ESCoE Economic Statistics Centre of Excellence

A collaboration with Office for National Statistics

escoe.ac.uk @ESCoEorg

Censored Density Forecasts: Production and Evaluation

James Mitchell (Federal Reserve Bank of Cleveland)

Starting at 11.30 AM

ESCoE ECONOMIC MEASUREMENT WEBINARS

Censored Density Forecasts: Production and Evaluation[1]

James Mitchell[2] and
Martin Weale[3]

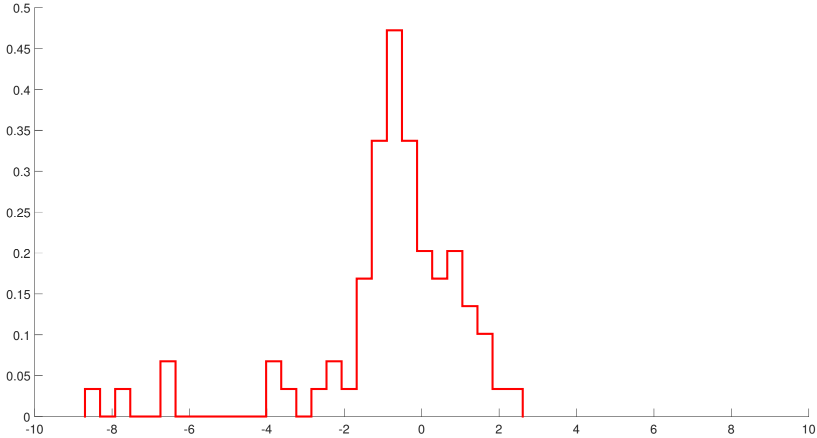
ESCoE Research Seminar

25 March 2021

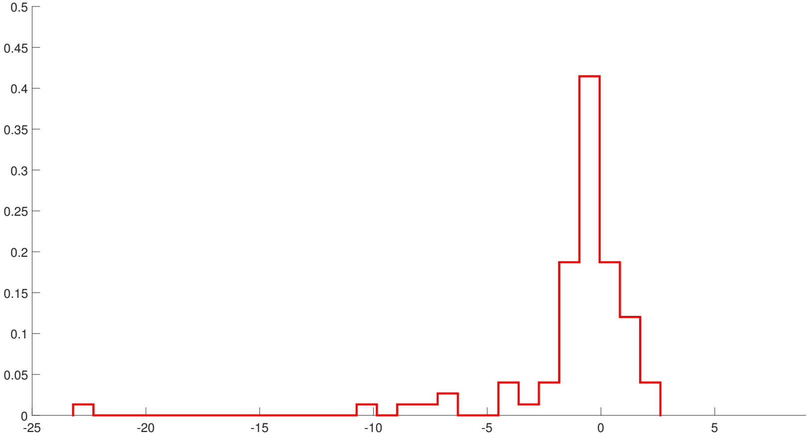
[1]Portions of this research were conducted under independent contract for the Federal Reserve Bank of Cleveland; the views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System

[2]Federal Reserve Bank of Cleveland; Economic Statistics Centre of Excellence
[3]King's College London; Economic Statistics Centre of Excellence

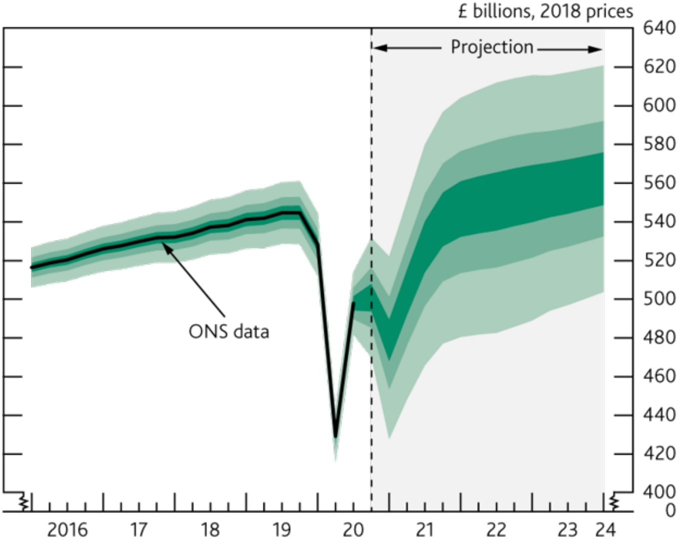
The unexpected happens: the MPC's historical forecast errors for 2-year-ahead GDP growth (% p.a)



The unexpected continues to happen...



Latest MPC Fan Chart for GDP: 90% best-critical-region



Quantifying uncertainty after shocks

- Assessments of future uncertainties are commonly informed, at least in part, by monitoring past forecast errors
- As Reifschneider and Tulip (2019 IJF) review, this is the general approach to gauging forecast uncertainty at the Federal Reserve, ECB, Bank of England, Reserve Bank of Australia, Bank of Canada and the Riksbank
- Practice varies across institutions
  - MPC use a two-piece normal parameterisation
  - FRB emphasise a 70% interval forecast, based on an underlying Gaussianity assumption
- If and how should extreme errors or outliers be accommodated when producing and evaluating probabilistic forecasts from historical forecast errors?

Allowing for unknown tail probabilities

- In May 2010 BoE *Inflation Report* stated, for the first time, that "on the remaining 10 out of 100 occasions GDP growth can fall anywhere outside the green area of the fan chart"
- MPC does not have to place equal probability in each tail
- The MPC is producing, what we call, a *censored* density forecast

- We can also interpret the confidence intervals published by FRB, and others, around their forecasts as ignoring the magnitude but not the frequency of outer observations

Censored density forecasts

- This paper develops methods for the production and evaluation of censored density forecasts
- Censored density forecasts quantify forecast risks in a middle region of the density covering a specified probability, but ignore the magnitude but not the frequency of outlying observations
- Places centre-stage the distinction between known and unknown unknowns, between Knightian uncertainty and risk
- Fits with ideas in Orlik and Veldkamp (2014, NBER) that economic agents know more about the probabilities of everyday events than (black swan) events in the tails of a distribution, given these are rarely observed

Production of censored density forecasts

- We consider how to estimate the parameters of the distribution of past forecast errors when censoring
- We do so allowing for:
  - Fat-tails: extreme events appear to happen more frequently than implied by a normal distribution
  - Skewness: forecast asymmetries emphasised post-GFC; e.g. see Adrian, Boyarchenko and Giannone (2019 AER)
- Is the evidence for fat-tails and skewness influenced by how extreme forecast errors are treated?

Skewed distributions

- Consider a general family of skew distributions as defined in Arellano-Valle et al. (2005 J.Stat.Plan.Infer)
  - Like the two-piece normal, these involve joining two distributions - but not necessarily normal with different scale (and perhaps shape, see Rubio and Steel, 2015) parameters
- A leading density within this family, that we focus on, is the two-piece t distribution described by Fernandez and Steel (1998 JASA)
- Focus on ML estimation

A fat tailed generalisation of the 2PN: the 2Pt

- The density function of the two-piece t distribution is

$$f(y_t) = \frac{2}{\sigma(\gamma + 1/\gamma)} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{1/2}} \left[1 + \frac{(y_t - \mu)^2}{\gamma^2 \nu \sigma^2}\right]^{-(\nu+1)/2} \quad \text{if } y_t < \mu$$

$$f(y_t) = \frac{2}{\sigma(\gamma + 1/\gamma)} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{1/2}} \left[1 + \frac{\gamma^2(y_t - \mu)^2}{\nu \sigma^2}\right]^{-(\nu+1)/2} \quad \text{if } y_t \geq \mu$$

- The mode of the distribution is μ but is no longer the same as the mean, unless $\gamma = 1$
- The probability mass to the left of the mode is $\gamma^2/(\gamma^2 + 1)$ while that to the right of the mode is $1/(\gamma^2 + 1)$ so with $\gamma < 1$ the distribution is skewed to the right and with $\gamma > 1$ it is skewed to the left

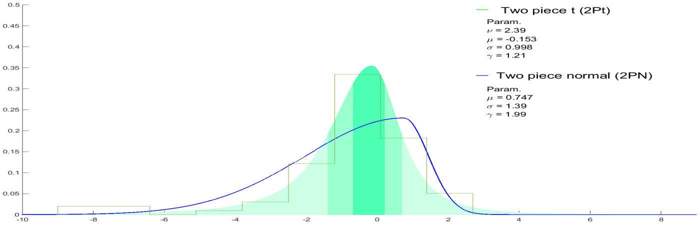
The 2Pt density (cont.)

- The log-likelihood function of a sequence of observations y_t, $t = 1, \ldots, T$, is

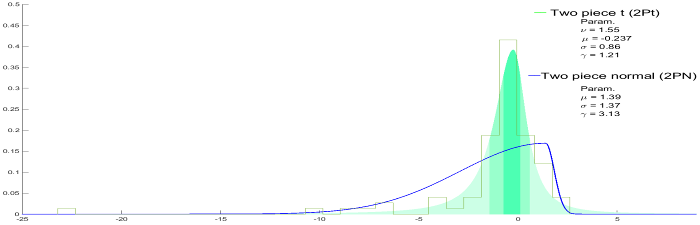
$$\log L = T \ln \left(\frac{2}{\sigma\left(\gamma + 1/\gamma\right)} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left(\pi\nu\right)^{1/2}} \right) + \sum_{t=1}^{T} \mathbf{I}\left(y_t - \mu\right) \ln \left[1 + \frac{\gamma^2 (y_t - \mu)^2}{\nu\sigma^2} \right]^{-(\nu+1)/2} + \sum_{t=1}^{T} \mathbf{I}\left(\mu - y_t\right) \ln \left[1 + \frac{(y_t - \mu)^2}{\gamma^2 \nu\sigma^2} \right]^{-(\nu+1)/2}$$

where $\mathbf{I}(x)$ is an Indicator function, $\mathbf{I}(x) = 1$ if $x \geq 0$ and $\mathbf{I}(x) = 0$ if $x < 0$

90% (uncensored) 2Pt fitted to MPC GDP errors

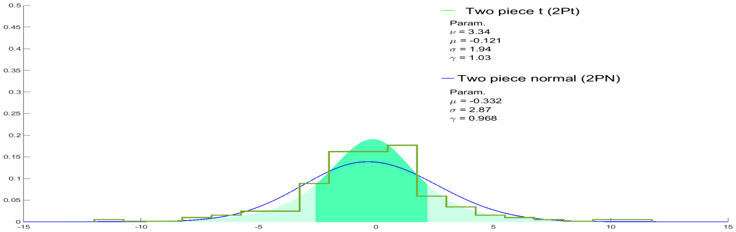


(a) Pre-pandemic

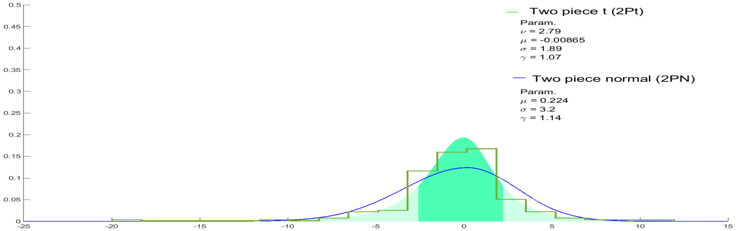


(b) Post-pandemic

70% uncens 2Pt fitted to FRB GDP errors (1-yr-ahead)



(a) Pre-pandemic



(b) Simulated post-pandemic

Censored 2Pt Distribution

- For these fitted distributions (so assuming the density applies to all observations including those in the censored region) we can estimate the 90% best-critical-region (BCR)
- But we don't want to make any distributional assumption about the censored observations
- If the lower cut point is y_L and the upper cut point is y_U, the censored log likelihood is:

$$\log L_A^C = \left\{ \begin{array}{ll} \log(F(y_L) & \text{if } (y \ < y_L) \\ \log L & \text{if } (y_L \leq y < y_U) \\ \log(1 - F(y_U)) & \text{if } (y > y_U) \end{array} \right\}$$

- Estimation of μ, σ, γ and ν becomes more difficult when the censor points are treated as endogenous (or proportionate) due to discontinuities (boundary problems) as movements in the BCR cut points place observations either in the censored or the uncensored region

Agnostic Censoring

- L_A^C assumes the forecaster has a view on whether points are likely to be in the upper or the lower tail of the distribution, notwithstanding that the density function within those tails is not specified
- A likelihood function which is completely agnostic as to whether observations are going to be above the upper cut point or below the lower cut point is

$$\log L_B^C = \left\{ \begin{array}{ll} \log(F(y_L) + 1 - F(y_U)) & \text{if } (y < y_L) \\ \log L & \text{if } (y_L \leq y < y_U) \\ \log(F(y_L) + 1 - F(y_U)) & \text{if } (y > y_U) \end{array} \right\}$$

Full ML Estimation is Degenerate (i)

- The BCR satisfies

$$f(y_U, \beta) - f(y_L, \beta) = 0$$

We might then expect to solve the problem

$$\underset{\beta}{Max} \sum \log L^C(y_i, \beta) + \lambda_1 \left(F(y_U, \beta) - F(y_L, \beta) - 0.9 \right)$$
$$+ \lambda_2 \left(f(y_U, \beta) - f(y_L, \beta) \right)$$

Full ML Estimation is Degenerate (ii)

- Suppose that we set $\mu = y_A$; some arbitrary observation, and σ is so small that all other observations are in the censored region - with T_1 observations below y_L and T_2 observations above y_U. Then, when the constraints are met

$$\log L^C = T_1 \log F(y_L) + \log f(y_A, \beta) + T_2 \log(1 - F(y_U))$$

But

$$\log L^C \to \infty \text{ as } \sigma \to 0$$

y_L and y_U change as σ shrinks but $F(y_L)$ and $F(y_U)$ do not

- The "regularity conditions" needed to prove asymptotic normality of ML estimators are violated when the support of the density depends on its parameters; e.g., see Woodroofe (1972) and Smith (1985)

A censored 2Pt distribution: a fixed-point estimator

- In large samples, estimates (for μ, σ, γ and ν) produced by maximising $\log L^C$, with *fixed censor points,* are independent of the censor points, provided all the uncensored observations are genuinely drawn from the specified distribution
- In finite samples, we propose an iterative fixed point estimator:

Step 1: $\beta_{r+1} = \max_{\beta} \sum \log L^C(y_t, \beta, y_{L,r}, y_{U,r})$

Step 2: compute BCR of $f(y_t \mid \beta_{r+1}) \Rightarrow y_{L,r+1}, y_{U,r+1}$

where $\beta = [\nu, \mu, \sigma, \gamma]$ and where we iterate in r between Steps 1 and 2 to minimise $(y_{L,r+1} - y_{L,r})^2 + (y_{U,r+1} - y_{U,r})^2$

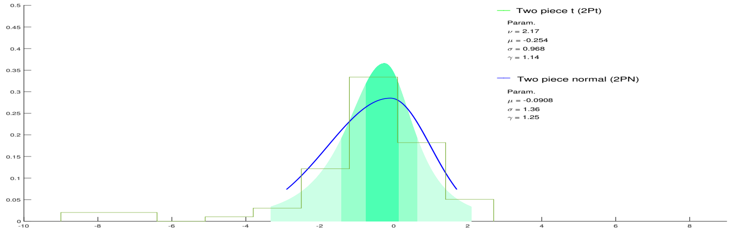
Monte Carlo

- Results confirm good performance of the fixed point estimator in large samples
- In small-samples
  - The standard errors of the parameter estimates are larger
  - Accuracy deteriorates the more highly skewed the true data: increased risk of divergent estimates for γ
  - L_A^C outperforms L_B^C
  - We find it can be helpful to use a penalised estimator

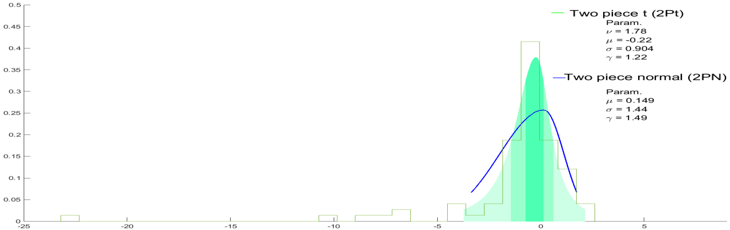
$$PL_A^C(y_i, \beta) = \sum \log L_j^C(y_i, \beta) - \frac{1}{2} P_\lambda(|(\gamma - 1)|)$$

  and $P_\lambda(|(\gamma - 1)|)$ is a nonnegative penalty function. We use the Lasso penalty, $P_\lambda(|(\gamma - 1)|) = \lambda\,|(\gamma - 1)|$ where λ is a tuning parameter

90% censored 2Pt fitted to MPC GDP errors

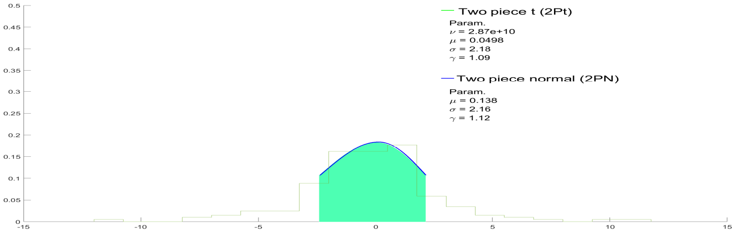


(a) Pre-pandemic

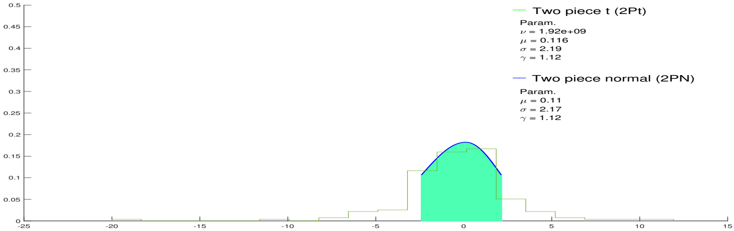


(b) Post-pandemic

70% censored 2Pt fitted to FRB GDP errors (1-yr-ahead)



(a) Pre-pandemic



(b) Simulated post-pandemic

Implications of censoring

- Parameter estimates change when we censor - the tails do behave differently so there is a danger that shocks in the tails do “wag” the whole density (Kozlowski, Veldkamp and Venkateswaran, 2020 JPE)

- For the UK: there is less evidence for skew when we censor
- For the US: there is less evidence for fat-tails when we censor

- We shall also see, censoring delivers forecast error densities which exhibit fewer temporal instabilities than when an uncensored Gaussian density is used

Evaluation of censored density forecasts

- For uncensored densities, typically calibration tests look at the PITS
- PITS are uniform over (0,1) for well-calibrated densities
- Point of departure for censored densities is that tail probabilities and hence tail PITS are unknown
- A well-calibrated censored density delivers uniform PITS in the uncensored region and satisfies a Christoffersen (1998 IER)-type coverage rate condition in the censored region
- Need a joint test

PITS evaluation of censored density forecasts I

- We develop, and assess via Monte Carlo, a moments-based test, generalising Knüppel (2015 JBES) + Rossi & Sekhposyan (2019 JoE)
- Is suitable for multi-step-ahead forecasts
- PITs are defined as $z_t = F(y_t)$
- Given censoring, define PIT thresholds $z_{L,t} = F(y_{L,t})$, $z_{U,t} = F(y_{U,t})$
    - e.g., $z_{L,t} = 0.05$ and $z_{U,t} = 0.95$ for 10% censoring with symmetric thresholds (about the mean)
- The censored density forecast $f(y_t)$ is well-calibrated when the subset of PITS, z_t^c, defined as

$$z_t^c = z_t \in [z_{L,t}, z_{U,t}] \tag{1}$$

is uniformly distributed between $(z_{L,t}, z_{U,t})$

PITS evaluation of censored density forecasts II

- Consider a censored form of the standardised PITS

$$v_t = H(z_t^c) = \sqrt{12/\left(z_{U,t} - z_{L,t}\right)^2}\left(z_t^c - 0.5(z_{L,t} + z_{U,t})\right) \qquad (2)$$

- Under the null of calibration, v_t is uniformly distributed with an expectation of 0, variance of 1, skewness of 0 and kurtosis of 1.8, resp.
- Defines $N = 4$ moment conditions à la Knüppel (2015 JBES), where a HAC estimator accommodates serial dependence for multi-step-ahead densities, and delivers a χ_N^2 test based on the test statistic

$$\hat{\beta}_{s_1, s_2, \ldots, s_N}$$

PITS evaluation of censored density forecasts III

- For censored densities, we add an additional moment condition for coverage:

$$D_t = w_t\left(z_t\right) - \left(1-\alpha\right). \tag{3}$$

The test statistic $\hat{\beta}_c \xrightarrow{d} \chi_1^2$ then follows as the square of the standard normal test statistic of a sample proportion, i.e.

$$\hat{\beta}_c = T\left(\frac{1}{T}\sum_{t=1}^{T} D_t\right)^2 / (1-\alpha)\alpha \tag{4}$$

- An overall test for calibration of censored density forecasts is

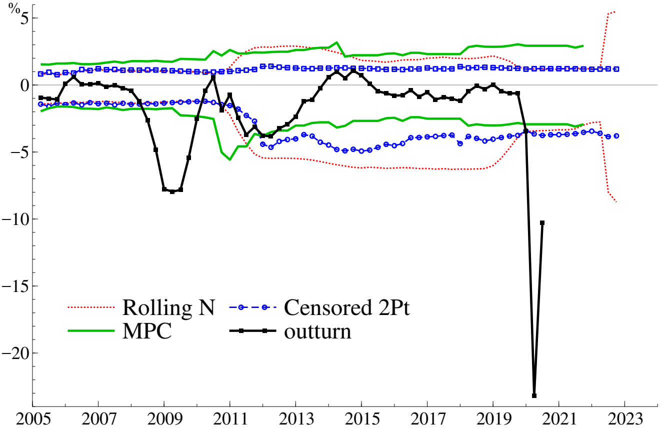
$$B = \hat{\beta}_{s_1,s_2,\ldots,s_N} + \hat{\beta}_c \xrightarrow{d} \chi^2_{N+1} \tag{5}$$

Out-of-sample

- Illustrate and evaluate behaviour of the proposed censored density forecasts in a real-time application using FRB and MPC historical forecasting errors for GDP growth
- Compare with uncensored Gaussian density forecasts fitted to rolling samples of 20 years (US) and 10 years (UK) - as used in practice at FRB and BoE
- Data revisions and the choice of outturn have a big effect, especially in the UK
- Focus today on visual examination of the densities, with the paper containing evaluation tests

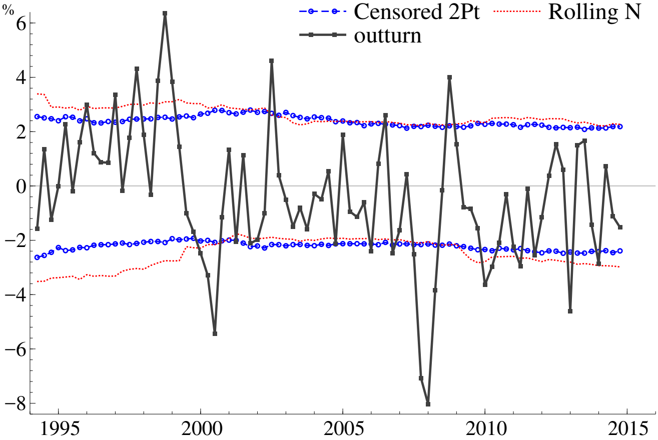
Temporal evolution of the censored intervals: UK

Figure: Real-time UK GDP 90% ex ante forecast error intervals and ex post forecast errors (2 year ahead forecasts of outturns from 2005q1-2020q4; second vintage defines GDP outturns)



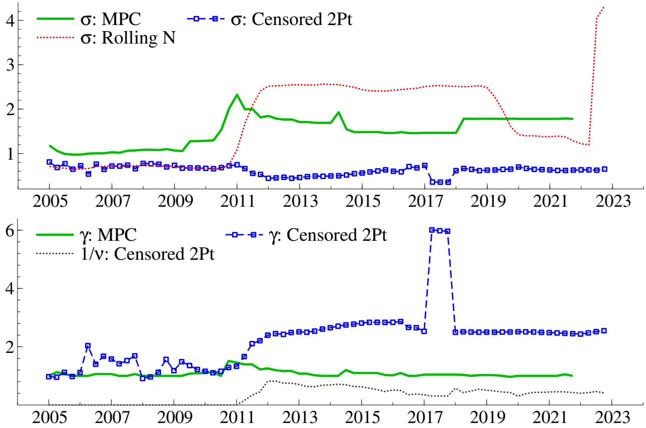
Temporal evolution of the censored intervals: US

Figure: Real-time US GDP 70% ex ante forecast error intervals and ex post forecast errors (1 year ahead forecasts of outturns from 1994q2-2014q4; second vintage defines GDP outturns)



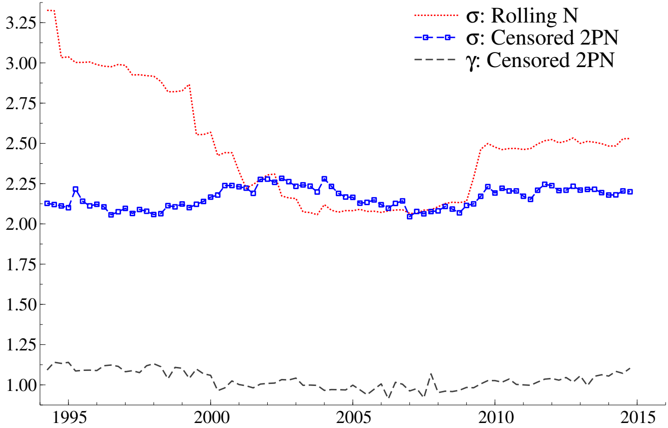
Temporal evolution of the densities: UK

Figure: Real-time and MPC parameters for the UK GDP error density



Temporal evolution of the densities: US

Figure: Real-time parameters for the US GDP error density



Conclusion: Use Censored Density Forecasts

- We propose the (wider) use of censored density forecasts
  - Outlying forecast errors reflect (realised) unknown unknowns or events not expected to recur that should be censored before quantifying known unknowns
  - Alternative way of quantifying forecast risks after shocks like the GFC and COVID-19
  - Motivation comes from the censored nature of the MPC's fan charts and the apparent preference of many central banks for interval forecasts
- We develop the tool-kit: propose methods for the production and evaluation of density forecasts acknowledging that the outlying observations may be drawn from a different but unknown distribution to the inner observations
- Censoring delivers error densities that are both more Gaussian and exhibit fewer temporal variations - but that can be well-calibrated even in the presence of large shocks