Density Forecasting: Growth at Risk
Part I. Conceptual Framework

Romain Lafarguette (MCM)

IMF Institute for Capacity Development

14 September 2021

The views expressed in this presentation are those of the author and do not necessarily represent the views of the IMF, its Executive Board, or its management.

Table of Contents

Context

- Macrofinancial model initially developped by **Tobias Adrian et al.** at the NY Fed *(Vulnerable Growth, AER April 2019)*
- Operationalized for policy applications: cf. IMF WP 19/36 on *Growth at Risk: Concept and Application in IMF Country Surveillance (2019)*
- IMF GaR is coded in Python, first used internally. Later, public release of a user-friendly Excel interface (no knowledge of Python required)
- Open-source code, publicly released on Github. Internal and external review
- Internal use: GaR used for Article IV, FSAP scenario calibrations, technical assistance, etc.
- External use: More than 30 countries are using the tool now

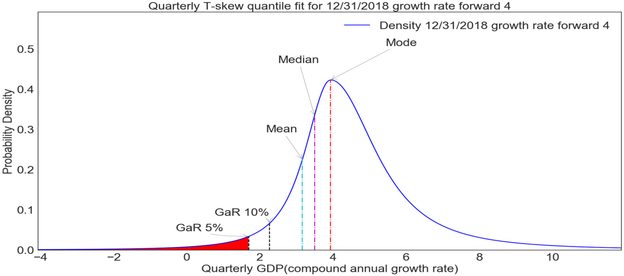
Growth at Risk: Overview

GaR is a reduced-form model

- Forecast a conditional **distribution** of future GDP growth, at different horizons
- Based on quantile regressions using a set of macro and financial regressors, customizable for each countries
- Useful to estimate potential ”tail” realizations of GDP (e.g. 5th percentile)
- **Probabilistic assessment**: associate a GDP potential realization from a given risk level (e.g. 5%)

Final output

Given current macrofinancial conditions, distribution of economic growth in the future



For $\tau \in \Omega \ : \ y_{t+h} = \alpha^{\tau} + \beta_1^{\tau} X_{1,t} + \cdots + \beta_k^{\tau} X_{k,t} + \epsilon_{t+h}^{\tau}$

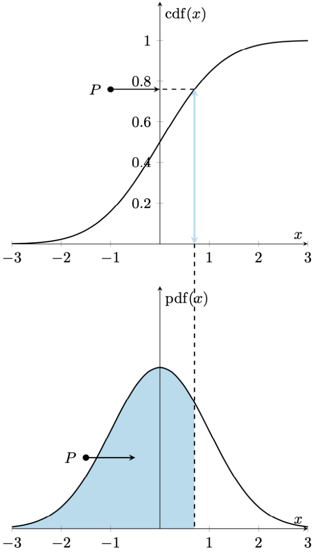
where Ω is a set of quantiles, e.g. 5th, 25th, median, etc.

Refresher: PDF and CDF for Continuous Variables

- The cumulative distribution function (CDF) gives the probability of X to be lower than a given value

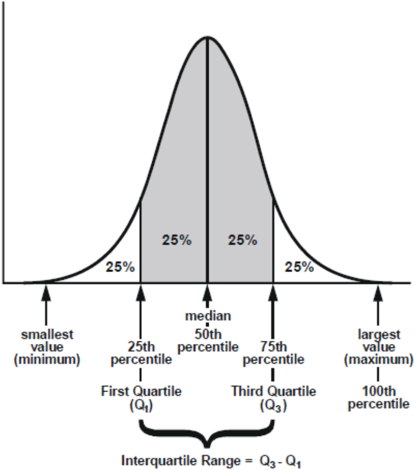
  $$F_x(a) = \mathbb{P}[X \leq a]$$

- The probability density function (PDF) is the derivative of the CDF $f_X(a) = \frac{\partial F_X(a)}{\partial a}$
  - Probability of the random variable falling within a given interval: $\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(y)dy$



Refresher: Quantiles

- Quantiles are cut points dividing a PDF into intervals of same probability
- Quantile Q at probability p: $\mathbb{P}[X \leq Q] = p$
  - The quantile function $\mathcal{Q}(p) = \inf\{y \in \mathbb{R} : p \leq F_X(y)\}$ is the **inverse cumulative distribution function**
- The median is the quantile at 50% (half observations below)
- Distributions can be characterized by their PDF, CDF or quantile functions (among other)



Concept

Idea

Forecast the **conditional density** of future GDP growth Y_{t+h} at a given horizon h

- **Density**: Estimate the *probability distribution function (pdf)* $f_Y : P[Y_{t+h} \in \Theta] = \int_{\Theta} f_Y(y)dy$, for Θ a measurable interval
- **Conditional**: density depends on a regressors set, observed at period t or in the past: $f_{Y_{t+h}|X_t}$

Intuition

Based on the current set of macrofinancial conditions observed recently, what is the probability that future GDP growth will be around 2 percent in the next year?

From VaR to GaR

- In finance, the Value at Risk (VaR) is defined as ”the lowest value of a portfolio with a given probability, based on historical or conditional data”
- The idea of Growth at Risk is to transpose the concept in macroeconomics
- GaR at 5% is the future GDP value so that only 5% of the potential growth realizations will be below it: represents a ”lower bound” in terms of risk
- Note that the tool also estimates the associated concepts with VaR, such as **Expected Shortfall** (average of all losses which are greater or equal than VaR)

Originality from a Statistical Perspective

- Estimating conditional distribution is not new: copulas (multivariate pdf) and non-parametric approaches are popular in finance

- However, interpretability is difficult and the estimation requires large datasets

GaR Contribution: a VaR Model for Macro Data

- Simple (linear) and parsimonious model to estimate density forecasts
- Based on regressions: easy to intepret, familiar to economists
- Flexible parametric approach: can be consistent with standard Article IV point forecasts
- Robust to outliers and can be used on macro quarterly data

Empirical Strategy

1. **Data partitioning**: aggregate a large number of variables into few regressors, using either unsupervised or supervised technics
2. **Quantile regressions**: Estimate the forecasting equation

   $$y_{t+h} = \beta^{\tau} X_t + \epsilon_t^{\tau}$$

   for different quantiles at probability τ and a given horizon h
3. **Parametric fit**: minimize the distance between the **empirical conditional quantiles** $\hat{Q}_{y_{t+h}}(\tau|X_t) = \hat{\beta}^{\tau} X_t$ and the theoretical quantiles from a parametric family

Empirical Strategy in Plain English

- From past data, find the relationship over time between a few number of components in t and growth in $t+h$.
- This relationship is "quantile dependent": e.g. housing prices might not matter in normal times, but might do in bad times
- For a given date, X_{2019}, use $\hat{\beta}^{\tau}$ to infer the quantiles of future growth $Q_{Y_{2020}}(\tau|X_{2019}) = \hat{\beta}^{\tau} X_{2019}$
- $\hat{\beta}^{\tau}$ is long-term: statistical relationships over time. X_{2019} is real-time: conditions the forecast to current macrofinancial conditions
- From a discrete set of quantiles, infer a "smooth" distribution, parametrized from a known family distribution

Data Requirements

Rule of thumb

- GaR is based on quantile regressions: needs more points than standard OLS to be accurate
- Rule of thumb: strict minimum is at least 60 points (15 years of quaterly data), accurately measured. I noticed that around 100 data points (25 years of quarterly data) is a comfortable level
- GaR is about crisis: the sample should contain some crisis episodes to be meaningful. Can be a problem for some countries (e.g. China). Work in progress: panel estimations
- **Ergodicity** matters: in case of structural breaks, if the economy radically changed, then past relationships might not be informative for future risks

Table of Contents

1st Step: Data Partitioning and Dimensionality Reduction

Idea

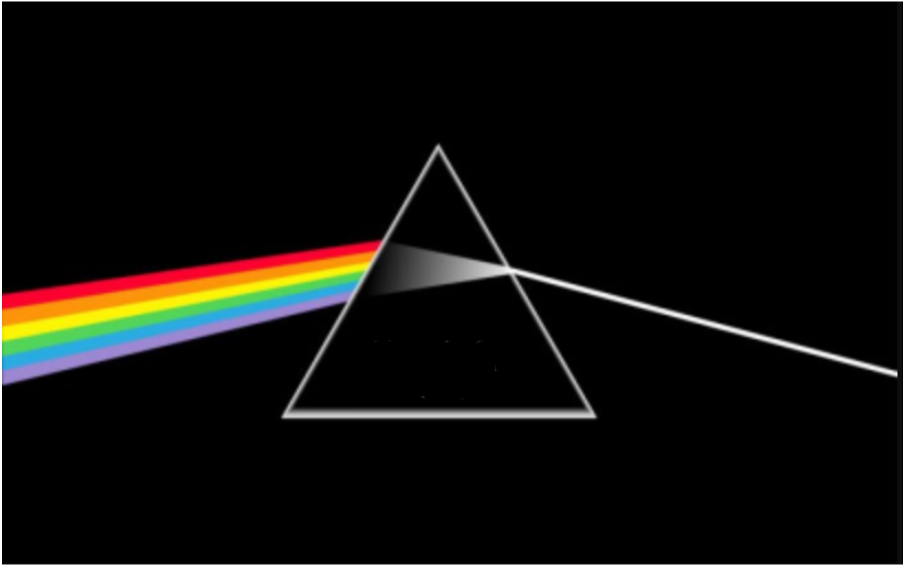
Extract the information from a large set of macro and financial variables and summarize it into few components

This step is useful because:

- Having too many variables in a regression increases parametric noise and risk of overfit
- Some individual variables can be noisy: extracting a common trend/component improves accuracy
- Attrition issue: some variables might be available only over a recent time period

PS: this step is not necessary and is not in the original paper by Adrian et al. (AER, 2019)

Concept of Dimensionality Reduction



Source: www.neuraldesigner.com

Unsupervised vs. Supervised Data Reduction Methods

The tool currently allows to run three data reduction methods:

1. **Unsupervised**: Principal Component Analysis (PCA)
2. **Supervised on categorical**: Linear Discriminant Analysis (LDA)
3. **Supervised on continuous**: Projection on Latent Structures (PLS), also called partial least squares

Tip

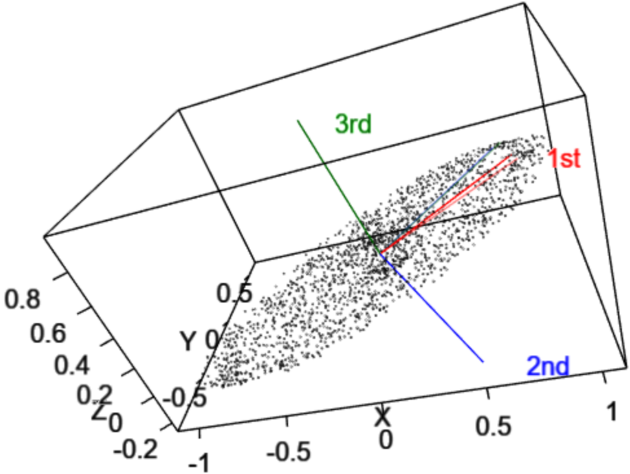
Although PCA is more familiar to economists, I found PLS to perform better and has also more intuitive features.
LDA is available in the tool as a legacy method but I don't recommend using it anymore

Principal Component Analysis: Derivation

Mathematically, for the first component:

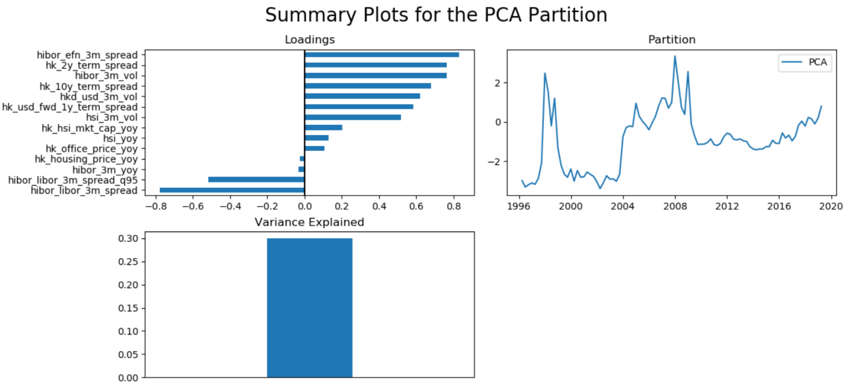
1. First, need to scale X_N so that the variables are centered, and with the same unit
2. PCA is a linear projection, so it looks for a set of weights $w*$ maximizing the variance of X_N, $\quad w* = \text{argmax}\{w^T X^T X w\}$
3. By construction, $X^T X$ is semi-positive definite (it's a variance). A standard result is that the eigenvector associated with the largest eigenvalue solves the optimization program above
4. Intuitively, both eigenvectors and eigenvalues are providing us with information about the distortion of a linear transformation:
    - the eigenvectors are basically the direction of this distortion
    - the eigenvalues are the scaling factor for the eigenvectors that describing the magnitude of the distortion.

PCA applied to an ellipsoidically-shaped points cloud



Source: http://joyofdata.de

GaR Tool Output: PCA First Component and Loadings (example)



Source: IMF Staff

Projection on Latent Structures (PLS) - also called Partial Least Squares -

Intuition

- PLS is a regression-based method where both the Y and X are projected onto new subspaces (this is why PLS is not ”ordinary” least square)
- Note that Y can be multivariate, as the difference of an OLS
- PLS is used to find the fundamental relations between X and Y, i.e. a **latent variable** approach to modeling the covariance structures between X and Y
- A PLS model will try to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space.

PLS Derivation

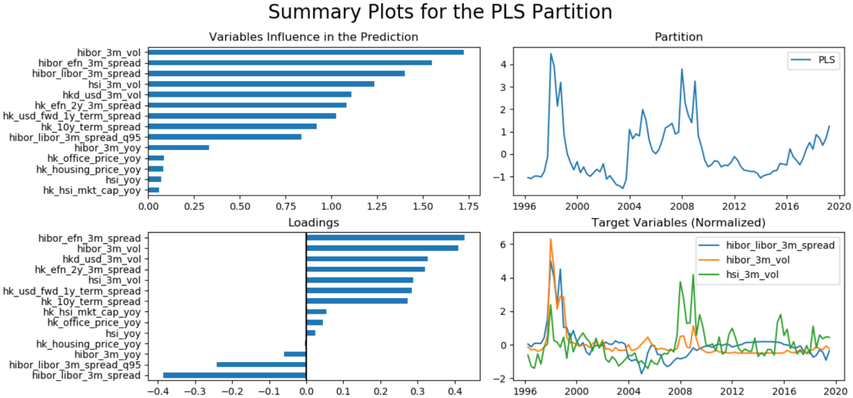
1. Start from the standard linear model $Y = XB + \epsilon$
2. Project Y and X such that:
   - $X = TP^T + E$ where T is the projection matrix and P the orthogonal loadings
   - $Y = UQ^T + F$ where U is the projection matrix and Q the orthogonal loadings
3. The objective is to find T, P, U, Q so that to **maximize the covariance between the "latent" structure** T and U

The estimation is done via an iterative algorithm

Why PLS is interesting for GaR

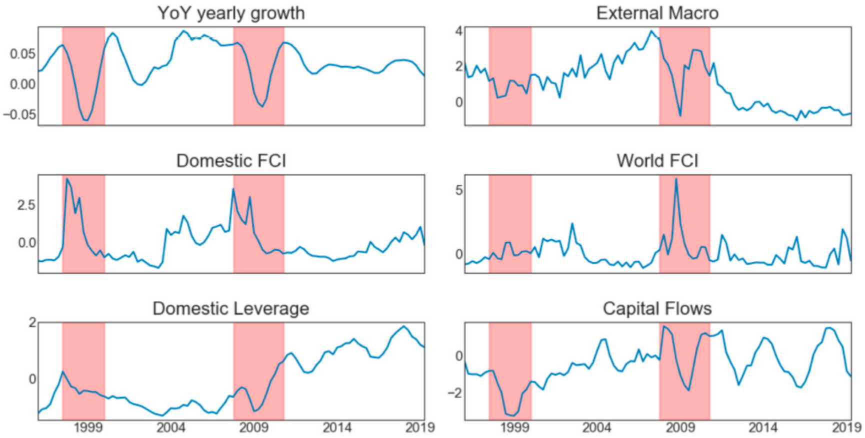
- PLS regression is particularly suited when the matrix of predictors X has more variables than observations, which often happens with economic data
- And also when there is multicollinearity among X values (often the case with financial variables)
- Useful for interpretation: the PLS aggregate will follow one or more target variables (e.g. term spread)
- Important to anchor the extrapolation backward (see after)

Example of PLS: Domestic FCI with Multiple Targets



Source: IMF Staff

Model with Six Partitions



Source: IMF Staff

Data Partitioning in Practice

Important to keep in mind

- It is absolutely crucial that the regressors generated with the partitioning tool **make economic sense**
- They should capture the largest movements, be in line with the historical developments in the country, etc.
- Spend time to customize the partition groups, add or remove variables, detrend them, etc.
- Avoid to partition with non-stationary variables: the partition might end-up as a trend, which will create problems for the quantile regressions (see after)
- Uninformative regressors reduce the accuracy and relevance of the model

Data Partitioning in Practice : Rule of Thumb

1. As a general rule, don't use more than 5 or 6 partitions. Else the model will overfit and add too much parametric noise
2. Don't throw a lot of heterogeneous variables in one partition: else the partition will be difficult to interpret
3. Distinguish between fast-impact variables (typically spreads) and slow-moving ones (quantities, leverage)
4. Distinguish when possible between domestic and foreign variables: their dynamics are often different

Example: Indonesia

Table 1. Indonesia: List of Macrofinancial Variables for GaR Analysis	
Financial conditions	• Real long-term interest rates • Term spread • Sovereign spreads • Corporate spreads • CEMBI market cap • Equity returns • Change in foreign exchange rate • VIX
Macrofinancial vulnerability	• Credit growth • Credit gap • NPL ratios in the banking system • Ratio of external debt to GDP • Ratio of currency account balance to GDP • House price growth
External condition	• China's GDP growth

Source: Indonesia IMF Selected Issues Papers (2019)

Example: Albania

Table 2. Albania: List of Variables Entering each Partition

Domestic Financial Conditions	Domestic Leverage	Main Trading Partners Macro Conditions	Euro Area Financial Conditions	World Financial Conditions
One week repo rate	Credit to GDP	Italy growth rate	Euro Area VIX	VIX
One week repo rate, first difference	Loans to deposits ratio	Italy unemployment rate	10 year Italian sovereign rate	Oil prices
Volatility of the repo rate	NPLs to total loans ratio	Greece growth rate	Euro Area headline inflation	Oil prices, implied volatility
Inflation rate, yearly	Tier 1 capital as percent of risk-weighted assets	Greece unemployment rate	One week Euribor	US Treasuries yield, implied volatility
Deviation from the interest rate corridor	Share of non-resident liabilities			Bonds flows to Central Europe
Volatility of the 2 year sovereign bond				
Exchange rate volatility				

Table of Contents

2nd Step: Quantile Regressions vs. OLS

Intuition

Most of the macrofinancial literature uses OLS regressions, focusing on marginal impact on the **mean.** Density models allow to think about the **full distribution**

- Much richer policy analysis: some variables might not matter to explain the dynamic on average, but might matter a lot in crisis time (example: housing prices)
- The density estimation model is a natural framework: looking at **conditional distribution** and their implications
- However, at this stage, I didn't solve the problem of identification: shocks to the partitions are supposed to be reduced-forms (no joint dynamic among partitions)

Quantile Regressions: Intuition

- An OLS regression fits **the conditional mean**:

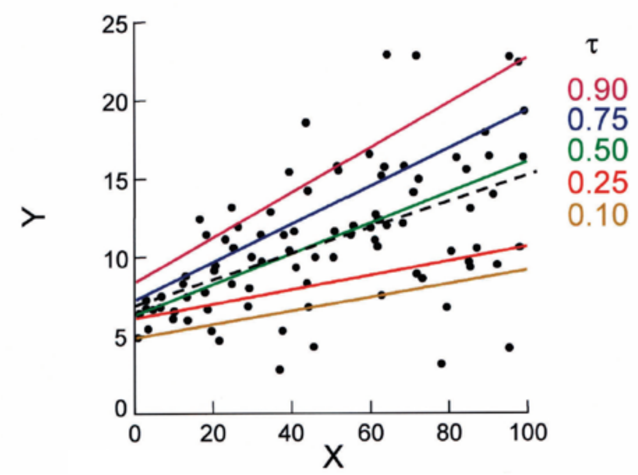
$$Y = \beta^{OLS} X + \epsilon \quad \Longleftrightarrow \quad E[Y|X] = \hat{\beta}^{OLS} X$$

- For a given probability τ, a quantile regression fits the **conditional quantile at** τ:

$$Y = \beta^{\tau} X + \epsilon^{\tau} \quad \Longleftrightarrow \quad Q_Y(\tau|X) = \hat{\beta}^{\tau} X$$

- $\hat{\beta}^{\tau}$ is the marginal effect of X on the conditional quantile of Y at probability τ
- Example: estimating a Mincer equation with quantile regressions. How adding one year more of education improves the income for rich people? For poor people? etc.

Quantile Regression vs. OLS



Source: www.datasciencecentral.com

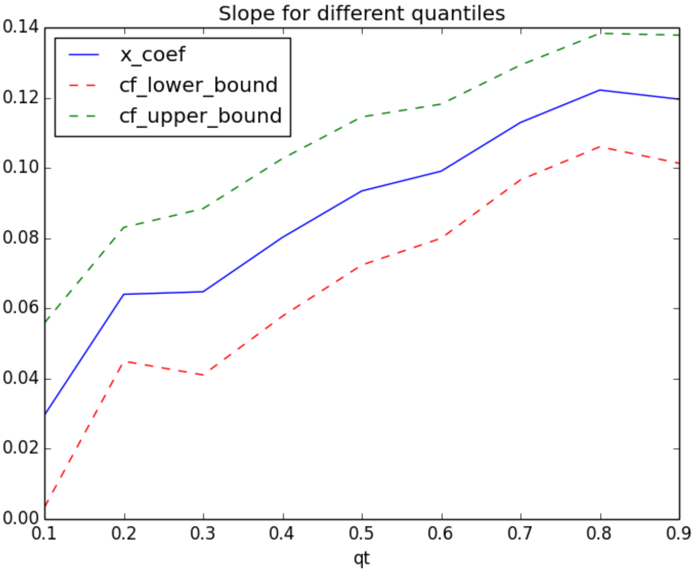
Linear Specification with Non-Linear Effects

- GaR uses linear specification for the quantile regression, for instance: $Y = \beta_1^\tau X_1 + \beta_2^\tau X_2 + \beta_3^\tau X_3 + \beta_4^\tau X_4 + \epsilon^\tau$
- However, because the model estimates the quantile regressions for different values of τ (e.g. 5%, 25%, 50%, etc.), the marginal impact of X on Y is non-linear: it varies with the distribution of Y: $\frac{\partial Q(Y,\tau)}{\partial X_1} = \beta_1^\tau X_1$

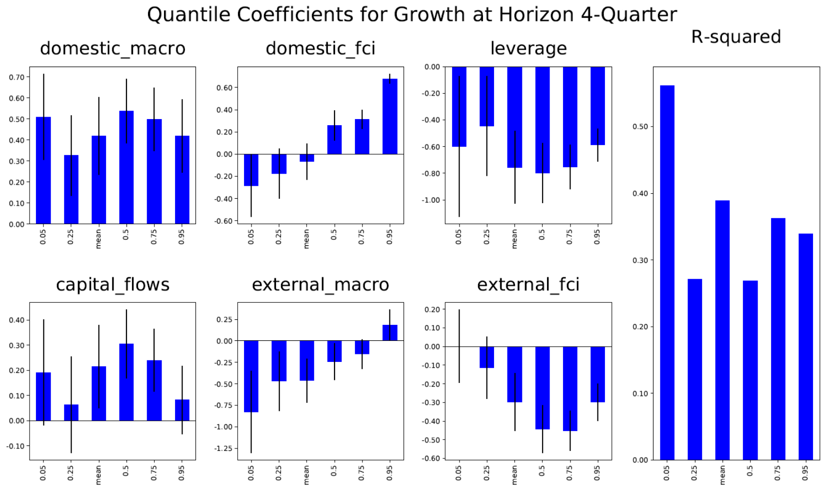
Important Remark

- A quantile regression is not a threshold regression: the variation is in the quantile of Y, not of X
- The interpretation of the coefficient is therefore: "how an increase by 1 unit **on average** of X impacts **the quantile** τ of Y"

Non-Linearities in Quantile Regressions Coefficients



Quantile Regressions Output



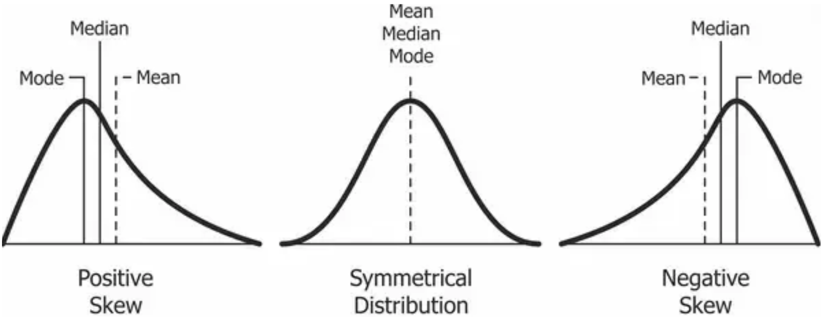
Source: IMF Article IV (2018)

Table of Contents

3rd Step: From Empirical Conditional Quantiles to Parametric Fit

- In principle, I could estimate the empirical cdf simply by interpolating and inverting the empirical quantile function (a good paper doing that is *Quantile Spacing*, *mimeo*, Schmidt and Zhu 2016)
- However, this might lead to very unsmooth distributions
- The idea is therefore to **parametrize a well-known distribution** so that its quantiles are as close as possible as the empirical quantiles estimated by GaR
- The tool uses a Tskew distribution for the fit, paramatrized with:
  1. Location (the mode)
  2. Scale (variance)
  3. Kurtosis ('fatness' of the tails) - also defined via the degrees of freedom
  4. Skewness ('asymmetry')

Skewed Distributions



Source: www.datasciencecentral.com

Pros and Cons of Parametric Fitting

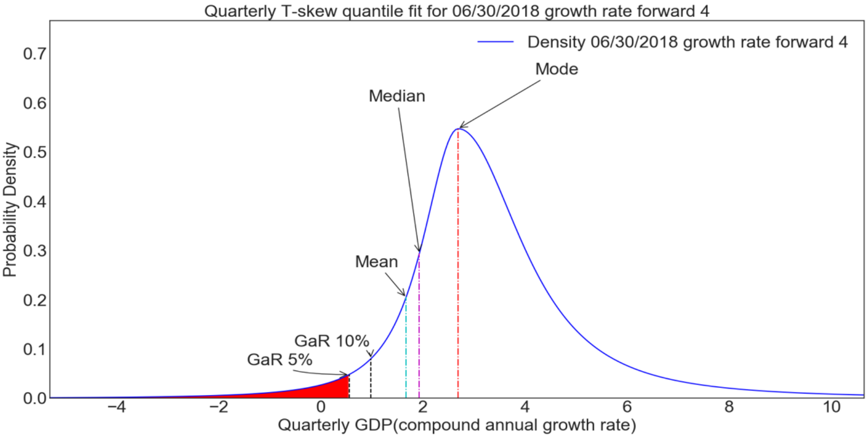
Pros

- Parsimonious way to summarize the information (4 parameters)
- Capture higher conditional moments, in particular skewness and kurtosis: **riche policy implications**
- Smooth the extreme tails
- t distribution widely used in finance (fat tails)

Cons

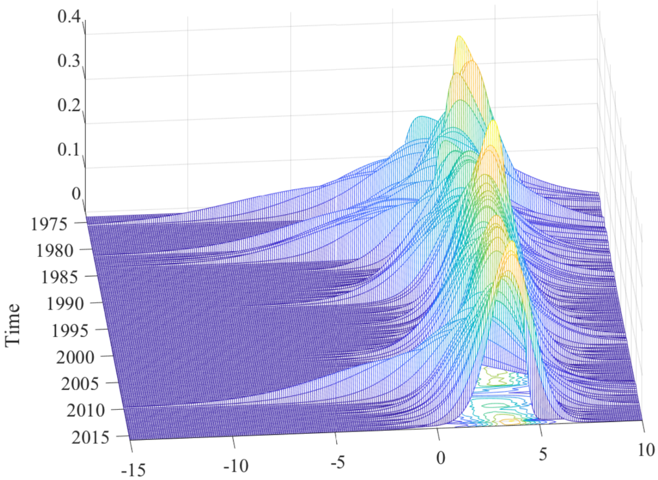
- Extra statistical layer with distribution fitting error, on top of the quantile estimation error
- Ignores potential bi-modality in the data (or more)
- No clear asymptotic properties

Fitted Distribution



Source: IMF Staff

GaR Time Series



Source:

Adrian et al. (2019)

Constrained Optimization

Consistence with baseline or WEO forecasts

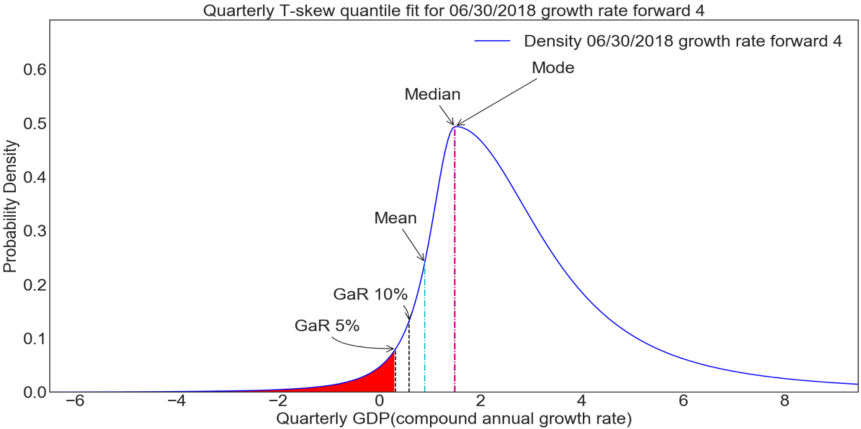
- GaR estimates the full distribution. However, often a central tendency forecast is available (WEO or authorities' forecast)
- The tool gives the possibility to forecast under the constraint that the mode of the density coincides with an ad-hoc value (decided by the user)

The program optimized under fixed location:

$$\text{argmin} \sum_{\tau \in \Theta} ||\text{TskQ}(\{\overline{\text{loc}}, \text{scale}, \text{skew}, \text{kurtosis}\}, \tau) - \hat{Q}_Y(Y_{t+h}, \tau)||^2$$

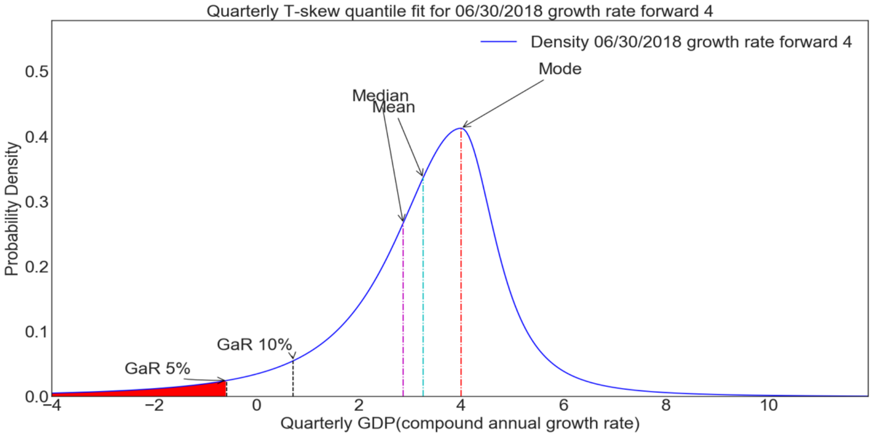
Note that the model is internally consistent: if the user imposes a mode far away from the unconstrained mode, then the full distribution will be highly distorted to accomodate user's assumptions.

Constrained Fit: on the Left



Source: IMF Staff

Constrained Fit: on the Right



Source: IMF Staff

Table of Contents

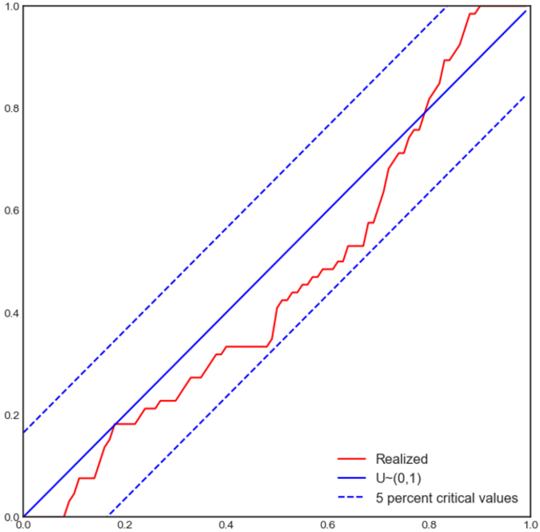
Probability Integral Transform Test (PIT)

- A probability integral transform is simply the evaluation of the cdf of a random variable (F_x) on its own values (X).
- Mathematically, the random variable $Y = F_X(X)$ is uniformally distributed

Intuition

- If the density forecast model is correctly specified, the PIT follows an IID uniform distribution on the unit interval
- The departure to the IID hypothesis can be quantified by the Kullback-Leibler (1951) information criterion
- Thus the test statistic measures the distance of a candidate model to the unknown true model.

Probability Integral Transform



Source: IMF Staff

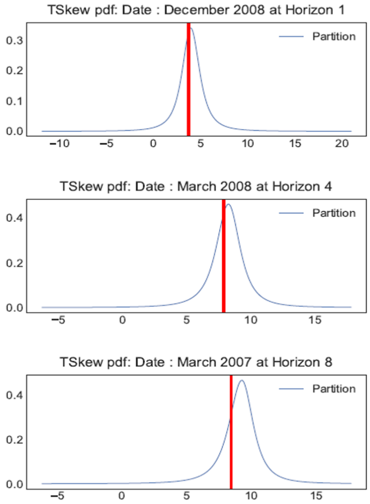
Scoring Tests

The literature on the topic is vast (cf. Timmerman et al., **Handbook of Economic Forecasting** 2013). One intuitive way is to use scoring rules, possibly asymmetric:

Intuition

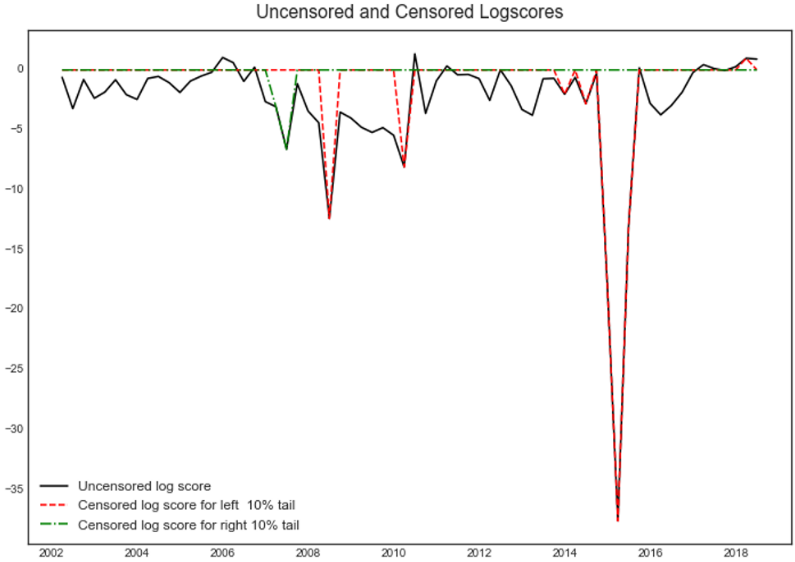
- Idea: was it the ex-ante probability of the ex-post realization?
- Scores are usually taken in log-form: $S^l(\hat{f}^t; y_{t+h}) = \log \hat{f}^t(y_{t+h})$
- Can give more weights to models which provide more accurate density in the tails (risk models) than in the central tendency

Ex-Ante Distributions and Ex-Post Realizations



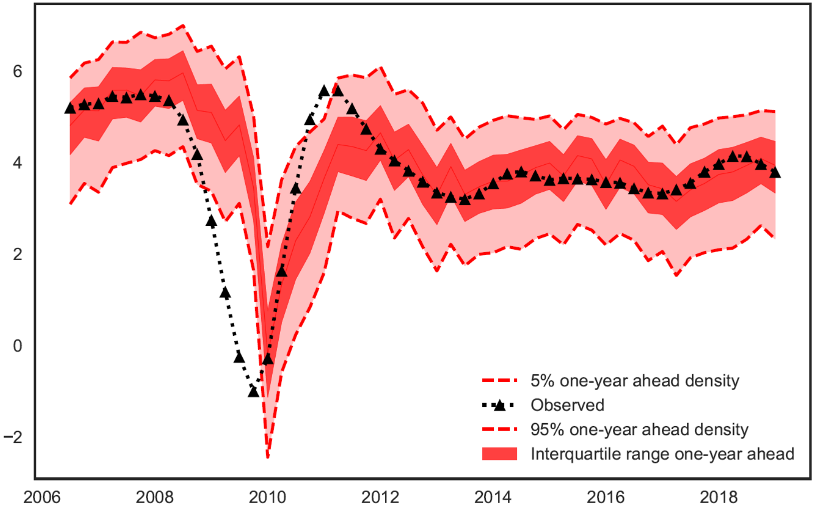
Source: IMF Staff

Logscore Across Time



Source: IMF Staff

GaR Ex-Ante and Ex-Post Assessment



Source: IMF Staff