observable vector } \tilde{y}_t = y_t - Ax_t = [a(L)y_t - a_i(L)r_t; b_\pi(L)\pi_t - b_yy_t]', \text{ where } b_\pi(L) = 1 - b_\pi L - \frac{1}{2}(1 - b_\pi)(L^2 + L^3 + L^4) \text{ is the lag polynomial capturing the dynamics of inflation. Alternatively, the steady-state } P \text{ matrix could be computed recursively as in equation 13.5.3 in Hamilton (1994) to replace } P_{t|t-1} \text{ in the recursions for } \hat{\xi}_t|t. \text{ The relation in (59) would then yield } \hat{\xi}_t|t = \Phi\xi_0|0 + \sum_{i=1}^{t-1} \Phi^iG\tilde{y}_{t-i}, \text{ where } \Phi = (I - GH)F \text{ and } G = PH'(HP'H' + R)^{-1} \text{ would be the steady-state analogue to } \Phi_t \text{ and } G_t, \text{ with } P_{t|t-1} \text{ replaced by } P \text{ from the steady-state } P \text{ matrix.} \]
follows as a direct consequence from the other.

Second, because of the one-sided moving average nature of the Kalman Filtered estimates of the state vector, any outliers, structural breaks or otherwise ‘extreme’ observations at the beginning (or end) of the sample period can have a strong impact on these filtered estimates. For the (two-sided) Hodrick and Prescott (1997) filter, such problems (and other ones) are well known and have been discussed extensively in the literature before. Kalman Filter based (one-sided) estimates will also be affected. A simple way to appreciate this is by re-estimating the model using four different starting dates, while keeping the end of the sample period the same at 2019:Q2. In Figure 9 I show filtered estimates of $r_t^*, g_t, z_t$ and $\tilde{y}_t$ for the four starting dates 1967:Q1, 1972:Q1, 1952:Q2 and 1947:Q1 (smoothed estimates are shown in Figure 10), together with Holston et al.’s (2017) estimates using 1961:Q1 as the starting date.

Why did I choose these starting dates? The period following the April 1960 to February 1961 recession was marked by temporarily (and unusually) high GDP growth, yielding an annualized mean of 6.07% (median 6.47%), with a low standard deviation of 2.67% from 1961:Q2 to 1966:Q1 (see panel (b) of Figure 2). Having such ‘excessive’ growth at the beginning of the sample period has an unduly strong impact on the filtered (less so on the smoothed) estimate of trend growth $g_t$ in the model. Since both $g_t$ and $z_t$ enter the natural rate, this affects the estimate of $r_t^*$. To illustrate the sensitivity of these estimates to this time period, I estimate the model with data starting 6 years later in 1967:Q1. Also, Holston et al.’s (2017) Euro Area estimates of $r_t^*$ are negative from around 2013 onwards (see the bottom panel of Figure 3 on page S63 of their paper). To show that one can get the same negative estimates of $r_t^*$ for the U.S., I estimate the model with data starting in 1972:Q1 to match the sample period of the Euro Area. Lastly, I extend Holston et al.’s (2017) data back to 1947:Q1 to have estimates from a very long sample, using total PCE inflation prior to 1959:Q2 in place of Core PCE inflation and the Federal Reserve Bank of New York discount rate from 1965:Q1 back to 1947:Q1 as a proxy for the Federal Funds rate, as was done in Laubach and Williams (2003). Since inflation was rather volatile from 1947 to 1952, I also estimate the model with

---

64There exists a large literature on the HP filter and its problems (one of the more recent papers is by Hamilton (2018)), and it is not the goal to review or list them here. The paper by Phillips and Jin (2015) is, nonetheless, interesting to single out here, in particular the introduction section on pages 2 to 9, as it highlights the recent public debates by James Bullard, Paul Krugman, Tim Duy and others on the use (and misuse) of the HP filter in the construction of output gaps for policy analysis. Phillips and Jin (2015) show also that the HP filter fails to recover the underlying trend asymptotically in models with breaks (see section 4 in their paper), and they further propose alternative filtering/smoothing methods. In an earlier study, Schlicht (2008) describes how to deal with structural breaks and missing data. In the policy literature, one rarely sees problems such as structural breaks being addressed when using the HP filter.

65In all computations, I use Holston et al.’s (2017) R-Code and follow exactly their three stage procedure as before to estimate the factors of interest.

66This negative estimate in $r_t^*$ is driven by an excessively large and volatile estimate of ‘other factor’ $z_t$. Some commentators have attributed the larger decline in the natural rate to a stronger manifestation of ‘secular stagnation’ in the Euro Area than in the U.S.

67Note that from the quarterly CORE PCE data it will be possible to construct annualized inflation only from 1947:Q2 onwards. To have an inflation data point for 1947:Q1, annual core PCE data (BEA Series ID: DPC-CRG3A086NBEA) that extends back to 1929 was interpolated to a quarterly frequency and subsequently used
data beginning in 1952:Q2 to exclude this volatile inflation period from the sample.

Panel (a) in Figure 9 shows how sensitive the natural rate estimates to the different starting dates are, particularly at the beginning of Holston et al.’s (2017) sample, namely, from 1961 until about 1980, and at the end of the sample from 2009 onwards. Negative natural rate estimates are now also obtained for the U.S. when the sample starts in 1972:Q1 (or 1967:Q1). From panel (b) in Figure 9 it is evident that the filtered trend growth estimates are the primary driver of the excessive sensitivity in $r^*_t$ over the 1961 to 1980 period. For instance, in 1961:Q1, these estimates can be as high as 6 percent, or as low as 3 percent, depending on the starting date of the sample. Also, the differences in the estimates stay sizeable until 1972:Q1, before converging to more comparable magnitudes from approximately 1981 onwards. Apart from the estimate using the very long sample beginning in 1947:Q1 (see the blue line in panel (b) of Figure 9), the other four series remain surprisingly similar, even during and after the financial crisis period, that is, from mid 2007 to the end of the sample in 2019:Q2. Thus, the ‘front-end’ variability of the natural rate estimates are driven by the ‘front-end’ variability in the estimates of trend growth $g_t$.

In panel (b) of Figure 9, I add MUE and MMLE (smoothed) estimates of trend growth from Stock and Watson’s (1998) model in (4), and also estimates from the (correlated) UC model in (28) to provide long-sample benchmarks of trend growth from simple univariate models to Holston et al.’s (2017) estimates. These are the same as plotted earlier in panels (b) and (c) of Figure 2. To avoid cluttering the plot with additional lines, I do not plot the mean and median estimates computed over the more recent expansion periods as was done in Figure 2. Note, however, that the MUE estimate overlaps with the mean and median of GDP growth from 2009:Q3 onwards and can thus be used as a representative for these model free ‘average’ estimates of trend growth since the end of the financial crisis. Comparing the Kalman Filter based estimates from the various starting dates to the (smoothed) MUE, MMLE, and UC ones shows how different these are, particularly, from 2009:Q3 until the end of the sample. In the immediate post-crisis period, the (one-sided) filter based estimates are ‘pulled down’ excessively by the sharp decline in GDP and ‘converge’ only slowly at the very end of the sample period towards the three long-sample benchmarks. Trend growth is severely underestimated from 2009:Q3 onwards, and this affects the estimate of $r^*_t$.

In Figure A.7 in the Appendix, I show plots of (real) GDP growth and the recursively estimated mean of GDP growth over the post financial crisis period from 2009:Q3 to 2019:Q2. Trend growth stays rather stable between 2% and 3% over nearly the entire period, settling at around 2.25% in 2014:Q2 and remaining at that level. Moreover, it is never close to the filtered estimate of Holston et al. (2017) from 2009:Q3 to 2014:Q3. In Figure A.8, I plot the mean as well as median 10 year ahead annual-average (real) GDP growth forecasts from the Survey of Professional Forecasters (SPF) from 1992 to 2020 to have another benchmark to compute (annualized) quarterly inflation data for 1947:Q1. Since Holston et al.’s (2017) R-Code requires 4 quarters of GDP data prior to 1947:Q1 as initial values, annual GDP (BEA Series ID: GDPCA) was interpolated to quarterly data for the period 1946:Q1 1946:Q4.
series. These forecasts also remain fairly stable between 2% and 3% from 2008 until 2017, and drift only marginally lower towards the very end of the sample. Finally, in Figure A.9, Vanguard investor survey based 3 year and 10 year ahead expectations of (real) GDP growth from February 2017 to April 2020 are plotted. These are taken from Figure II on page 5 in Giglio et al. (2020). The 10 year expected growth rate shown in the right panel of Figure A.9 fluctuates (mainly) between 2.8% and 3.2%. All three plots suggest that, following the financial crisis, trend growth in GDP is unlikely to have dropped to the value reported in Holston et al. (2017).

Looking at the estimates of ‘other factor’ $z_t$ in panel (c) of Figure 9, we can see that it is the end of the sample, namely, from 2009:Q1 to 2019:Q2, that is most strongly affected by the different starting dates. In particular the two estimates that are based on the shorter samples starting in 1967:Q1 and 1972:Q1, which exclude the ‘excessive’ GDP growth period at the beginning of Holston et al.’s (2017) sample, generate substantially more negative $z_t$ estimates. For instance, in 2009:Q1, the 1972:Q1 based estimate is $-2.87$ while Holston et al.’s (2017) is $-1.22$. Also, the $z_t$ estimates from the shorter samples are well below $-2$ over nearly the entire 2014:Q4 to 2019:Q2 period. What is particularly surprising to observe here is how stable (and very close to zero) the estimates of $z_t$ from the four earlier sample starts are from 1947:Q1 until about 1971:Q3. Given the change in demographics and population growth, as well as factors related to savings and investment following the end of World War II, one would expect $z_t$ to reflect these changes. It is only from 1990:Q2 onwards that a decisive downward trend in $z_t$ becomes visible.

Holston et al. (2017) initialize the state vector for the $z_t$ elements of $\xi_t$ at zero. This choice leads to an anchoring effect and implies that the natural rate is driven solely by trend growth at the beginning of the sample. In the data, it acts as a normalisation. Since $z_t$ is specified to evolve as a driftless random walk, an initialisation at zero seems sensible from an econometric perspective. Nevertheless, if one is to view ‘other factor’ $z_t$ as a factor which is meant to capture underlying structural changes in demographics, saving and investment rates, and the likes, in the economy, it needs to be aligned with such trends in the empirical data of the economy that is modelled. That is, the normalization date needs to be justified from an economic perspective. Due to its large impact on the downward trend in the estimates of the natural rate, understanding how the zero initialisation affects the estimates and what exactly $z_t$ captures is crucial from a policy perspective.

Why does the estimate of $\sigma_z$ shrink towards zero, that is, from MLE and the ‘correctly

---

68 The data were downloaded from: https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/rgdp10 (accessed on the 27th of July, 2020).
69 There is also some variability beginning in the 70s until the 80s, but this variation seems to be largely due to the noisier nature of the filtered estimates and is not visible from the more efficient smoothed estimates shown in panel (c) of Figure 10. The differentiation here is not important. The point to take away from this discussion is that the period following the financial crisis yields very different estimates from the two shorter samples, irrespective of whether smoothed or filtered estimates are used in the construction of the natural rate.
70 This is even more pronounced in the smoothed estimates of $z_t$ shown in panel (c) of Figure 10.
Stage 2 model MUE? My conjecture is that the empirical data reject the $I(1)$ restriction on ‘other factor’ $z_t$. To test if the $z_t$ series indeed follows an $I(0)$ process, we can examine the time series properties of the difference between the observed GDP growth series $\Delta y_t$ and the real rate $r_t$, that is, the $(\Delta y_t - r_t)$ series. From Holston et al.’s (2017) structural model in (1) we know that:

$$r_t = r^*_t + \tilde{r}_t$$
$$= g_t + z_t + \tilde{r}_t$$
$$= g_{t-1} + \varepsilon^y_t + z_t + \tilde{r}_t,$$

(60)

where $\tilde{r}_t$ is the real rate gap defined earlier, which must be an $I(0)$ process for the output gap equation to be stationary. From the relations in (21) and (1d) we then obtain

$$\Delta y_t = g_{t-1} + \varepsilon^y_t + \Delta \tilde{y}_t,$$

(61)

so that the difference between real GDP growth $\Delta y_t$ in (61) and the real rate in (60) is:

$$(\Delta y_t - r_t) = (g_{t-1} + \varepsilon^y_t + \Delta \tilde{y}_t) - (g_{t-1} + \varepsilon^y_t + z_t + \tilde{r}_t)$$

$$= \Delta \tilde{y}_t - \tilde{r}_t + \varepsilon^y_t - \varepsilon^y_t - z_t,$$

(62)

where the variables in the first block on the right hand side of (62) will be the sum of two stationary ARMA terms (the two gaps) and two uncorrelated error terms ($\varepsilon^y_t$ and $\varepsilon^y_t$), while $z_t$ is $I(1)$. In the model we thus have that $(\Delta y_t - r_t)$ is $I(1)$.

In Figure 11 I show time series plots of $\Delta y_t$ with $r_t$ superimposed in Panel (a) in the top of the figure. The GDP growth minus real rate series $(\Delta y_t - r_t)$ is plotted in Panel (b), with the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of $(\Delta y_t - r_t)$ in Panel (c) in the bottom. Both, the time series plot in Panel (b) as well as the ACF and PACF plots in Panel (c) give the visual impression of a stationary $(\Delta y_t - r_t)$ series, with a fast decaying correlation structure. The first order autocorrelation coefficient is 0.576. More formal unit-root tests confirm that the null hypothesis of a unit-root in $(\Delta y_t - r_t)$ is strongly rejected.\(^{71}\) Either ‘other factor’ $z_t$ is stationary, or it does not appear in $(\Delta y_t - r_t)$ at all, which is only possible if $\sigma_z$ is zero.\(^{72}\) This supports the MLE’s and the correct Stage 2 model’s MUE results. From Panel (a) in Figure 11 it is also visible that the real rate remained well below real GDP growth for an extended period of time in the aftermath of the financial crisis; arguably, the longest stretch in the sample, with only the period following the Dotcom bubble showing some similarities.

A final comment on Holston et al.’s (2017) choice of reporting filtered (as opposed to

\(^{71}\)Augmented Dickey-Fuller and Elliott et al. (1996) DF-GLS $P_T$ and $t$-tests yield test statistics of $-5.33$, $0.58$, and $-5.30$, which are all substantially lower than their respective 1% critical values of $-3.46$, $1.92$, and $-2.58$.

\(^{72}\)Note again that $z_t$ is initialized at 0, so that a $\sigma_z = 0$ implies $z_t = 0$ for all $t$. 

Electronic copy available at: https://ssrn.com/abstract=3897838
smoothed) estimates of the latent states is in order. It is well known that the mean squared error (MSE) of the filtered states is in general larger than from the smoothed ones (see the discussion on page 151 in Harvey (1989)). The larger variability in the filtered states is visible from the estimates of $r^*_t$, $g_t$ and $z_t$, and less so from the output gap (cycle) estimates. Moreover, reporting filtered estimates precludes the use of a diffuse prior for the $I(1)$ variables in the state vector, since it creates extreme volatility in the filtered estimates of the states at the beginning of the sample period. This is not the case with the smoothed estimates. While it is sometimes stated that filtered states are ‘real time’ estimates, and are thus more relevant for policy analysis, one can see that this argument is invalid in this model. Not only are the parameter estimates of the model, ie., the $\theta_3$ in (58), based on full sample information, the GDP and PCE inflation data are not real time data, that is, data that were available to policy makers at time $t < T$. Reporting filtered (one-sided) estimates of the states as in Holston et al. (2017) (or on the FRBNY website) gives a misleading visual impression of the magnitude of the natural rate and trend growth.

6. Conclusion

Holston et al.’s (2017) implementation of Stock and Watson’s (1998) Median Unbiased Estimation (MUE) in Stages 1 and 2 of their procedure to estimate the natural rate of interest from a larger structural model is unsound. I show algebraically that their procedure cannot recover the ratios of interest $\lambda_g = \sigma_g / \sigma_{y^*}$ and $\lambda_z = a_z \sigma_z / \sigma_{\tilde{y}}$ needed for the estimation of the full structural model of interest. Holston et al.’s (2017) implementation of MUE in Stage 2 of their procedure is particularly problematic, because it is based on a misspecified model as well as an incorrect MUE procedure that spuriously amplifies their estimate of the signal-to-noise $\lambda_z$. This has a direct and consequential effect on the severity of the downward trending behaviour of ‘other factor’ $z_t$ and thereby the magnitude of the estimate of the natural rate.

Correcting their Stage 2 model and the implementation of MUE leads to a substantially smaller estimate of $\lambda_z$ of close to zero, and a substantial reduction in the downward trend of ‘other factor’ $z_t$ and the natural rate. The correction that is applied is quantitatively important. It shows that the estimate of $\lambda_z$ based on the correctly specified Stage 2 model is statistically highly insignificant. The resulting filtered estimates of $z_t$ are very close to zero for the entire sample period, highlighting the lack of evidence of ‘other factor’ $z_t$ being important for the determination of the natural rate in this model. Obtaining an accurate estimate of trend growth for the measurement of the natural rate is therefore imperative. Simple alternative estimates of trend growth that I construct suggest that the ones in Holston et al.’s (2017) are too low, particularly in the immediate aftermath of the global financial crisis.

I discuss various other issues with Holston et al.’s (2017) model of the natural rate that make it unsuitable for policy analysis. For instance, Holston et al.’s (2017) estimates are extremely sensitive to the starting date of the sample used to estimate the model. Using data beginning in 1972:Q1 (or 1967:Q1) leads to negative estimates of the natural rate, just
as with their Euro Area estimates. These negative estimates are again driven purely by
the exaggerated downward trending behaviour of ‘other factor’ $z_t$. Moreover, due to the
Kalman Filtered (or Smoothed) estimates of the state vector being a function of all observable
variables that enter into the model, with the central bank controlled nominal interest rate
being one of these, a confounding effect between $r^*_t$ and $i_t$ will arise, because any central
bank induced change in the policy rate $i_t$ is mechanically transferred to the natural rate.
This makes it impossible to answer ‘causal’ questions regarding the relationship between $r^*_t$
and $i_t$, as one responds to changes in the other. The entire set-up of the model is such that
‘other factor’ $z_t$ acts as a ‘plug-variable’ to ensure a zero mean in the output gap equation. In
this model, policy makers simply need to raise the policy rate for a higher natural rate to be
realized.
References

Andrews, Donald W. K. and Werner Ploberger (1994): “Optimal Tests when a Nuisance Parameter is Present only under the Alternative,” *Econometrica, 62*(6), 1383–1414.

Buncic, Daniel (2020): “On a standard Method for Measuring the Natural Rate of Interest,” SSRN Working Paper No. 3725151. Available from: http://ssrn.com/abstract=3725151.

Buncic, Daniel and Oliver Müller (2017): “Measuring the Output Gap in Switzerland with Linear Opinion Pools,” *Economic Modelling, 74*(August), 153–177.

Chow, Gregory C. (1960): “Tests of Equality between Sets of Coefficients in two Linear Regressions,” *Econometrica, 28*(3), 591–605.

Clark, Peter K. (1987): “The Cyclical Component of U.S. Economic Activity,” *Quarterly Journal of Economics, 102*(4), 797–814.

Durbin, James and Siam J. Koopman (2012): *Time Series Analysis by State Space Methods, 2nd Edition*, Oxford University Press.

Elliott, Graham, Thomas J. Rothenberg and James H. Stock (1996): “Efficient Tests for an Autoregressive Unit Root,” *Econometrica, 64*(4), 813–836.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel and Stephen Utkus (2020): “Inside the Mind of a Stock Market Crash,” NBER Working Paper No. 27272, National Bureau of Economic Research.

Grant, Angela L. and Joshua C. C. Chan (2017): “Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter,” *Journal of Economic Dynamics and Control, 75*, 114–121.

Hamilton, James D. (1994): *Time Series Analysis*, Princeton University Press.

——— (2018): “Why You Should Never Use the Hodrick-Prescott Filter,” *Review of Economics and Statistics, 100*(5), 831–843.

Harvey, Andrew C. (1985): “Trends and Cycles in Macroeconomic Time Series,” *Journal of Business and Economic Statistics, 3*(3), 216–227.

——— (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Hodrick, Robert and Edward C. Prescott (1997): “Post-war U.S. business cycles: A descriptive empirical investigation,” *Journal of Money, Credit, and Banking, 29*(1), 1–16.

Holston, Kathryn, Thomas Laubach and John C. Williams (2017): “Measuring the Natural Rate of Interest: International Trends and Determinants,” *Journal of International Economics, 108*(Supplement 1), S59–S75.

Kuttner, Kenneth N. (1994): “Estimating Potential Output as a Latent Variable,” *Journal of Business and Economic Statistics, 12*(3), 361–368.

Laubach, Thomas (2001): “Measuring The NAIRU: Evidence From Seven Economies,” *Review of Economics and Statistics, 83*(2), 218–231.

Electronic copy available at: https://ssrn.com/abstract=3897838
Laubach, Thomas and John C. Williams (2003): “Measuring the Natural Rate of Interest,” *Review of Economics and Statistics, 85*(4), 1063–1070.

Morley, James C., Charles R. Nelson and Eric Zivot (2003): “Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?” *Review of Economics and Statistics, 85*(2), 235–243.

Muth, John F. (1960): “Optimal properties of exponentially weighted forecasts,” *Journal of the American Statistical Association, 55*(290), 299–306.

Nyblom, Jukka (1989): “Testing for the Constancy of Parameters over Time,” *Journal of the American Statistical Association, 84*(405), 223–230.

Pagan, Adrian and Michael Wickens (2019): “Checking If the Straitjacket Fits,” *CEPR Discussion Paper No.: DP14140*.

Phillips, Peter C. B. and Sainan Jin (2015): “Business cycles, trend elimination, and the HP filter,” *Cowles Foundation Discussion Paper No. 2005, Yale University.* Available from: [http://cowles.yale.edu/sites/default/files/files/pub/d20/d2005.pdf](http://cowles.yale.edu/sites/default/files/files/pub/d20/d2005.pdf).

Quandt, Richard E. (1960): “Tests of the Hypothesis that a Linear Regression System obeys two Separate Regimes,” *Journal of the American Statistical Association, 55*(290), 324–330.

Schlicht, Ekkehart (2008): “Trend Extraction from Time Series with Structural Breaks and Missing Observations,” *Journal of the Japan Statistical Society, 38*(2), 285–292.

Shephard, Neil G. and Andrew C. Harvey (1990): “On the probability of estimating a deterministic component in the local level model,” *Journal of Time Series Analysis, 11*(4), 339–347.

Stock, James H. and Mark W. Watson (1998): “Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model,” *Journal of the American Statistical Association, 93*(441), 349–358.

——— (2011): *Introduction to Econometrics, 3rd Edition*, Addison-Wesley.
Figure 2: Inflation, interest rates, and GDP growth (annualized) from 1947:Q1 to 2019:Q2.
**Table 1:** Summary statistics of GDP growth over various sub-periods and expansion periods only

| Time period       | Mean   | Median | Stdev  | T     | Stderr | HAC-Stderr |
|-------------------|--------|--------|--------|-------|--------|------------|
| 1947:Q2 – 1981:Q3 | 3.5746 | 3.4501 | 4.6261 | 138   | 0.3938 | 0.4841     |
| 1983:Q1 – 2001:Q1 | 3.6419 | 3.6862 | 2.2956 | 73    | 0.2687 | 0.4016     |
| 1947:Q2 – 2001:Q1 | 3.4674 | 3.4809 | 4.0406 | 216   | 0.2749 | 0.3530     |
| 1947:Q2 – 1948:Q4 | 2.7952 | 2.2850 | 3.4072 | 7     | 1.2878 | 1.2530     |
| 1950:Q1 – 1953:Q2 | 7.3471 | 7.1194 | 4.9276 | 14    | 1.3169 | 1.7060     |
| 1954:Q3 – 1957:Q3 | 3.9390 | 3.9005 | 3.6537 | 13    | 1.0134 | 1.2985     |
| 1958:Q3 – 1960:Q2 | 5.3858 | 8.2510 | 4.7796 | 8     | 1.6898 | 1.7682     |
| 1961:Q2 – 1969:Q4 | 4.7809 | 4.3428 | 2.9850 | 35    | 0.5046 | 0.6104     |
| 1971:Q1 – 1973:Q4 | 4.9622 | 4.0550 | 3.8234 | 14    | 1.1037 | 0.9041     |
| 1975:Q2 – 1980:Q1 | 4.1795 | 2.9423 | 3.6637 | 20    | 0.8192 | 0.7484     |
| 1980:Q4 – 1981:Q3 | 4.2348 | 6.0763 | 4.9889 | 4     | 2.4944 | 1.9604     |
| 1983:Q1 – 1990:Q3 | 4.1711 | 3.8099 | 2.2194 | 31    | 0.3986 | 0.6300     |
| 1991:Q2 – 2001:Q1 | 3.5522 | 3.6291 | 1.8873 | 40    | 0.2984 | 0.3073     |
| 2002:Q1 – 2007:Q4 | 2.8546 | 2.4710 | 1.4995 | 24    | 0.3061 | 0.3324     |
| 2009:Q3 – 2019:Q2 | 2.2864 | 2.2528 | 1.4740 | 40    | 0.2331 | 0.1946     |

**Notes:** This table reports estimates of trend growth computed as the ‘average’ of annualized GDP growth computed over various sub-periods and expansion periods only. Columns 2 to 5 report means, medians, standard deviations (Stdev) and sample sizes (T) for the different sub-periods that are listed in column 1. The last two columns provide simple (Stderr) and HAC robust (HAC-Stderr) standard errors of the sample mean. The first three rows show time periods that include recession as well as expansion periods over which GDP growth was larger on average and/or more volatile than the last two rows (excluding the global financial crisis recession period). The ten rows in the middle provide summary statistics from 1947:Q2 to 2001:Q1 for expansion periods only.
### Table 2: Replicated results of Tables 4 and 5 in Stock and Watson (1998)

| Test | Statistic | \( p \)-value | \( \lambda \) | 90% CI | \( \sigma_{\Delta \beta} \) | 90% CI |
|------|-----------|---------------|-------------|--------|-----------------|--------|
| L    | 0.2094    | 0.2500        | 4.0559      | [0, 19.36] | 0.1303          | [0, 0.62] |
| MW   | 1.1588    | 0.2850        | 3.4335      | [0, 18.76] | 0.1103          | [0, 0.60] |
| EW   | 0.6821    | 0.3250        | 3.0712      | [0, 17.01] | 0.0987          | [0, 0.54] |
| QLR  | 3.3105    | 0.4800        | 0.7786      | [0, 13.26] | 0.0250          | [0, 0.42] |

| Parameter | MPLE | MMLE | MUE(0.13) | MUE(0.62) | SW.GAUSS |
|-----------|------|------|-----------|-----------|-----------|
| \( \sigma_{\Delta \beta} \) | 0    | 0.04440098 | 0.13 | 0.62 | 0.13 |
| \( \sigma_{\epsilon} \) | 3.85199480 | 3.85859423 | 3.84661923 | 3.78210658 | 3.84661917 |
| AR(1)     | 0.33708321 | 0.34025234 | 0.33501453 | 0.31544472 | 0.33501454 |
| AR(2)     | 0.12890328 | 0.13074607 | 0.12742313 | 0.12015642 | 0.12742309 |
| AR(3)     | −0.00917384 | −0.00725108 | −0.01017060 | −0.01488988 | −0.01017052 |
| AR(4)     | −0.08564442 | −0.08247862 | −0.08680297 | −0.09156066 | −0.08680298 |
| \( \beta_{00} \) | 1.79589936 | — | 2.44099926 | 2.67150007 | 2.44099940 |

Log-likelihood: \(-539.77274703\), \(-547.48046450\), \(-540.69267706\), \(-544.90718114\), \(-540.69267706\)

**Notes:** This table reports replication results that correspond to Tables 4 and 5 in Stock and Watson (1998) on page 354. The top part of the table shows the 4 different structural break test statistics together with their \( p \)-values in the first two columns, followed by the corresponding MUE estimates of \( \lambda \) with 90% CIs in square brackets. The last two columns show the implied \( \sigma_{\Delta \beta} \) estimate computed from \( T^{-1}\lambda \times \sigma_{\epsilon}/a(1) \) and 90% CIs in square brackets. The first two columns of the bottom part of the table report results from Maximum Likelihood based estimation, where MPLE estimates the initial value of the state vector \( \beta_{00} \), while MMLE uses a diffuse prior for the initial value of the state vector with mean zero and the variance set to \( 10^6 \). Columns under the heading MUE(0.13) and MUE(0.62) show Median Unbiased Estimates when \( \sigma_{\Delta \beta} \) is held fixed at 0.13, respectively, 0.62, which correspond to the estimate of \( \sigma_{\Delta \beta} \) when \( \lambda \) is computed using Nyblom’s (1989) \( L \) test (and its upper 90% CI). The last column under the heading SW.GAUSS lists the corresponding MUE(0.13) estimates obtained from running Stock and Watson’s (1998) GAUSS code. The row Log-likelihood displays the value of the log-likelihood at the reported parameter estimates. The Matlab file SW1998_MUE_replication.m replicates these results.
Figure 3: Smoothed trend growth estimates of US real GDP per capita.
Table 3: Broader replication results of Tables 4 and 5 in Stock and Watson (1998) using per capita real GDP data from the Federal Reserve Economic Data database (FRED2)

| Test   | Statistic | $p$-value | $\lambda$ | 90% CI     | $\sigma_{\Delta \beta}$ | 90% CI     |
|--------|-----------|-----------|-----------|------------|--------------------------|------------|
| L      | 0.0467    | 0.8950    | 0.0000    | [0, 4.099] | 0.0000                   | [0, 0.1092]|
| MW     | 0.2514    | 0.8900    | 0.0000    | [0, 4.296] | 0.0000                   | [0, 0.1145]|
| EW     | 0.1321    | 0.9000    | 0.0000    | [0, 3.910] | 0.0000                   | [0, 0.1042]|
| QLR    | 0.8834    | 0.9800    | 0.0000    | [0, 0.000] | 0.0000                   | [0, 0.0000]|

| Parameter | MPLE | MMLE | MUE($\sigma_{\Delta \beta}^L$) | MUE(CI – $\sigma_{\Delta \beta}^L$) |
|-----------|------|------|-------------------------------|-------------------------------------|
| $\sigma_{\Delta \beta}$ | 0    | 0    | 0                             | 0.10926099                          |
| $\sigma_{\epsilon}$       | 3.86603366 | 3.87619022 | 3.86603367          | 3.87574722                          |
| AR(1)  | 0.31646541 | 0.32120674 | 0.31646541          | 0.32138794                          |
| AR(2)  | 0.14652905 | 0.14903845 | 0.14652905          | 0.14924197                          |
| AR(3)  | −0.11122061 | −0.10873408 | −0.11122061         | −0.10846721                         |
| AR(4)  | −0.09512645 | −0.09050024 | −0.09512645         | −0.08983094                         |
| $\beta_{00}$  | 2.12011198 | 2.12011200 | 2.12011200          | 2.07784473                          |

Log-likelihood: $-540.49919714$  $-548.38308851$  $-540.49919714$  $-541.89394940$

Notes: This table reports replication results that correspond to Tables 4 and 5 in Stock and Watson (1998) on page 354, but now using real GDP per capita data (2012 chained dollars) obtained from the Federal Reserve Economic Data database (FRED2) with series ID: A939RX0Q048SBEA. The top part of the table shows the 4 different structural break test statistics together with their $p$-values in the first two columns, followed by the corresponding MUE estimates of $\lambda$ with 90% CIs in square brackets. The last two columns show the implied $\sigma_{\Delta \beta}$ estimate computed from $T^{-1} \lambda \times \sigma_{\epsilon}/a(1)$ and 90% CIs in square brackets. The first two columns of the bottom part of the table report results from Maximum Likelihood based estimation, where MPLE estimates the initial value of the state vector $\beta_{00}$, while MMLE uses a diffuse prior for the initial value of the state vector with mean zero and the variance set to $10^6$. Columns under the heading MUE($\sigma_{\Delta \beta}^L$) and MUE(CI – $\sigma_{\Delta \beta}^L$) show Median Unbiased Estimates when $\sigma_{\Delta \beta}$ is held fixed again at Nyblom’s (1989) $L$ test statistic based structural break estimate, respectively, when the upper 90% CI value is used. The row Log-likelihood displays the value of the log-likelihood at the reported parameter estimates. The sample period is the same as in Stock and Watson (1998), that is, from 1947:Q2 to 1995:Q4. The Matlab file estimate_per capita_trend_growth_v1.m replicates these results.
### Table 4: Stage 1 parameter estimates

| $\theta_1$ | HLW Prior | Diffuse Prior |
|-------------|------------|---------------|
|             | HLW.R-File | $b_y \geq 0.025$ | Alt.Init.Vals | $b_y \geq 0.025$ | $b_y$ Free | $b_y \geq 0.025$ | $b_y$ Free |
| $a_{y,1}$   | 1.517069   | 1.517069      | 1.557667      | 1.459697      | 1.646444 | 1.567830 |
| $a_{y,2}$   | -0.528804  | -0.528804     | -0.622443     | -0.463828     | -0.672732 | -0.577830 |
| $b_\pi$     | 0.712494   | 0.712494      | 0.669957      | 0.729089      | 0.717871 | 0.733063 |
| $b_y$       | 0.025000   | 0.025000      | 0.097185      | 0.005741      | 0.025000 | -0.000947 |
| $s$         | 0.776397   | 0.776396      | 0.743775      | 0.779472      | 0.709483 | 0.604555 |
| $\sigma_y$  | 0.534943   | 0.534943      | 0.405380      | 0.617190      | 0.406360 | 0.474077 |
| $\sigma_{\pi}$ | 0.807736 | 0.807736      | 0.790683      | 0.811801      | 0.809171 | 0.813024 |
| $\sigma_{y^*}$ | 0.511911 | 0.511910      | 0.617689      | 0.418977      | 0.574747 | 0.530756 |

Log-likelihood $-531.874714$ $-531.874714$ $-531.451446$ $-531.051066$ $-536.980336$ $-535.959612$

**Notes:** This table reports replication results for the Stage 1 model parameter vector $\theta_1$ of Holston et al. (2017). The table is split in two blocks. The left block (under the heading HLW Prior) reports estimation results of the Stage 1 model using the initialisation of Holston et al. (2017) for the state vector $\xi_t$, where $\xi_{00} = [806.45, 805.29, 804.12]$ and $P_{00}$ as defined in (12c). The right block (under the heading Diffuse Prior) uses a diffuse prior for $\xi_t$ with $P_{00} = 10^6 \times I_3$, where $I_3$ is a 3 dimensional identity matrix. In the left block, 4 sets of results are reported. The first column (HLW.R-File) reports estimates obtained by running Holston et al.’s (2017) R-Code for the Stage 1 model. The second column ($b_y \geq 0.025$) shows estimation results using Holston et al.’s (2017) initial values for parameter vector $\theta_1$ in the optimisation routine, together with the lower bound restriction $b_y \geq 0.025$. Footnote 32 describes how these initial values were found. The third column (Alt.Init.Vals) shows estimates when alternative initial values for $\theta_1$ are used, with the $b_y \geq 0.025$ restriction still in place. The fourth column ($b_y$ Free) reports estimates when the restriction on $b_y$ is removed. The right column block displays estimates of $\theta_1$ with and without the restriction on $b_y$ being imposed, but with a diffuse prior on the state vector. The last row (Log-likelihood) reports the value of the log-likelihood function at these parameter estimates. The Matlab file Stage1_replication.m computes these results.
| $\lambda_g$ | HLW Prior | | | | Diffuse Prior |
|---|---|---|---|---|
| | HLW.R-File | $b_y \geq 0.025$ | Alt.Init.Vals | $b_y$ Free | $b_y \geq 0.025$ | $b_y$ free |
| $L$ | — | 0.0732880 | 0.0941991 | 0.0328624 | 0.0475203 | 0 |
| MW | 0.0651806 | 0.0651807 | 0.0894533 | 0.0314654 | 0.0418274 | 0 |
| EW | 0.0538690 | 0.0538691 | 0.0806758 | 0.0253835 | 0.0423790 | 0 |
| QLR | 0.0493818 | 0.0493818 | 0.0792015 | 0.0194289 | 0.0411877 | 0 |

**Notes:** This table reports Stage 1 estimates of the ratio $\lambda_g = \sigma_g / \sigma_y$ which is equal to Stock and Watson’s (1998) MUE $\lambda/T$ for the various estimates of $\theta_1$ reported in Table 4 and the four different structural break tests. The table is split into left and right column blocks as in Table 4. Under the heading HLW.R-File, estimates of $\lambda_g$ obtained from running Holston et al.’s (2017) R-Code are reported for reference. These are computed for the MW, EW and QLR structural break tests only. The remaining columns report the replicated $\lambda_g$ from the various $\theta_1$ estimates from Table 4.
Table 6: Stage 1 MUE results of $\lambda_g$ after AR(1) filtering $\Delta \hat{y}^*_{i|T}$ as in Stock and Watson (1998)

| Test | Statistic | $p$-value | $\lambda$ | 90% CI      | $\lambda_g = \frac{\lambda}{T}$ |
|------|-----------|-----------|-----------|-------------|---------------------------------|
| $L$  | 2.2815    | 0.0050    | 20.3833   | [4.36, 80.00] | 0.0906                         |
| MW   | 15.3544   | 0.0050    | 20.5840   | [4.47, 80.00] | 0.0915                         |
| EW   | 8.4581    | 0.0050    | 15.9903   | [3.53, 52.81] | 0.0711                         |
| QLR  | 20.7596   | 0.0050    | 14.8164   | [3.14, 48.48] | 0.0659                         |

Notes: This table reports Stock and Watson’s (1998) MUE estimation results after the constructed $\Delta \hat{y}^*_{i|T}$ variable was AR(1) filtered to remove the serial correlation. The first two columns report the 4 different structural break test statistics together with the corresponding $p$-values, followed by the implied MUE estimates of $\lambda$ with 90% CIs in square brackets. The last column lists Holston et al.’s (2017) $\lambda_g = \frac{\lambda}{T}$ to facilitate the comparison to the results listed under column one in Table 5.
Table 7: MUE estimates of the transformed Stage 1 model using an AR(4) model for $u_t$

| Test  | Statistic | $p-$value | $\lambda$  | 90% CI       | $\sigma_g$ | 90% CI       |
|-------|-----------|-----------|-------------|--------------|-------------|--------------|
| L     | 0.3162    | 0.1200    | 5.914619    | [0, 23.95]   | 0.154213    | [0.0, 0.62]  |
| MW    | 1.7875    | 0.1450    | 5.650431    | [0, 23.88]   | 0.147325    | [0.0, 0.62]  |
| EW    | 1.0663    | 0.1800    | 4.883719    | [0, 20.97]   | 0.127335    | [0.0, 0.54]  |
| QLR   | 4.6029    | 0.2850    | 3.511961    | [0, 17.65]   | 0.091568    | [0.0, 0.46]  |

| Parameter | MPLE     | MMLE     | MUE($\lambda_{EW}$) | MUE($\lambda_{EW}^{Up}$) |
|-----------|----------|----------|----------------------|---------------------------|
| $\sigma_g$| 0        | 0.10621861| 0.12733451           | 0.54678211                |
| $\sigma_\varepsilon$| 2.99782490| 2.98030099| 2.97346405           | 2.90800215                |
| AR(1)     | 0.28603147| 0.27433173| 0.26988229           | 0.24291126                |
| AR(2)     | 0.16828174| 0.16079307| 0.15789805           | 0.14866124                |
| AR(3)     | -0.02046076| -0.02734562| -0.02996691          | -0.03106235               |
| AR(4)     | 0.06570210| 0.05750551| 0.05423838           | 0.06119692                |
| $g_{00}$  | 3.02198581| —        | 4.09740642           | 5.17204700                |

Notes: This table reports MUE estimation results of the transformed (expressed in local level model form) Stage 1 model, using an AR(4) process for $u_t$. The top part of the table shows the 4 different structural break test statistics together with their $p-$values in the first two columns, followed by the corresponding MUE estimates of $\lambda$ with 90% CIs in square brackets. The last two columns show the implied $\sigma_g$ estimate computed from $T^{-1} \sum \sigma_\varepsilon / a(1)$ and 90% CIs in square brackets. The first two columns of the bottom part of the table report results from Maximum Likelihood based estimation, where MPLE estimates the initial value of the state vector $g_{00}$, while MMLE uses a diffuse prior for the initial value of the state vector with mean zero and the variance set to $10^6$. Columns under the heading MUE($\lambda_{EW}$) and MUE($\lambda_{EW}^{Up}$) show Median Unbiased Estimates when $\sigma_\varepsilon$ is held fixed at its MUE point estimate and upper 90% CI, respectively, from the EW structural break test. The row Log-likelihood displays the value of the log-likelihood at the reported parameter estimates. The Matlab file Stage1_local_level_model_SW98_MUE_Clark_UC.m replicates these results.
| Parameter | Clark’s UC0 | Std.error | Clark’s UC | Std.error |
|-----------|-------------|-----------|------------|-----------|
| $a_{y,1}$ | 1.66886173  | 0.10948741| 1.29544818 | 0.23535955|
| $a_{y,2}$ | -0.72428051 | 0.11242749| -0.56748691| 0.21688350|
| $\sigma_{y^*}$ | 0.58984175 | 0.05842091| 1.15753826 | 0.22509014|
| $\sigma_{g}$  | 0.04632149 | 0.02276935| 0.03219018 | 0.02221788|
| $\sigma_{\tilde{y}}$ | 0.34626037 | 0.09727028| 0.80950722 | 0.36461143|
| Corr($\tilde{y}_t$, $y^*_t$) | 0 — | -0.94263135| 0.09714541|

Log-likelihood: -270.00071839 — — -269.87504061 —

**Notes:** This table reports parameter estimates of Clark’s (1987) UC model. Two sets of results are reported. In the left part of Table 8, parameter estimates and standard errors (Std. errors) from Clark’s UC0 model which assumes Corr($\tilde{y}_t$, $y^*_t$) = 0 are reported. In the right part, parameter estimates and standard errors for Clark’s correlated UC model are shown, where Corr($\tilde{y}_t$, $y^*_t$) is explicitly estimated. Standard errors are computed from the inverse of the Hessian matrix of the log-likelihood. I use a diffuse prior for the I(1) part of the state vector, with the variance set to $10^6$. The stationary part of the state vector is initialized at its unconditional mean and variance. I do not estimate the initial value of the state vector. This is analogous to MMLE in Stock and Watson (1998). The Matlab file Stage1_local_level_model_SW98_MUE_Clark_UC.m replicates these results.
Figure 4: Smoothed trend growth estimates from the modified Stage 1 model.
## Table 9: Stage 2 parameter estimates

| $\theta_2$ | HLW.R-File | Replicated | MLE($\sigma_g$) | MLE($\sigma_g$).$M_0$ |
|------------|------------|------------|-----------------|-----------------------|
| $a_{y,1}$  | 1.5139909  | 1.5139909  | 1.4735945       | 1.4947611             |
| $a_{y,2}$  | -0.5709339 | -0.5709339 | -0.5321668      | -0.5531451            |
| $a_r$      | -0.0736647 | -0.0736647 | -0.0831539      | -0.0755563            |
| $a_0$      | -0.2630694 | -0.2630694 | -0.2548597      | —                     |
| $a_g$      | 0.6078666  | 0.6078666  | 0.6277124       | —                     |
| $b_\pi$    | 0.6627428  | 0.6627428  | 0.6655286       | 0.6692919             |
| $b_y$      | 0.0844720  | 0.0844720  | 0.0819058       | 0.0802934             |
| $\sigma_g$ | 0.3582701  | 0.3582702  | 0.3636498       | 0.3742316             |
| $\sigma_\pi$ | 0.7872280 | 0.7872280  | 0.7881906       | 0.7895137             |
| $\sigma_y^*$ | 0.5665698 | 0.5665698  | 0.5534537       | 0.5526273             |
| $\sigma_g$ (implied) | (0.0305205) | (0.0305205) | 0.0437061       | 0.0448689             |
| $\lambda_g$ (implied) | 0.0538690 | 0.0538690  | (0.0789697)     | (0.0811920)           |

Notes: This table reports replication results for the Stage 2 model parameter vector $\theta_2$ of Holston et al. (2017). The first column (HLW.R-File) reports estimates obtained by running Holston et al.’s (2017) R-Code for the Stage 2 model. The second column (Replicated) shows the replicated results using the same set-up as in Holston et al.’s (2017). The third column (MLE($\sigma_g$)) reports estimates when $\sigma_g$ is freely estimated by MLE together with the other parameters of the Stage 2 model, rather than imposing the ratio $\lambda_g = \sigma_g/\sigma_y^* = 0.0538690378$ obtained from Stage 1. The last column (MLE($\sigma_g$).$M_0$) provides estimates of the “correctly specified” Stage 2 model in (42), with $\sigma_g$ again estimated directly by MLE. Values in round brackets give the implied $\sigma_g$ or $\lambda_g$ values when either $\lambda_g$ is fixed or when $\sigma_g$ is estimated. The last row (Log-likelihood) reports the value of the log-likelihood function at these parameter estimates. The Matlab file Stage2_replication.m replicates these results.
Figure 5: Sequence of $\{F(\tau)\}_{\tau=\tau_0}^{T_1}$ statistics on the dummy variable coefficients $\{\check{\xi}_1(\tau)\}_{\tau=\tau_0}^{T_1}$ used in the construction of the structural break test statistics.
| $\lambda_z$ | 'Time varying $\phi$' | 'Constant $\phi$' |
|-----------|-----------------------|-------------------|
|           | HLW.R-File    | Replicated  | [90% CI] | MLE($\sigma_g$) | [90% CI] | MLE($\sigma_g$).M0 | [90% CI] | Replicated | [90% CI] | MLE($\sigma_g$) | [90% CI] | MLE($\sigma_g$).M0 | [90% CI] |
| $L$       | 0.024969      | 0.024969     | [0, 0.11] | 0.032997      | [0, 0.14] | 0.008920           | [0, 0.07] | 0            | [0, 0.02] | 0            | [0, 0.00] | 0            | [0, 0.05] |
| MW        | 0.030217      | 0.030217     | [0, 0.11] | 0.033798      | [0, 0.12] | 0.007796           | [0, 0.06] | 0            | [0, 0.03] | 0            | [0, 0.02] | 0.000754       | [0, 0.06] |
| EW        | 0.034265      | 0.034265     | [0, 0.12] | 0.038942      | [0, 0.14] | 0.017199           | [0, 0.08] | 0            | [0, 0.05] | 0            | [0, 0.04] | 0.014703       | [0, 0.07] |

**Structural break test statistics ($p$-values in parenthesis)**

| $\lambda_z$ | 'Time varying $\phi$' | 'Constant $\phi$' |
|-----------|-----------------------|-------------------|
|           | HLW.R-File    | Replicated  | MLE($\sigma_g$) | MLE($\sigma_g$).M0 |
| $L$       | 0.050851      | 0.050851     | (0.8700) | (0.8700) | 0.108301 | (0.5450) |
| MW        | 1.870561      | 1.870561     | (0.1300) | 2.683140 | (0.0600) | 0.807469 | (0.4300) |
| EW        | 1.693014      | 1.693015     | (0.0800) | 2.120527 | (0.0450) | 0.506165 | (0.4300) |
| QLR       | 8.714461      | 8.714463     | (0.0450) | 10.303035 | (0.0250) | 4.751292 | (0.2700) |

**Notes:** This table reports the Stage 2 estimates of $\lambda_z$ for the different $\theta_2$ estimates corresponding to the "misspecified" and "correctly specified" Stage 2 models reported in Table 9. The table is split into two column blocks, showing the results for the 'Time varying $\phi$' and 'Constant $\phi$' scenarios in the left and right blocks, respectively. In the bottom half of the table, the four different structural break test statistics for the considered models are shown. The results under the heading 'HLW.R-File' show the $\lambda_z$ estimates obtained from running Holston et al.'s (2017) R-Code for the Stage 2 model as reference values. The second column 'Replicated' shows my replicated results. Under the heading 'MLE($\sigma_g$)', results for the "misspecified" Stage 2 model are shown with $\sigma_g$ estimated directly by MLE rather than from the first stage estimate of $\lambda_z$. Under the heading 'MLE($\sigma_g$).M0', results for the "correctly specified" Stage 2 model are reported where $\sigma_g$ is again estimated by MLE. The values in square brackets in the top half of the table report 90% confidence intervals for $\lambda_z$ computed from Stock and Watson's (1998) tabulated values provided in their GAUSS files. These were divided by sample size $T$ to make them comparable to $\lambda_z$. In the bottom panel, $p$-values of the various structural break tests are reported in round brackets. These were also extracted from Stock and Watson's (1998) GAUSS files.
Table 11: Summary statistics of the $\lambda_z$ estimates obtained from applying Holston et al.’s (2017) Stage 2 MUE procedure to simulated data

| Summary Statistics | DGPs when $\theta_2$ held fixed at $\hat{\theta}_2$ | DGPs when $\theta_2$ is re-estimated |
|--------------------|-----------------------------------------------|-------------------------------------|
|                    | $r_t^* = 4g_t$                                | $r_t^* = 4g_t + z_t$                |
| Minimum            | 0                                             | 0                                   |
| Maximum            | 0.101220                                      | 0.116886                            |
| Standard deviation | 0.016245                                      | 0.018512                            |
| Mean               | 0.028842                                      | 0.025103                            |
| Median             | 0.028394                                      | 0.022215                            |
| $\Pr(\hat{\lambda}_z > 0.030217)$ | 0.457000                                      | 0.339000                            |

Notes: This table reports summary statistics of the $\lambda_z$ estimates that one obtains from implementing Holston et al.’s (2017) Stage 2 MUE procedure on artificial data that was simulated from two different data generating processes (DGPs). The first DGP simulates data from the full structural model in (1) under the parameter estimates of Holston et al. (2017), but where the natural rate is determined solely by trend growth. That is, in the output gap equation in (1c), $r_t^* = 4g_t$. The second DGP simulates data from the full model of Holston et al. (2017) where $r_t^* = 4g_t + z_t$. The summary statistics that are reported are the minimum, maximum, standard deviation, mean, median, as well as the empirical frequency of observing a value larger than the estimate of 0.030217 obtained by Holston et al. (2017), denoted by $\Pr(\hat{\lambda}_z > 0.030217)$. The table shows four different estimates, grouped in 2 block pairs. The left block under the heading ‘DGPs when $\theta_2$ is held fixed’ shows the simulation results for the two DGPs when the Stage 2 parameter vector $\theta_2$ is held fixed at the Stage 2 estimates and is not re-estimated on the simulated data. The right block under the heading ‘DGPs when $\theta_2$ is re-estimated’ shows the simulation results when $\theta_2$ is re-estimated for each simulated series. Simulations are performed on a sample size equivalent to the empirical data, with 1000 repetitions.
Figure 6: Histograms of the estimated $\{\hat{\lambda}_s\}_{s=1}^{S}$ sequence corresponding to the summary statistics shown in Table 11. On the left and right columns, histograms for the two different DGPs are shown. To top two histograms show the results when $\theta_2$ is held fixed in the simulations and is not re-estimated, while the bottom plots show the results when $\theta_2$ is re-estimated on each simulated series that is generated.
### Table 12: Stage 3 parameter estimates

| θ_3    | HLW.R-File | Replicated | MLE(σ_g|λ_HLW^z) | MLE(σ_g|λ_M^z) | MLE(σ_g, σ_z) |
|--------|------------|------------|----------------|----------------|---------------|
| a_y,1  | 1.52957249 | 1.52957247 | 1.49442462     | 1.49566712     | 1.49566147    |
| a_y,2  | −0.58756415| −0.58756414| −0.55370268    | −0.55448942    | −0.55448212   |
| a_r    | −0.07119569| −0.07119569| −0.07941598    | −0.07525496    | −0.07525240   |
| b_π    | 0.66820705 | 0.66820705 | 0.67128197     | 0.66919468     | 0.66919993    |
| b_y    | 0.07895778 | 0.07895778 | 0.07593604     | 0.08054901     | 0.08054716    |
| σ_y    | 0.35346845 | 0.35346847 | 0.36043114     | 0.37381376     | 0.37382935    |
| σ_π    | 0.78919487 | 0.78919487 | 0.79029982     | 0.78948921     | 0.78949094    |
| σ_y*   | 0.57241925 | 0.57241924 | 0.55915743     | 0.55293818     | 0.55293018    |
| σ_g (implied) | (0.03083567) | (0.03083567) | 0.04583852 | 0.04497450 | 0.04497414 |
| σ_z (implied) | (0.15002080) | (0.15002080) | (0.13714150) | (0.00374682) | 0.00000000 |
| λ_g (implied) | 0.05386904 | 0.05386904 | (0.08197784) | (0.08133730) | (0.08133782) |
| λ_z (implied) | 0.03021722 | 0.03021722 | 0.03021722     | 0.00075430     | (0.00000000) |

**Log-likelihood** -515.14470528 -515.14470599 -514.83070544 -514.28987426 -514.28958969

**Notes:** This table reports replication results for the Stage 3 model parameter vector θ_3 of Holston et al. (2017). The first column (HLW.R-File) reports estimates obtained by running Holston et al.’s (2017) R-Code for the Stage 3 model. The second column (Replicated) shows the replicated results using the same set-up as in Holston et al.’s (2017). The third column (MLE(σ_g|λ_HLW^z)) reports estimates when σ_g is directly estimated by MLE together with the other parameters of the Stage 3 model, while λ_z is held fixed at λ_HLW^z = 0.030217 obtained from Holston et al.’s (2017) “misspecified” Stage 2 procedure. In the forth column (MLE(σ_g|λ_M^z)), σ_g is again estimated directly by MLE together with the other parameters of the Stage 3 model, but with λ_z now fixed at λ_M^z = 0.000754 obtained from the “correctly specified” Stage 2 model in (42). The last column (MLE(σ_g, σ_z)) shows estimates when all parameters are computed by MLE. Values in round brackets give the implied {σ_g, σ_z} or {λ_g, λ_z} values when either is fixed or estimated. The last row (Log-likelihood) reports the value of the log-likelihood function at these parameter estimates. The Matlab file Stage3_replication.m replicates these results.
Figure 7: Filtered estimates of the natural rate $r_t^*$, annualized trend growth $g_t$, 'other factor' $z_t$, and the output gap (cycle) variable $\tilde{y}_t$.
Figure 8: Smoothed estimates of the natural rate $r_t^*$, annualized trend growth $g_t$, 'other factor' $z_t$, and the output gap (cycle) variable $\tilde{y}_t$. Electronic copy available at: https://ssrn.com/abstract=3897838
Figure 9: Filtered estimates of annualized trend growth $g_t$, ‘other factor’ $z_t$ and the natural rate $r_t^\ast$ based on different starting dates
Figure 10: Smoothed estimates of annualized trend growth $g_t$, ‘other factor’ $z_t$ and the natural rate $r_t^*$ based on different starting dates.

Electronic copy available at: https://ssrn.com/abstract=3897838
Figure 11: Time series and autocorrelation plots of real GDP growth and the real interest rate
Appendix

This appendix provides additional information on the Holston et al. (2017) model, their estimation procedure as well as snippets of R-Code. Matrix details regarding the three stages of their procedure are taken from the file HLW_Code_Guide.pdf which is contained in the HLW_Code.zip file available from John Williams’ website at the Federal Reserve Bank of New York: https://www.newyorkfed.org/medialibrary/media/research/economists/williams/data/HLWCode.zip.

The state-space model notation is:

\[ y_t = A \xi_t + H \xi_{t-1} + \nu_t, \quad \text{where} \quad \begin{bmatrix} \nu_t \\ \xi_t \end{bmatrix} \sim \text{MNorm} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & W \end{bmatrix} \right), \quad (A.1) \]

where \( \xi_t = \epsilon_t \), so that \( \text{Var}(S \epsilon_t) = \text{Var}(\epsilon_t) = S W S' = Q \), with \( \epsilon_t \) and \( Q \) being the notation used in the online appendix of Holston et al. (2017) for the state vector’s disturbance term and its variance-covariance matrix.

### A.1. Stage 1 Model

The first Stage model is defined by the following system matrices:

\[
\begin{align*}
    y_t &= \begin{bmatrix} y_t \\ \pi_t \end{bmatrix}' , \\
    x_t &= \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \pi_{t-1} \\ \pi_{t-2} \end{bmatrix}' , \\
    \xi_t &= \begin{bmatrix} y^*_t \\ y^*_{t-1} \end{bmatrix}' , \\
    A &= \begin{bmatrix} a_{y,1} & a_{y,2} & 0 & 0 \\ b_y & 0 & b_{\pi} & (1 - b_{\pi}) \end{bmatrix} , \\
    H &= \begin{bmatrix} 1 & -a_{y,1} & -a_{y,2} \\ 0 & -b_y & 0 \end{bmatrix} , \\
    F &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} , \\
    S &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} .
\end{align*}
\]

From this, the measurement relations are:

\[
\begin{align*}
    y_t = Ax_t + H \xi_{t-1} + \nu_t \\
    \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} a_{y,1} & a_{y,2} & 0 & 0 \\ b_y & 0 & b_{\pi} & (1 - b_{\pi}) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \pi_{t-1} \\ \pi_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & -a_{y,1} & -a_{y,2} \\ 0 & -b_y & 0 \end{bmatrix} \begin{bmatrix} y^*_t \\ y^*_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^\pi \end{bmatrix} .
\end{align*}
\]

with the corresponding state equations being:

\[
\begin{align*}
    \xi_t = F \xi_{t-1} + S \epsilon_t \\
    \begin{bmatrix} y^*_t \\ y^*_{t-1} \\ y^*_{t-2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y^*_t \\ y^*_{t-1} \\ y^*_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^\pi \end{bmatrix} .
\end{align*}
\]
Expanding (A.3) and (A.4) yields:

\[ y_t = y^*_t + a_{y,1}(y_{t-1} - y^*_{t-1}) + a_{y,2}(y_{t-2} - y^*_{t-2}) + \varepsilon^y_t \]

\[ \pi_t = b_y(y_{t-1} - y^*_{t-1}) + b_\pi \pi_{t-1} + (1 - b_\pi) \pi_{t-24} + \varepsilon^\pi_t \]

and

\[ y_t' = y^*_{t-1} + \varepsilon^y_t \]
\[ y^*_t = y^*_{t-1} \]
\[ y^*_{t-2} = y^*_{t-2} \]

respectively, for the measurement and state equations. Defining output \( y_t \) as trend plus cycle, and ignoring the identities, yields then the following relations for the Stage 1 model:

\[ y_t = y^*_t + \tilde{y}_t \quad \text{(A.6a)} \]

\[ \pi_t = b_\pi \pi_{t-1} + (1 - b_\pi) \pi_{t-24} + b_y \tilde{y}_{t-1} + \varepsilon^\pi_t \quad \text{(A.6b)} \]

\[ \tilde{y}_t = a_{y,1} \tilde{y}_{t-1} + a_{y,2} \tilde{y}_{t-2} + \varepsilon^\pi_t \quad \text{(A.6c)} \]

\[ y^*_t = y^*_{t-1} + \varepsilon^y_t \quad \text{(A.6d)} \]

If we disregard the inflation equation (A.6b) for now, the decomposition of output into trend and cycle can be recognized as the standard Unobserved Component (UC) model of Harvey (1985), Clark (1987), Kuttner (1994), Morley et al. (2003) and others. Holston et al. (2017) write on page S64: ". . . we follow Kuttner (1994) and apply the Kalman filter to estimate the natural rate of output, omitting the real rate gap term from Eq. (4) [our Equation (A.6c)] and assuming that the trend growth rate, g, is constant."

One key difference is, nevertheless, that no drift term is included in the trend specification in (A.6d), so that \( y^*_t \) follows a random walk without drift. Evidently, this cannot match the upward sloping pattern in the GDP series. The way that Holston et al. (2017) deal with this mismatch is by ‘detrending’ output \( y_t \) in the estimation. This is implemented by re-placing \( \{y_{t-j}\}_{j=0}^2 \) in \( y_t \) and \( x_t \) in (A.2) by \( (y_t - gt) \), where \( g \) is a parameter (and not a trend growth state variable) to be estimated, and \( t \) is a linear time trend defined as \( t = [1, \ldots, T]' \). This is hidden away from the reader and is not described in the documentation in either text or equation form. Only from the listing of the vector of parameters to be estimated by MLE, referred to as \( \theta_1 \) in the middle of page 10 in the documentation, does it become evident that an additional parameter — confusingly labelled as \( g \) — is included in the estimation. That is, the vector of Stage 1 parameters to be estimated is defined as:

\[ \theta_1 = [a_{y,1}, a_{y,2}, b_\pi, b_y, g, \sigma_g, \sigma_\pi, \sigma_y]' \quad \text{(A.7)} \]

Note that the parameter \( g \) in \( \theta_1 \) is not found in any of the system matrices that describe the Stage 1 model on page 10 of the documentation. This gives the impression that it is a typographical error in the documentation, rather than a parameter that is added to the model in the estimation. However, from their R-Code file unpack.parameters.stage1.R, which is reproduced in R-Code 3, one can see that part of the unpacking routine, which is later called by the log-likelihood estimation function,
‘detrends’ the data (see the highlighted lines 29 to 31 in R-Code 3, where * parameter [5] refers to parameter g in θ₁). Due to the linear time trend removal in the estimation stage, it has to be added back to the Kalman Filter and Smoother extracted trends \( y^*_t \), which is is done in \texttt{kalman.states.wrapper.R} (see the highlighted lines 29 to 30 in R-Code 4, where the if statement: \texttt{if (stage == 1) \{ \} \}} on line 28 of this file ensures that this is only done for the Stage 1 model). The actual equation for the trend term \( y^*_t \) is thus:

\[
y^*_t = g + y^*_{t-1} + \varepsilon^y_t
\]

\[
y^*_0 + gt + \sum_{s=1}^{t} \varepsilon^y_s,
\]

where \( g \) is an intercept term that captures constant trend growth, and \( y^*_0 \) is the initial condition of the state vector set to 806.45 from the HP filter output as discussed in Footnote 27. Why Holston et al. (2017) prefer to use this way of dealing with the drift term rather than simply adding an intercept term to the state equation in (A.4) is not clear, and not discussed anywhere.

In the estimation of the Stage 1 model, the state vector \( \xi_t \) is initialized using the same procedure as outlined in (12a) and Footnote 27 with the numerical value of \( \xi_{00} \) and \( P_{00} \) set at:

\[
\xi_{00} = [806.4455, 805.2851, 804.1248]
\]

\[
P_{00} = \begin{bmatrix}
0.4711 & 0.2 & 0.0 \\
0.2 & 0.2 & 0.0 \\
0.0 & 0.0 & 0.2
\end{bmatrix}
\]

A.2. Stage 2 Model

The second Stage model of Holston et al. (2017) is defined by the following model matrices:

\[
y_t = [y_t, \pi_t]^\prime
\]

\[
x_t = [y_{t-1}, y_{t-2}, r_{t-1}, r_{t-2}, \pi_{t-1}, \pi_{t-2,4}, 1]^\prime
\]

\[
\xi_t = [y^*_t, y^*_{t-1,}, y^*_{t-2}, \pi_{t-1}]^\prime
\]

\[
A = \begin{bmatrix}
a_{y,1} & a_{y,2} & a_g & 0 & 0 & a_0 \\
b_y & 0 & 0 & 0 & b_\pi & (1 - b_\pi) & 0
\end{bmatrix}, \quad H = \begin{bmatrix}
1 & -a_{y,1} & -a_{y,2} & a_g \\
0 & -b_y & 0 & 0
\end{bmatrix},
\]

\[
F = \begin{bmatrix}
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}, \quad S = \begin{bmatrix}
1 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 1
\end{bmatrix}
\]

The measurement and state relations are given by:

\[
y_t = Ax_t + H\xi_t + \nu_t
\]
\[
\begin{bmatrix}
y_t \\
p_t
\end{bmatrix} = \begin{bmatrix}
a_{y1} & a_{y2} & \frac{a}{2} & \frac{a}{2} & 0 & 0 & a_0 \\
b_y & 0 & 0 & 0 & b_{\pi} & (1 - b_{\pi}) & 0
\end{bmatrix} \begin{bmatrix}
y_{t-1} \\
y_{t-2} \\
r_{t-1} \\
r_{t-2} \\
p_{t-1} \\
p_{t-2} \\
p_{t-2,4}
\end{bmatrix} + \begin{bmatrix}
1 - a_{y1} & -a_{y2} & a_g \\
0 & -b_y & 0 & 0
\end{bmatrix} \begin{bmatrix}
y_t^* \\
y_{t-1}^* \\
y_{t-2}^* \\
g_{t-1}^*
\end{bmatrix} + \begin{bmatrix}
\epsilon_t^y \\
\epsilon_{t-1}^\pi
\end{bmatrix}
\]  
(A.15)

and

\[
\xi_t = F \xi_{t-1} + S \epsilon_t
\]
\[
\begin{bmatrix}
y_t^* \\
y_{t-1}^* \\
y_{t-2}^* \\
g_{t-1}^*
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix} \begin{bmatrix}
y_{t-1}^* \\
y_{t-2}^* \\
y_{t-3}^* \\
g_{t-2}^*
\end{bmatrix} + \begin{bmatrix}
1 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 1
\end{bmatrix} \begin{bmatrix}
\epsilon_t^y \\
\epsilon_t^g \\
\epsilon_{t-1}^y \\
\epsilon_{t-1}^g
\end{bmatrix}.
\]  
(A.16)

Note that $\sigma_\gamma^2$ in $\text{Var}(\epsilon_t) = W = \text{diag}(\sigma_{y^*}^2, \sigma_\gamma^2)$ is replaced by $(\hat{\lambda}_\gamma \sigma_{y^*})^2$ where $\hat{\lambda}_\gamma$ is the estimate from the first Stage, so that we obtain:

\[
\text{Var}(S \epsilon_t) = S W S' = \begin{bmatrix}
1 & 0 \\
0 & \sigma_{y^*}^2 \\
0 & 0 \\
0 & 0
\end{bmatrix} \begin{bmatrix}
\sigma_{y^*}^2 & 0 & 0 & 0 \\
0 & (\hat{\lambda}_\gamma \sigma_{y^*})^2 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & (\hat{\lambda}_\gamma \sigma_{y^*})^2
\end{bmatrix} \begin{bmatrix}
1 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 1
\end{bmatrix}
\]  
(A.17)

which is then used in the Kalman Filter routine and ML to estimate the Stage 2 model parameters.

Expanding the relations in (A.15) and (A.16) leads to the measurement:

\[
y_t = y_t^* + a_{y1}(y_{t-1} - y_{t-1}^*) + a_{y2}(y_{t-2} - y_{t-2}^*) + \frac{a}{2}(r_{t-1} + r_{t-2}) + a_0 + a_g g_{t-1} + \epsilon_t^y
\]  
(A.18a)

\[
p_t = b_y(y_{t-1} - y_{t-1}^*) + b_\pi p_{t-1} + (1 - b_{\pi}) p_{t-2,4} + \epsilon_t^\pi
\]  
(A.18b)

and corresponding state relations

\[
y_t^* = y_{t-1}^* + g_{t-2} + \epsilon_t^{y^*}
\]  
(A.19a)

\[
y_{t-1}^* = y_{t-1}^*
\]  
(A.19b)

\[
y_{t-2}^* = y_{t-2}^*
\]  
(A.19c)

\[
g_{t-1} = g_{t-2} + \epsilon_{t-1}^g.
\]  
(A.19d)
Defining output $y_t$ as before as trend plus cycle, dropping identities, simplifying and rewriting gives the following Stage 2 system relations:

\begin{align}
    y_t &= y^*_t + \tilde{y}_t \quad (A.20a) \\
    \pi_t &= b_\pi \pi_{t-1} + (1 - b_\pi) \pi_{t-2,4} + b_y \tilde{y}_{t-1} + \varepsilon^\pi_t \quad (A.20b) \\
    a_y(L) \tilde{y}_t &= a_0 + \frac{a_2}{2} (r_{t-1} + r_{t-2}) + a_2 \tilde{g}_{t-1} + \varepsilon^\theta_t \quad (A.20c) \\
    y^*_t &= y^*_{t-1} + \tilde{g}_{t-2} + \varepsilon^\gamma_t \quad (A.20d) \\
    \tilde{g}_{t-1} &= \tilde{g}_{t-2} + \varepsilon^\delta_{t-1} \quad (A.20e)
\end{align}

where the corresponding vector of parameters to be estimated by MLE is:

$$
\theta_2 = [a_{y,1}, a_{y,2}, a_r, a_0, a_g, b_\pi, b_y, \sigma_g, \sigma_\pi, \sigma_y]^\prime. \quad (A.21)
$$

The state vector $\xi_t$ in the estimation of the Stage 2 model is initialized using the procedure outlined in (12a) and Footnote 27, with the numerical value of $\xi_{00}$ and $P_{00}$ set at:

\begin{align}
    \xi_{00} &= [806.4455, 805.2851, 804.1248, 1.1604] \quad (A.22) \\
    P_{00} &= \begin{bmatrix}
        0.7185 & 0.2 & 0.0 & 0.2 \\
        0.2 & 0.2 & 0.0 & 0.0 \\
        0.2 & 0.2 & 0.0 & 0.0 \\
        0.2 & 0.0 & 0.2 & 0.2009
    \end{bmatrix}. \quad (A.23)
\end{align}

Notice from the trend specification in (A.20d) that $g_{t-2}$ instead of $g_{t-1}$ is included in the equation. This is not a typographical error, but rather a ‘feature’ of the Stage 2 model specification of Holston et al. (2017), and is not obvious until the Stage 2 model relations are written out as above in equations (A.15) to (A.20). I use the selection matrix $S$ to derive what the variance-covariance matrix of $S\varepsilon_t$, that is, $\text{Var}(S\varepsilon_t) = \text{Var}(\varepsilon_t) = S\gamma S' = \gamma$, should look like. Holston et al. (2017) only report the $\gamma$ matrix in their online appendix included in the R-Code zip file (see page 10, lower half of the page in Section 7.4).

In the Stage 3 model, Holston et al. (2017) use a ‘trick’ to arrive at the correct trend specification for $y^*_t$ by including both, the $\varepsilon^\delta_{t-1}$ as well as the $\varepsilon^\gamma_t$ error terms in the equation for $y^*_t$ (see (A.32) below). This can also be seen from the $\gamma$ matrix on page 11 in Section 7.5 of their online appendix or (A.31) below, which now includes off-diagonal terms in the Stage 3 model.

### A.2.1. Getting the correct Stage 2 Model from the Stage 3 Model

We can apply this same ‘trick’ for the Stage 2 model, by taking the Stage 3 model state-space form and deleting the row, respectively, column entries of the $F$, $H$, and $S$ matrices to make them conformable with the required Stage 2 model. The state and measurement equations of the correct Stage 2 model then look as follows:

$$
y_t = A\xi_t + H\xi_t + \nu_t
$$
which, upon expanding and dropping of identities, yields:

\[
\begin{bmatrix}
\epsilon_t \\
y_t
\end{bmatrix} = \begin{bmatrix}
y_{t-1} \\
y_{t-2} \\
\pi_{t-1} \\
\pi_{t-2,4}
\end{bmatrix} + \begin{bmatrix}
1 - a_y,1 & -a_y,2 \\
0 & 0
\end{bmatrix} \begin{bmatrix}
\epsilon_t \\
y_{t-1} \mu_t \\
y_{t-2} \mu_t \\
\pi_{t-1} \mu_t \\
\pi_{t-2,4} \mu_t
\end{bmatrix} + \begin{bmatrix}
1 \\
y_{t-1} \mu_t \\
y_{t-2} \mu_t \\
\pi_{t-1} \mu_t \\
\pi_{t-2,4} \mu_t
\end{bmatrix} + \begin{bmatrix}
\epsilon_t \gamma \\
\epsilon_t \gamma
\end{bmatrix}
\]

(A.24)

\[
\begin{bmatrix}
\pi_t,1 \\
\pi_t,2 \\
\pi_t,3 \\
\pi_t,4
\end{bmatrix} = b_\pi \pi_{t-1} + (1 - b_\pi) \pi_{t-2,4} + b_\pi \hat{\pi}_{t-1} + \epsilon_t \gamma
\]

These last relations correspond to (42), with \( \epsilon_t \gamma \) being the counterpart to \( \epsilon_t \gamma = -a_r(L)z_t + \epsilon_t \gamma \) if we take the full Stage 3 model as the true model.

Using the Stage 3 state-space form and simply adjusting it as shown above yields the correct Stage 2 equations for trend \( y_t \) and the output gap \( \hat{\gamma} \). With this form of the state-space model, it is also clear that the variance-covariance matrix \( Q = \text{Var}(S\epsilon_t) \) will be:

\[
Q = SWS'
\]

(A.25)
where \((\hat{\lambda}_\sigma \sigma_y)\)^2 again replaces \(\sigma^2\), as before. Since the \(Q\) matrix in Holston et al. (2017) takes the form of (A.17) and not (A.25), we can see that this ‘trick’ of rewriting the trend growth equation as in the Stage 3 model specification was not applied to the Stage 2 model. Given that the correct Stage 2 model is easily obtained from the full Stage 3 model specification, it is not clear why the Stage 2 model is defined incorrectly as in (36).

A.3. Stage 3 Model

The third and final Stage model is defined as follows:

\[
y_t = \begin{bmatrix} y_t, \pi_t \end{bmatrix}' \tag{A.26}
\]
\[
x_t = \begin{bmatrix} y_{t-1}, y_{t-2}, r_{t-1}, r_{t-2}, \pi_{t-1}, \pi_{t-2,4} \end{bmatrix}' \tag{A.27}
\]
\[
\xi_t = \begin{bmatrix} y_{t}^*, y_{t-1}^*, y_{t-2}^*, g_{t-1}, g_{t-2}, z_{t-1}, z_{t-2} \end{bmatrix}' \tag{A.28}
\]

\[
A = \begin{bmatrix} a_y,1 & a_y,2 & a_r,2 & a_r,2 & 0 & 0 & 0 \[0.5cm] b_y & 0 & 0 & b_{\pi} (1 - b_{\pi}) \end{bmatrix},
H = \begin{bmatrix} 1 - a_y,1 & -a_y,2 & -a_r,2 & -a_r,2 & -a_r,2 & -a_r,2 & -a_r,2 \[0.5cm] 0 - b_y & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.'
\]

\[
F = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \[0.5cm] 1 & 0 & 0 & 0 & 0 & 0 & 0 \[0.5cm] 0 & 1 & 0 & 0 & 0 & 0 & 0 \[0.5cm] 0 & 0 & 0 & 1 & 0 & 0 & 0 \[0.5cm] 0 & 0 & 0 & 0 & 1 & 0 & 0 \[0.5cm] 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},
S = \begin{bmatrix} 0 & 1 & 0 \[0.5cm] 0 & 0 & 0 \[0.5cm] 0 & 0 & 1 \[0.5cm] 0 & 0 & 0 \[0.5cm] 0 & 0 & 0 \end{bmatrix}.
\]

The measurement and state relations are:

\[
y_t = Ax_t + H\xi_t + v_t
\]

\[
\begin{bmatrix} y_t \[y_t] \end{bmatrix} = \begin{bmatrix} a_y,1 & a_y,2 & a_r,2 & a_r,2 & 0 & 0 & 0 \[0.5cm] b_y & 0 & 0 & b_{\pi} (1 - b_{\pi}) \end{bmatrix} + \begin{bmatrix} 1 - a_y,1 & -a_y,2 & -a_r,2 & -a_r,2 & -a_r,2 & -a_r,2 \[0.5cm] 0 - b_y & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} 
\end{bmatrix}
\]

\[
\xi_t = F\xi_{t-1} + S\epsilon_t
\]

\[
\begin{bmatrix} y_{t-1} \[y_{t-1}] \end{bmatrix}
\]

A-7

Electronic copy available at: https://ssrn.com/abstract=3897838
In the Stage 3 model, Holston et al. (2017) replace $\sigma_y^2$ and $\sigma_\pi^2$ in $\text{Var}(\varepsilon_i) = W = \text{diag}(\sigma_y^2, \sigma_\pi^2, \sigma_\pi^2)$ with $(\lambda_8 \sigma_y^2)$ and $(\lambda_2 \sigma_y/a_r)^2$, respectively, from the two previous estimation steps, so that:

$$\text{Var}(S\varepsilon_i) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_y^2 & 0 & 0 \\ 0 & (\lambda_8 \sigma_y^2)^2 & 0 \\ 0 & 0 & (\lambda_2 \sigma_y/a_r)^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$  \quad (A.31)

which enters the Kalman Filter routine and ML estimation of the final Stage 3 parameters.

Expanding the relations in (A.29) and (A.30) leads to the following measurement:

$$y_t = y_t^* + a_{y,1}(y_{t-1} - y_{t-1}^*) + a_{y,2}(y_{t-2} - y_{t-2}^*) + \frac{\sigma_y}{2}(r_{t-1} - g_{t-1} - z_{t-1}) + \frac{\sigma_y}{2}(r_{t-2} - g_{t-2} - z_{t-2}) + \varepsilon_i^*$$

$$\pi_t = b_y(y_{t-1} - y_{t-1}^*) + b_\pi \pi_{t-1} + (1 - b_\pi) \pi_{t-2} + \varepsilon_\pi^*$$

and corresponding state relations

$$y_t^* = y_{t-1}^* + g_{t-2} + \varepsilon_i^{y_t^*} + \varepsilon_i^{y_t^*}_{t-1}$$

$$y_{t-1}^* = y_{t-1}^*$$

$$y_{t-2}^* = y_{t-2}^*$$

$$g_{t-1} = g_{t-2} + \varepsilon_i^{g_{t-1}}$$

$$g_{t-2} = g_{t-2}$$

$$z_{t-1} = z_{t-2} + \varepsilon_i^{z_{t-1}}$$

$$z_{t-2} = z_{t-2}.$$  \quad (A.32)
Defining output $y_t$ once again as trend plus cycle, dropping identities and simplifying gives the following system of Stage 3 relations:

$$y_t = y_t^* + \tilde{y}_t$$  \hspace{1cm} (A.33a)

$$\tau_t = b_\pi \tau_{t-1} + (1 - b_\pi) \tau_{t-2} + b_y \tilde{y}_{t-1} + \epsilon_t^\pi$$  \hspace{1cm} (A.33b)

$$a_y(L)\tilde{y}_t = \frac{a_y}{2} (r_t - g_t - z_{t-1}) + \frac{a_y}{2} (r_{t-2} - g_{t-2} - z_{t-2}) + \epsilon_t^g$$  \hspace{1cm} (A.33c)

$$y_t^* = y_{t-1}^* + \tilde{y}_{t-1} + \epsilon_t^{y^*}$$  \hspace{1cm} (A.33d)

$$g_{t-1} = g_{t-2} + \epsilon_{t-1}^g$$  \hspace{1cm} (A.33e)

$$z_{t-1} = z_{t-2} + \epsilon_{t-1}^z$$  \hspace{1cm} (A.33f)

with the corresponding vector of Stage 3 model parameters to be estimated by MLE being:

$$\theta_3 = [a_{y,1}, a_{y,2}, a_r, b_\pi, b_y, \sigma_{\tilde{y}}, \sigma_{\tau}, \sigma_{\epsilon^y}]'.$$  \hspace{1cm} (A.34)

For the Stage 3 model, the variance of the state vector $\xi_t$ is initialized once more as outlined in (12a) and Footnote 27, with the numerical value of $\xi_{00}$ and $P_{00}$ being:

$$\xi_{00} = [806.4455, 805.2851, 804.1248, 1.1604, 1.1603, 0, 0]$$  \hspace{1cm} (A.35)

$$P_{00} = \begin{bmatrix}
0.7272 & 0.2 & 0 & 0.2009 & 0.2 & 0 & 0 \\
0.2 & 0.2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.2 & 0 & 0 & 0 & 0 \\
0.2009 & 0 & 0 & 0.2009 & 0.2 & 0 & 0 \\
0.2 & 0 & 0 & 0.2 & 0.2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.2227 & 0.2 \\
0 & 0 & 0 & 0 & 0 & 0.2 & 0.2
\end{bmatrix}.$$  \hspace{1cm} (A.36)

### A.4. Additional simulation results

As an additional experiment, I simulate entirely unrelated univariate time series processes as inputs into the $Y_t$ and $X_t$ vector series needed for the structural break regressions in (54). As before, the simulated inputs that are required are the cycle variable $\tilde{y}_t$, trend growth $g_t$ as well as the real rate $r_t$. To avoid having to use the observed exogenous interest rate series that makes up the real rate via the relation $r_t = i_t - \pi^e_t$ ($\pi^e_t$ is expected inflation as defined in (3)) as it will be function of $r^*_t$ and hence $g_t$ and $z_t$, I fit a low order ARMA process to $r_t$. I then use the coefficients from this estimated ARMA model to generate a simulated sequence of $T$ observations from the real interest rate. I follow the same strategy to generate a simulated series for $\tilde{y}_t$. Note that I do not simply use the AR(2) model structure for the cycle series $\tilde{y}_t$ as is implied by the left hand side of (36c) together with the $a_{y,1}$ and $a_{y,2}$ estimates from the Stage 2 model in the simulation. The reason for this is that the empirical $\hat{\tilde{y}}_{||T}$ series that Holston et al. (2017) use in their procedure is the Kalman Smoother based estimate of $\tilde{y}_t$ which portrays a more complicated autocorrelation pattern than an AR(2) process. In order to match the autocorrelation pattern of the $\hat{\tilde{y}}_{||T}$ series as closely as possible, I fitted the best (low order) ARMA process to $\hat{\tilde{y}}_{||T}$, and used those coefficients to generate the simulated cycle series.
For the $g_t$ series, I use three different simulation scenarios. First, I replace the trend growth estimate in $\mathbf{X}_t$ by the Kalman Smoother estimate of $g_t$ denoted by $\hat{g}_{1:T}$ above. This is the same series that Holston et al. (2017). Second, I simulate $g_{t-1}$ from a pure random walk (RW) process with the standard deviation of the error term set equal to $\hat{\sigma}_s = 0.0305205$, the implied estimate reported in column 1 of (9). Third, I simulate a simple (Gaussian) white noise (WN) for $g_{t-1}$. And last, I fit a low order ARMA process to the first difference of $\hat{g}_{1:T}$. The "empirical" $\hat{g}_{1:T}$ series is very persistent and its dynamics are not sufficiently captured by a pure RW. I therefore use the coefficients from a fitted ARMA model to $\Delta \hat{g}_{1:T}$ to simulate the first difference process $\Delta g_{t-1}$, and then construct the $g_{t-1}$ as the cumulative sum of $\Delta g_{t-1}$ in $\mathbf{X}_t$. All simulation scenarios are based on 1000 repetitions of sample size $T$ and the EW structural break test.

In Table A.1, summary statistics of the $\lambda_z$ estimates obtained from implementing Holston et al.’s (2017) MUE procedure in Stage 2 are shown. The summary statistics are means and medians, as well as empirical probabilities of observing an $\lambda_z$ estimate computed from the simulated data being larger than the Stage 2 estimate of 0.030217 from Holston et al. (2017). In Figure A.1, I show histogram plots corresponding to the summary statistics of the $\lambda_z$ estimates computed from the simulated data. These are shown as supplementary information to complement the summary statistics in Table A.1 and to avoid concerns related to unusual simulation patterns.

**Table A.1**: Summary statistics of $\lambda_z$ estimates of the Stage 2 MUE procedure applied to data simulated from unrelated univariate ARMA processes

| Summary Statistic | $g_{t-1} = \hat{g}_{1:T}$ | $g_{t-1} \sim$ RW | $g_{t-1} \sim$ WN | $\Delta g_{t-1} \sim$ ARMA |
|-------------------|--------------------------|-------------------|-------------------|--------------------------|
| Minimum           | 0                        | 0                 | 0                 | 0                        |
| Maximum           | 0.097019                 | 0.095914          | 0.096789          | 0.093340                 |
| Standard deviation| 0.015240                 | 0.015858          | 0.016803          | 0.016335                 |
| Mean              | 0.031798                 | 0.029708          | 0.026117          | 0.030449                 |
| Median            | 0.030165                 | 0.028647          | 0.024254          | 0.029435                 |
| $\Pr(\lambda_z > 0.030217)$ | 0.498000 | 0.456000          | 0.384000          | 0.482000 |

**Notes:** This table reports summary statistics of the Stage 2 estimates of $\lambda_z$ that one obtains when applying Holston et al.’s (2017) MUE procedure to simulated data without the $z_t$ process. The summary statistics that are reported are the minimum, maximum, standard deviation, mean, median, as well as the empirical frequency of observing a value larger than the estimate of 0.030217 obtained by Holston et al. (2017), denoted by $\Pr(\lambda_z > 0.030217)$. The columns show the estimates for the four different data generating processes for trend growth $g_t$. The first column reports results when the Kalman Smoother estimate $\hat{g}_{1:T}$ is used for $g_{t-1}$. The second and third columns show estimates when $g_{t-1}$ is generated as pure random walk (RW) or (Gaussian) white noise (WN) process. The last column reports results when $g_{t-1}$ is computed as the cumulative sum of $\Delta g_{t-1}$, which is simulated from the coefficients obtained from a low order ARMA process fitted to $\Delta g_{t-1}$. The cycle and real rate series are also constructed by first finding the best fitting low order ARMA processes to the individual series and then simulating from fitted coefficients.

Looking over the results in Table A.1 and histograms in Figure A.1, it is clear that there are many instances where the estimates of $\lambda_z$ from the simulated data are not only non-zero, but rather sizeable, being larger than the estimate of $\lambda_z = 0.030217$ that Holston et al. (2017) compute from the empirical data. Note that there is no $z_t$ process simulated, yet with Holston et al. (2017) Stage 2 MUE procedure one can recover an estimate that is at least as large as the empirical one around 40 to 50 percent of the time, depending on how $g_t$ is simulated. This simulation exercise thus highlights how spurious Holston et al.’s (2017) MUE procedure to estimate $\lambda_z$ is. As the downward trend in the $z_t$ process drives the movement in the natural rate, where the severity of the downward trend is related to the
Figure A.1: Histograms of the estimated $\{\hat{\lambda}^S_z\}_{s=1}^S$ sequence corresponding to the summary statistics shown in Table A.1

magnitude of $\sigma_z$, which is through $\lambda_z$, Holston et al.’s (2017) estimates of the natural rate are likely to be downward biased.

A.5. Additional figures and tables

This section presents additional figures and tables to complement the results reported in the main text. Some of these results are based on an expanded sample period using data that ends in 2019:Q2.
Figure A.2: Kalman smoothed estimates of (annualized) trend growth $g_t$ and output gap (cycle) $\tilde{y}_t$ from Holston et al.’s (2017) ‘misspecified’ Stage 2 model (HLW blue solid line) and the ‘correctly specified’ Stage 2 model (MLE($\sigma_g$),$\mathcal{M}_0$ red dashed lined). These are used as inputs into the structural break dummy variable regression in (54).
Figure A.3: Sequences of $\{F(\tau)\}_{\tau=\tau_0}$ statistics from the structural break dummy variable regressions in (54) for the different scenarios that are considered.
Misspecified: $GY_t = a_g(L)\tilde{y}_t - a_r(L)r_t - a_g y_{t-1} - a_0$

Correctly specified: $GY_t = a_g(L)\tilde{y}_t - a_r(L)[r_t - 4 \gamma_t]$

Figure A.4: Sequence of $\{F(\tau)\}_{\tau = \tau_0}^{T_1}$ statistics on the dummy variable coefficients $\{\hat{\zeta}_1(\tau)\}_{\tau = \tau_0}^{T_1}$ used in the construction of the structural break test statistics.
Table A.2: Stage 2 MUE results of $\lambda_z$ with corresponding 90% confidence intervals, structural break test statistics and $p-$values using data up to 2019:Q2

| $\lambda_z$ | Time varying $\phi$ | Constant $\phi$ |
|-------------|---------------------|-----------------|
|             | HLW.R-File | Replicated | [90% CI] | MLE($\sigma_g$) | [90% CI] | MLE($\sigma_g$).$M_0$ | [90% CI] | Replicated | [90% CI] | MLE($\sigma_g$) | [90% CI] | MLE($\sigma_g$).$M_0$ | [90% CI] |
| $L$         | —         | 0          | [0, 0.02] | 0               | [0, 0.00] | 0.011597               | [0, 0.07] | 0          | [0, 0.02] | 0               | [0, 0.00] | 0.011597               | [0, 0.05] |
| MW          | 0.031699  | 0.031699  | [0, 0.14] | 0.039255         | [0, 0.17] | 0.014622               | [0, 0.08] | 0          | [0, 0.03] | 0               | [0, 0.02] | 0.011044               | [0, 0.06] |
| EW          | 0.035202  | 0.035202  | [0, 0.13] | 0.040416         | [0, 0.14] | 0.014773               | [0, 0.07] | 0          | [0, 0.03] | 0               | [0, 0.02] | 0.012272               | [0, 0.07] |
| QLR         | 0.044029  | 0.044029  | [0, 0.15] | 0.048009         | [0, 0.16] | 0.022654               | [0, 0.09] | 0          | [0, 0.05] | 0               | [0, 0.04] | 0.020062               | [0, 0.07] |

Structural break test statistics ($p-$values in parenthesis)

| $L$       | —         | 0.049851  | (0.8750)  | 0.037369         | (0.9450)  | 0.159840               | (0.3600)  | 0.049851  | (0.8750)  | 0.037369         | (0.9450)  | 0.159840               | (0.3600)  |
| MW        | 2.678576  | 2.678577  | (0.0600)  | 3.796062         | (0.0200)  | 1.107323               | (0.3050)  | 0.330273  | (0.8100)  | 0.254845         | (0.8850)  | 0.928099               | (0.3750)  |
| EW        | 2.486621  | 2.486622  | (0.0300)  | 3.140059         | (0.0100)  | 0.736389               | (0.3000)  | 0.200885  | (0.7900)  | 0.150082         | (0.8750)  | 0.642384               | (0.3450)  |
| QLR       | 12.221517 | 12.221521 | (0.0100)  | 13.648720        | (0.0050)  | 5.987862               | (0.1600)  | 2.767909  | (0.5900)  | 2.174813         | (0.7250)  | 5.432014               | (0.2000)  |

Notes: This table reports the Stage 2 estimates of $\lambda_z$ for the different $\theta_2$ estimates corresponding to the “misspecified” and “correctly specified” Stage 2 models reported in Table 9 using data updated to 2019:Q2. The table is split into two column blocks, showing the results for the “time varying $\phi$” and “constant $\phi$” scenarios in the left and right blocks, respectively. In the bottom half of the table, the four different structural break test statistics for the considered models are shown. The results under the heading ‘HLW.R-File’ show the $\lambda_z$ estimates obtained from running Holston et al.’s (2017) R-Code for the Stage 2 model as reference values. The second column ‘Replicated’ shows my replicated results. Under the heading ‘MLE($\sigma_g$)’, results for the “misspecified” Stage 2 model are shown with $\sigma_g$ estimated directly by MLE rather than from the first stage estimate of $\lambda_g$. Under the heading ‘MLE($\sigma_g$).$M_0$’, results for the “correctly specified” Stage 2 model are reported where $\sigma_g$ is again estimated by MLE. The values in square brackets in the top half of the table report 90% confidence intervals for $\lambda_z$ computed from Stock and Watson’s (1998) tabulated values provided in their GAUSS files. These were divided by sample size $T$ to make them comparable to $\lambda_z$. In the bottom panel, $p-$values of the various structural break tests are reported in round brackets. These were also extracted from Stock and Watson’s (1998) GAUSS files.
Table A.3: Stage 3 parameter estimates using data up to 2019:Q2

| \( \theta_3 \) | HLW.R-File | Replicated | MLE(\( \sigma_g | \lambda_{HLW}^z \)) | MLE(\( \sigma_g | \lambda_{M0}^z \)) | MLE(\( \sigma_g, \sigma_z \)) |
|---|---|---|---|---|---|
| \( a_{y,1} \) | 1.53876458 | 1.53876459 | 1.51083223 | 1.51659115 | 1.51669620 |
| \( a_{y,2} \) | -0.59700264 | -0.59700265 | -0.57637540 | -0.57645931 | |
| \( a_r \) | -0.06854043 | -0.06854043 | -0.07029671 | -0.07000675 | |
| \( b_{\pi} \) | 0.67331545 | 0.67331545 | 0.67683450 | 0.67483411 | |
| \( b_y \) | 0.07755450 | 0.07755451 | 0.07885243 | 0.07885243 | |
| \( \sigma_g \) | 0.33590693 | 0.33590692 | 0.34747670 | 0.34826106 | |
| \( \sigma_{\pi} \) | 0.78812554 | 0.78812554 | 0.78854952 | 0.78862036 | |
| \( \sigma_{y^*} \) | 0.57577319 | 0.57577320 | 0.56359236 | 0.56327803 | |
| \( \sigma_{\tilde{g}} \) (implied) | (0.03082331) | (0.03082331) | (0.04378919) | (0.04378919) | |
| \( \sigma_{\tilde{g}} \) (implied) | (0.17251762) | (0.17251762) | (0.05295494) | (0.05295494) | |
| \( \lambda_{\tilde{g}} \) (implied) | 0.05353377 | 0.05353377 | (0.07774103) | |
| \( \lambda_{\tilde{z}} \) (implied) | 0.03520151 | 0.03520151 | (0.01055443) | |

Log-likelihood: -533.36984524

Notes: This table reports replication results for the Stage 3 model parameter vector \( \theta_3 \) of Holston et al. (2017). The first column (HLW.R-File) reports estimates obtained by running Holston et al.’s (2017) R-Code for the Stage 3 model. The second column (Replicated) shows the replicated results using the same set-up as in Holston et al.’s (2017). The third column (MLE(\( \sigma_g | \lambda_{HLW}^z \))) reports estimates when \( \sigma_g \) is directly estimated by MLE together with the other parameters of the Stage 3 model, while \( \lambda_z \) is held fixed at \( \lambda_z^{HLW} = 0.035202 \) obtained from Holston et al.’s (2017) “misspecified” Stage 2 procedure. In the forth column (MLE(\( \sigma_g | \lambda_{M0}^z \))), \( \sigma_g \) is again estimated directly by MLE together with the other parameters of the Stage 3 model, but with \( \lambda_z \) now fixed at \( \lambda_z^{M0} = 0.012272 \) obtained from the “correctly specified” Stage 2 model in (42). The last column (MLE(\( \sigma_g, \sigma_z \))) shows estimates when all parameters are computed by MLE. Values in round brackets give the implied \( \{ \sigma_g, \sigma_z \} \) or \( \{ \lambda_{\tilde{g}}, \lambda_{\tilde{z}} \} \) values when either is fixed or estimated. The last row (Log-likelihood) reports the value of the log-likelihood function at these parameter estimates. The Matlab file Stage3_replication.m replicates these results.
Figure A.5: Filtered estimates of the natural rate $r_t^*$, annualized trend growth $g_t$, ‘other factor’ $z_t$, and the output gap (cycle) variable $\tilde{y}_t$ up to 2019:Q2.
Figure A.6: Smoothed estimates of the natural rate $r^*_t$, annualized trend growth $g_t$, ‘other factor’ $z_t$, and the output gap (cycle) variable $\tilde{y}_t$ up to 2019:Q2.
Figure A.7: GDP growth and recursively estimated mean of GDP growth from 2009:Q3 to 2019:Q2.

Figure A.8: Mean and Median 10 year (real) GDP growth forecasts from the Survey of Professional forecasters (SPF) obtained from https://www.philadephiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/rgdp10. The blue shaded region marks the 25th to 75th percentile region of the cross-section of forecaster.

Figure A.9: GMSU-Vanguard survey based expected 3 year and 10 year (real) GDP growth from February 2017 to April 2020, taken from Figure II on page 5 in Giglio et al. (2020) (see the appendix in Giglio et al. (2020) for more details on the design of the client/investor survey).
A.6. R-Code Snippets

This sections shows various parts of the R-Code that is provided by Holston et al. (2017) in the zip file from https://www.newyorkfed.org/medialibrary/media/research/economists/williams/data/HLWCode.zip. Below the code next to each of the headers, the name of the R-file is listed from which the code is displayed.

```r
#--------------------------------------------------#
# File: rstar.stage3.R
# # Description: This file runs the model in the third stage of the HLW estimation.
#--------------------------------------------------#

rstar.stage3 <- function(log.output,
inflation,
real.interest.rate,
nominal.interest.rate,
lambda.g,
lambda.z,
a3.constraint=NA,
b2.constraint=NA,
run.se = TRUE) {

stage <- 3

# Data must start 4 quarters before the estimation period
T <- length(log.output) - 4

# Original output gap estimate
x.og <- cbind(rep(1,T+4), 1:(T+4))
y.og <- log.output
output.gap <- (y.og - x.og %*% solve(t(x.og) %*% x.og, t(x.og) %*% y.og)) * 100

# Initialization of state vector for Kalman filter using HP trend of log output
log.output.hp.trend <- hpfilter(log.output,freq=36000,type="lambda",drift=FALSE)$trend
g.pot <- log.output.hp.trend[(g.pot.start.index):length(log.output.hp.trend)]
g.pot.diff <- diff(g.pot)
xi.00 <- c(100*g.pot[3:1],100*g.pot.diff[2:1],0,0)

# IS curve
y.is <- output.gap[5:(T+4)]
x.is <- cbind(output.gap[4:(T+3)], output.gap[3:(T+2)],
(real.interest.rate[4:(T+3)] + real.interest.rate[3:(T+2)])/2,
rep(1,T))
b.is <- solve(t(x.is) %*% x.is, t(x.is) %*% y.is)
r.is <- as.vector(y.is - x.is %*% b.is)
s.is <- sqrt(sum(r.is^2) / (length(r.is)-(dim(x.is)[2])))

# Phillips curve
y.ph <- inflation[5:(T+4)]
x.ph <- cbind(inflation[4:(T+3)],
(inflation[3:(T+2)]+inflation[2:(T+1)]+inflation[1:T])/3,
output.gap[4:(T+3)])
b.ph <- solve(t(x.ph) %*% x.ph, t(x.ph) %*% y.ph)
r.ph <- y.ph - x.ph %*% b.ph
s.ph <- sqrt(sum(r.ph^2) / (length(r.ph)-(dim(x.ph)[2])))

y.data <- cbind(100 * log.output[5:(T+4)],
inflation[5:(T+4)])
x.data <- cbind(100 * log.output[4:(T+3)],
100 * log.output[3:(T+2)],
real.interest.rate[4:(T+3)],
real.interest.rate[3:(T+2)],
inflation[4:(T+3)],
(inflation[3:(T+2)]+inflation[2:(T+1)]+inflation[1:T])/3)

# Starting values for the parameter vector
```

Electronic copy available at: https://ssrn.com/abstract=3897838
initial.parameters <- c(b.is[1:3], b.ph[1], b.ph[3], s.is, s.ph, 0.7)

# Set an upper and lower bound on the parameter vectors:
# The vector is unbounded unless values are otherwise specified
theta.lb <- c(rep(-Inf,length(initial.parameters)))
theta.ub <- c(rep(Inf,length(initial.parameters)))

# Set a lower bound for the Phillips curve slope (b_2) of \textit{b2.constraint}, if not NA
# In HLW, \textit{b2.constraint} = 0.025
if (!is.na(b2.constraint)) {
  print(paste0("Setting a lower bound on b_2 of ",as.character(b2.constraint)))
  if (initial.parameters[5] < b2.constraint) {
    initial.parameters[5] <- b2.constraint

    theta.lb[5] <- b2.constraint
  }
}

# Set an upper bound for the IS curve slope (a_3) of \textit{a3.constraint}, if not NA
# In HLW, \textit{a3.constraint} = -0.0025
if (!is.na(a3.constraint)) {
  print(paste0("Setting an upper bound on a_3 of ",as.character(a3.constraint)))
  if (initial.parameters[3] > a3.constraint) {
    initial.parameters[3] <- a3.constraint

    theta.ub[3] <- a3.constraint
  }
}

# Set the initial covariance matrix (see footnote 6)
P.00 <- calculate.covariance(initial.parameters, theta.lb, theta.ub, y.data, x.data, stage, lambda.g, lambda.z, xi.00)

# Get parameter estimates via maximum likelihood
f <- function(theta) {return(-log.likelihood.wrapper(theta, y.data, x.data, stage, lambda.g, lambda.z, xi.00)$ll.cum)}
nloptr.out <- nloptr(initial.parameters, f, eval_grad_f=function(x) {gradient(f, x)},
  lb=theta.lb,ub=theta.ub,
  opts=list("algorithm"="NLOPT_LD_LBFGS","xtol_rel"=1.0e-8))
theta <- nloptr.out$solution

log.likelihood <- log.likelihood.wrapper(theta, y.data, x.data, stage, lambda.g, lambda.z, xi .00, P.00)$ll.cum

# Get state vectors (\textit{xi.tt}, \textit{xi.ttm1}, \textit{xi.T}, \textit{P.tt}, \textit{P.ttm1}, \textit{P.T}) via Kalman filter
states <- kalman.states.wrapper(theta, y.data, x.data, stage, lambda.g, lambda.z, xi.00, P.00)

# If \textit{run.se} = TRUE, compute standard errors for estimates of the states (see footnote 7) and
# report run time
if (run.se) {
  ptm <- proc.time()
  se <- kalman.standard.errors(T, states, theta, y.data, x.data, stage, lambda.g, lambda.z, xi .00, P.00, niter, a3.constraint, b2.constraint)
  print("Standard error procedure run time")
  print(proc.time() - ptm)
}

# One-sided (filtered) estimates
trend.filtered <- states$filtered$xi.tt[,4] * 4
z.filtered <- states$filtered$xi.tt[,6]
rstar.filtered <- trend.filtered + z.filtered
potential.filtered <- states$filtered$xi.tt[,1]/100
output.gap.filtered <- y.data[,1] - (potential.filtered * 100)

# Two-sided (smoothed) estimates
trend.smoothed <- states$smoothed$xi.tt[,4] * 4
z.smoothed <- states$smoothed$xi.tt[,6]
rstar.smoothed <- trend.smoothed + z.smoothed
potential.smoothed <- states$smoothed$xi.tt[,1]/100
output.gap.smoothed <- y.data[,1] - (potential.smoothed * 100)

R-Code 1: rstar.stage3.R
# File: calculate.covariance.R
#
# Description: This function calculates the covariance matrix of the
# initial state from the gradients of the likelihood function.
#
calculate.covariance <- function(initial.parameters, theta.lb, theta.ub, y.data, x.data, stage, lambda.g = NA, lambda.z = NA, xi.00) {
  n.state.vars <- length(xi.00)
  # Set covariance matrix equal to 0.2 times the identity matrix
  P.00 <- diag(0.2, n.state.vars, n.state.vars)
  # Get parameter estimates via maximum likelihood
  f <- function(theta) {return(-log.likelihood.wrapper(theta, y.data, x.data, stage, lambda.g, lambda.z, xi.00, P.00))}
  nloptr.out <- nloptr(initial.parameters, f, eval_grad_f=function(x) {gradient(f, x)},
                      lb=theta.lb, ub=theta.ub, opts=list("algorithm"="NLOPT_LD_LBFGS","xtol_rel" =1.0e-8))
  theta <- nloptr.out$solution
  # Run Kalman filter with above covariance matrix and corresponding parameter estimates
  states <- kalman.states.wrapper(theta, y.data, x.data, stage, lambda.g, lambda.z, xi.00, P.00)
  # Save initial covariance matrix
  P.00 <- states$filtered$P.ttm1[1:n.state.vars,]
  return(P.00)
}

R-Code 2: calculate.covariance.R
# File: unpack.parameters.stage1.R

# Description: This file generates coefficient matrices for the stage 1 state-space model for the given parameter vector.

# Stage 1 parameter vector: [a_y,1, a_y,2, b_pi, b_y, g, sigma_y~, sigma_pi, sigma_y*]

unpack.parameters.stage1 <- function(parameters, y.data, x.data, xi.00, P.00) {
  A <- matrix(0, 4, 2)
  A[1:2, 1] <- parameters[1:2] # a_y,1, a_y,2
  A[1, 2] <- parameters[4] # b_y
  A[3, 2] <- parameters[3] # b_pi
  A[4, 2] <- 1 - parameters[3] # 1 - b_pi

  H <- matrix(0, 3, 2)
  H[1, 1] <- 1
  H[2:3, 1] <- -parameters[1:2] # -a_y,1, -a_y,2
  H[2, 2] <- -parameters[4] # -b_y

  R <- diag(c(parameters[6]^2, parameters[7]^2)) # sigma_y~, sigma_pi
  Q <- matrix(0, 3, 3)
  Q[1, 1] <- parameters[8]^2 # sigma_y*

  F <- matrix(0, 3, 3)
  F[1, 1] <- F[2, 1] <- F[3, 2] <- 1

  # Make the data stationary
  y.data[, 1] <- y.data[, 1] - 1:dim(y.data)[1] * parameters[5] # g
  x.data[, 1] <- x.data[, 1] - 0:(dim(x.data)[1]-1) * parameters[5]
  x.data[, 2] <- x.data[, 2] - 1:(dim(x.data)[1]-2) * parameters[5]

  return(list("xi.00"=xi.00, "P.00"=P.00, "F"=F, "Q"=Q, "A"=A, "H"=H, "R"=R, "x.data"=x.data, "y.data"=y.data))
}

R-Code 3: unpack.parameters.stage1.R
# File: kalman.states.wrapper.R
#
# Description: This is a wrapper function for kalman.states.R that specifies
# inputs based on the estimation stage.
#
kalman.states.wrapper <- function(parameters, y.data, x.data, stage = NA,
lambda.g=NA, lambda.z=NA, xi.00=NA, P.00=NA){
  if (stage == 1) {
    out <- unpack.parameters.stage1(parameters, y.data, x.data,
    xi.00, P.00)
  } else if (stage == 2) {
    out <- unpack.parameters.stage2(parameters, y.data, x.data,
    lambda.g, xi.00, P.00)
  } else if (stage == 3) {
    out <- unpack.parameters.stage3(parameters, y.data, x.data,
    lambda.g, lambda.z, xi.00, P.00)
  } else {
    stop('You need to enter a stage number in kalman.states.wrapper.')</n  }
  for (n in names(out)) {
    eval(parse(text=paste0(n, "<-out$n)))
  }
  T <- dim(y.data)[1]
  states <- kalman.states(xi.00, P.00, F, Q, A, H, R, y.data, x.data)
  if (stage == 1) {
    states$filtered$xi.tt <- states$filtered$xi.tt + cbind(1:T,0:(T-1),-1:(T-2)) * parameters[5]
    states$smoothed$xi.tT <- states$smoothed$xi.tT + cbind(1:T,0:(T-1),-1:(T-2)) * parameters[5]
  }
  return(states)
}
# File: rstar.stage2.R
#
# Description: This file runs the model in the second stage of the HLW estimation.
#
#---------------------------------------------------------------------

rstar.stage2 <- function(log.output, inflation, real.interest.rate, lambda.g, 
a3.constraint=NA, 
b2.constraint=NA) {

  stage <- 2

  # Data must start 4 quarters before the estimation period
  T <- length(log.output) - 4

  # Original output gap estimate
  x.og <- cbind(rep(1,T+4), 1:(T+4))
  y.og <- log.output
  output.gap <- (y.og - x.og %*% solve(t(x.og) %*% x.og, t(x.og) %*% y.og)) * 100

  # Initialization of state vector for Kalman filter using HP trend of log output
  log.output.hp.trend <- hpfilter(log.output, freq=36000, type="lambda", drift=FALSE)
  g.pot <- log.output.hp.trend[(g.pot.start.index):length(log.output.hp.trend)]
  g.pot.diff <- diff(g.pot)
  xi.00 <- c(100*g.pot[3:1], 100*g.pot.diff[2])

  # IS curve
  y.is <- output.gap[5:(T+4)]
  x.is <- cbind(output.gap[4:(T+3)], output.gap[3:(T+2)],
                (real.interest.rate[4:(T+3)] + real.interest.rate[3:(T+2)])/2,
                rep(1,T))
  b.is <- solve(t(x.is) %*% x.is, t(x.is) %*% y.is)
  r.is <- as.vector(y.is - x.is %*% b.is)
  s.is <- sqrt(sum(r.is^2) / (length(r.is)-(dim(x.is)[2])))

  # Phillips curve
  y.ph <- inflation[5:(T+4)]
  x.ph <- cbind(inflation[4:(T+3)],
                (inflation[3:(T+2)]+inflation[2:(T+1)]+inflation[1:T])/3,
                output.gap[4:(T+3)])
  b.ph <- solve(t(x.ph) %*% x.ph, t(x.ph) %*% y.ph)
  r.ph <- y.ph - x.ph %*% b.ph
  s.ph <- sqrt(sum(r.ph^2) / (length(r.ph)-(dim(x.ph)[2])))

  # Phillips curve
  y.data <- cbind(100 * log.output[5:(T+4)],
                  inflation[5:(T+4)])
  x.data <- cbind(100 * log.output[4:(T+3)],
                  100 * log.output[3:(T+2)],
                  real.interest.rate[4:(T+3)],
                  real.interest.rate[3:(T+2)],
                  inflation[4:(T+3)],
                  (inflation[3:(T+2)]+inflation[2:(T+1)]+inflation[1:T])/3,
                  rep(1,T))

  # Starting values for the parameter vector
  initial.parameters <- c(b.is, -b.is[3], b.ph[1], b.ph[3], s.is, s.ph, 0.5)

  # Set an upper and lower bound on the parameter vectors:
  # The vector is unbounded unless values are otherwise specified
  theta.lb <- c(rep(-Inf,length(initial.parameters)))
  theta.ub <- c(rep(Inf,length(initial.parameters)))

  # Set a lower bound for the Phillips curve slope (b_2) of b2.constraint, if not NA
  # In HLW, b2.constraint = 0.025
  if (!is.na(b2.constraint)) {
    if (initial.parameters[7] < b2.constraint) {
      initial.parameters[7] <- b2.constraint
    }
  }
}
theta.lb[7] <- b2.constraint

# Set an upper bound for the IS curve slope (a_3) of a3.constraint, if not NA
# In HLW, a3.constraint = -0.0025
if (!is.na(a3.constraint)) {
  if (initial.parameters[3] > a3.constraint) {
    initial.parameters[3] <- a3.constraint
  }
  theta.ub[3] <- a3.constraint
}

# Set the initial covariance matrix (see footnote 6)
P.00 <- calculate.covariance(initial.parameters, theta.lb, theta.ub, y.data, x.data, stage, lambda.g, NA, xi.00)

# Get parameter estimates via maximum likelihood
f <- function(theta) {return(-log.likelihood.wrapper(theta, y.data, x.data, stage, lambda.g, NA, xi.00, P.00)$ll.cum)}
nloptr.out <- nloptr(initial.parameters, f, eval_grad_f=function(x) {gradient(f, x)}, lb=theta.lb,ub=theta.ub, opts=list("algorithm"="NLOPT_LD_LBFGS","xtol_rel"=1.0e-8))
theta <- nloptr.out$solution

log.likelihood <- log.likelihood.wrapper(theta, y.data, x.data, stage, lambda.g, NA, xi.00, P.00)$ll.cum

# Get state vectors (xi.tt, xi.ttm1, xi.tT, P.tt, P.ttm1, P.tT) via Kalman filter
states <- kalman.states.wrapper(theta, y.data, x.data, stage, lambda.g, NA, xi.00, P.00)

# Two-sided (smoothed) estimates
trend.smoothed <- states$smoothed$xi.tt[,4] * 4
potential.smoothed <- c(states$smoothed$xi.tT[1,3:2], states$smoothed$xi.tT[1,])
output.gap.smoothed <- 100 * log.output[3:(T+4)] - potential.smoothed

# Inputs for median.unbiased.estimator.stage2.R
y <- output.gap.smoothed[3:length(output.gap.smoothed)]
x <- cbind(output.gap.smoothed[2:(length(output.gap.smoothed)-1)], output.gap.smoothed[1:(length(output.gap.smoothed)-2)], (x.data[,3]+x.data[,4])/2, states$smoothed$xi.tT[,4], rep(1,T))
median.unbiased.estimator.stage2 <- function(y, x) {
  T <- dim(x)[1]
  stat <- rep(0, T-2*4+1)
  for (i in 4:(T-4)) {
    xr <- cbind(x, c(rep(0, i), rep(1, T-i)))
    xi <- solve(t(xr)%*%xr)
    b <- solve(t(xr)%*%xr, t(xr)%*%y)
    s3 <- sum((y-xr%*%b)^2)/(T-dim(xr)[2])
    stat[i+1-4] <- b[dim(xr)[2]]/sqrt(s3*xi[dim(xr)[2],dim(xr)[2]])
  }
  ew <- 0
  for (i in 1:length(stat)) {
    ew <- ew+exp((stat[i]^2)/2)
  }
  ew <- log(ew/length(stat))
  mw <- mean(stat^2)
  qlr <- max(stat^2)
  # Values are from Table 3 in Stock and Watson (1998)
  # Test Statistic: Exponential Wald (EW)
  valew <- c(0.426, 0.476, 0.516, 0.661, 0.826, 1.111,
             1.419, 1.762, 2.355, 2.91, 3.413, 3.868, 4.925,
             5.684, 6.707, 7.690, 8.477, 9.911, 10.693, 12.024,
             13.089, 14.440, 16.191, 17.332, 18.699, 20.464,
             21.667, 23.851, 26.538, 26.762, 27.874)
  # Test Statistic: Mean Wald (MW)
  valmw <- c(0.689, 0.757, 0.806, 1.015, 1.234, 1.632,
             2.018, 2.390, 3.081, 3.699, 4.222, 4.776, 5.767,
             6.586, 7.703, 8.683, 9.467, 10.101, 11.639, 13.039,
             13.900, 15.214, 16.806, 18.330, 19.020, 20.562,
             21.837, 24.350, 26.248, 27.089, 27.758)
  # Test Statistic: QLR
  valql <- c(3.198, 3.416, 3.594, 4.106, 4.848, 5.689,
             6.682, 7.626, 9.16, 10.66, 11.841, 13.098, 15.451,
             17.094, 19.423, 21.682, 23.342, 24.920, 28.174, 30.736,
             33.313, 36.109, 39.673, 41.955, 45.056, 48.647, 50.983,
             55.514, 59.278, 61.311, 64.016)
  lame <- NA
  lamm <- NA
  lamq <- NA
  # Median-unbiased estimator of lambda_g for given values of the test
  # statistics are obtained using the procedure described in the
  # footnote to Stock and Watson (1998) Table 3.
  if (ew <= valew[1]) {
    lame <- 0
  } else {
    for (i in 1:length(valew)-1) {
      if ((ew > valew[i]) & (ew <= valew[i+1])) {
        lame <- i-1+(ew-valew[i])/(valew[i+1]-valew[i])
      }
    }
  }
}

R-Code 6: median.unbiased.estimator.stage2.R