Uncertainty Quantification using a Discrepancy Term with a Gaussian Process Prior: An Example from Macroeconomics

Ivo Tavares

ISEG Lisbon School of Economics and Management

itavares@iseg.ulisboa.pt

May the 19th, 2021

Keywords: Non-Parametric Bayesian Modelling, Machine Learning, State-Space Models, Uncertainty Quantification.

Outline

What's Uncertainty Quantification?

Accounting for different sources of uncertainty (Parameter, Structural, Interpolation, code, etc).

An example of UQ from Computer Modelling

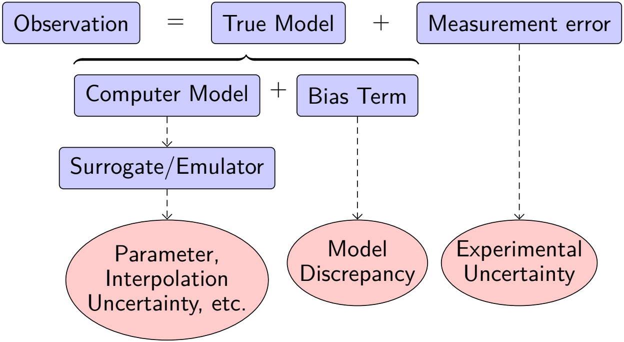

Figure: UQ in Computer Modelling

Structural Uncertainty / Model Discrepancy

Model Discrepancy may be present when:

- Using a simplified model, instead of the underlying model (e.g. linear instead of a non-linear model).
- Even in the presence of a deterministic model, with no other source of uncertainty except for the functional form of the model.

Definition

Model Discrepancy may be understood as the mismatch between the model being used and the data generating process being studied, even if all the other types of uncertainty did not exist.

Kennedy and O'Hagan (2001) Model Discrepancy Approach

Gaussian Processes

- Distribution on a Space of Functions. We write $f \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$
- Behaviour completely defined by $m(x) = E(f(x))$ and $k(x, x') = Cov(f(x), f(x'))$.

Non-parametric modelling.

UQ Approach(Regression Setting)

- The *Emulator/Surrogate Model* is a GP estimated from computer modelled data.
- *Bias Term* is another GP, given experimental data and the estimated emulator (modularization).

An example of a Gaussian Process

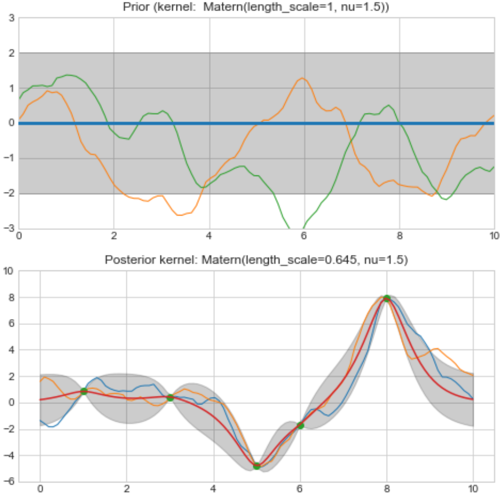

Figure: **Top Panel**: 2 realizations of GP, with zero mean in blue. Point-wise 95% prior credible interval in grey. **Bottom Panel**: The posterior mean function in red, plus 2 realizations of posterior GP. Point-wise 95% posterior credible intervals in grey.

Kennedy and O'Hagan (2001) Model Discrepancy Approach

Connection to Unobserved Heterogeneity(IV, Spatial Correlation, etc.)

- *Computer Model* alone is not able to fully represent the *True Model*.
- Hence, not using a *Bias Term* would allow for input vars' dynamics not in Emulator to be captured by measurement error.

$$\boxed{\text{Observation}} = \boxed{\text{Computer Model(X)}} + \underbrace{\boxed{\text{Correlated Measurement error}}}_{\boxed{\text{Bias Term (X)}} + \boxed{\text{Indep. Error Term}}}$$

Figure: Endogeneity when not accounting for bias term.

(Possible) Identification Issues

Confounding of True Model (via Bias term) and Measurement error

- Incorrect modelling of measurement error, may lead the *bias term* to capture more sources of uncertainty than *model discrepancy*.
- Probably more present in Economics, than natural sciences (observational data harder to model than experimental data).

(Possible) Identification Issues

Confounding of Computer Model and Discrepancy Term

- Different combinations of the emulator parameter inputs and discrepancy term parameters may give the same predicted experimental response. (It does not influence predictive performance)
- Possible solutions: Choice of more informative priors, Modularization (limits UQ interpretation), Multiple Output.

Confounding of non-zero Mean function and GP

- Usually the GP mean function is assumed to be 0. However, our economic model could be interpreted as the mean function.

(Possible)Identification Issues

$$\boxed{\text{Observation}} = \overbrace{\boxed{\text{Economic Model}} + \boxed{\text{Discrepancy Term}}}^{\boxed{\text{Non-Zero Mean GP}}} + \boxed{\text{Measurement Error}}$$

- Possible Solution: Use of Orthogonal Gaussian Processes (higher computational burden, and restrictive assumptions on shape of Mean function).

None of these last two types of identification issues seem to affect predictive performance.

Model Discrepancy Approach to SSM

A (usual) State-Space Model

$$x_t = g(x_{t-1}) + \eta_t$$
$$y_t = f(x_t) + \epsilon_t$$

- y_t is the vector of observed variables, x_t is the unobserved state, and η_t, ϵ_t iid error terms, with f, g being deterministic functions.

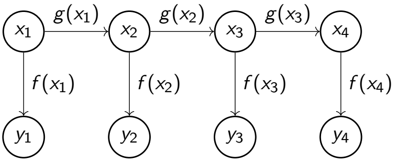

Figure: Dependence Structure of a SSM

Model Discrepancy Approach to SSM

$$
\begin{aligned}
x_t &= g(x_{t-1}) + \eta_t \\
y_t &= \underbrace{f(x_t) + b(x_t)}_{\tilde{b}(x_t)} + \epsilon_t
\end{aligned}
$$

- $\tilde{b} \sim \mathcal{GP}(f + m, k)$

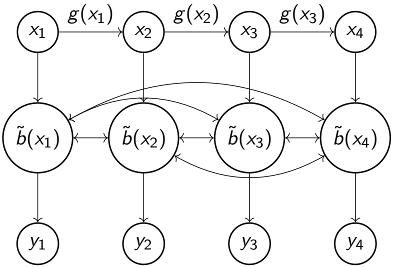

Figure: Dependence Structure of our GP SSM

Our Idea

Accounting for Model Discrepancy

$$
\begin{aligned}
x_t &= g(x_{t-1}) + \epsilon_t \\
y_t &= f(x_t) + \text{Bias}(x_t) + \text{Meas. Error}
\end{aligned}
$$

- Multivariate State-Space Model: y_t, x_t are multivariate
- Uncertainty Quantification: Frigola-Alcade et Al.(2013) studies a GP State-Space model, but in a regression setting only.
- To our knowledge, no one has used this technique in Macroeconomics, in any context.
- Possible to apply to non-linear $f(\cdot), g(\cdot)$.

Estimating $b(x_*)$ and y_*

A Monte Carlo Approximation

Let x_* be the state vector for which we want to predict $b(x_*)$.

$$p(\tilde{b}_* \mid x_*, y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N} N(\tilde{b}_* \mid \mu_{\tilde{b}_*}[i], \Sigma_{\tilde{b}_*}[i])$$

- $\mu_{\tilde{b}_*}[i]$ and $\Sigma_{\tilde{b}_*}[i]$ are obtained from the Multi-output GP regression formulas.
- Model discrepancy: $b_*[i] = \tilde{b}_*[i] - f(x_*)[i]$

Similarly for y_*, we have

$$p(y_* \mid x_*, y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N} N(y_* \mid \mu_{\tilde{b}_*}[i], \Sigma_{\tilde{b}_*}[i] + \Sigma_{\epsilon}[i])$$

How will we estimate the parameters and the states?

Learning the model

PGAS for Bayesian Learning of SSMs - Lindsten et Al(2014)

Particle Gibbs Ancestor Sampling (advantages)

- Particle Filtering $\rightarrow$ Preserves Non-Linear dynamics.
- Particle Gibbs has an easier implementation than Particle MH.
- Ancestor Sampling $\rightarrow$ Alleviates path degeneracy, allowing less particles to be used, improving computational performance

General Framework(Gibbs)

θ: collection of all parameters from the GP and the SSM (fully bayesian).

- Draw $\theta[i]$ from $p(\theta \mid y_{1:T}, x_{1:T}[i], \theta[i-1]) \rightarrow$ M-H step.
- Draw $x_{1:T}[i+1]$ from $p(x_{1:T} \mid y_{1:T}, \theta[i], x_{1:T}[i]) \rightarrow$ Particle Filtering.

Learning the model

PGAS for Bayesian Learning of SSMs - Lindsten et Al(2014)

Sampling θ - MH step

- To improve the acceptance ratio of the M-H step, we divided the θ into different blocks(economic, GP's hyperparameters,etc.)
- For each block j, draw θ_j^* from its proposal $q(\theta_j \mid \text{remaining blocks of } \theta, x_{1:T}[i], y_{1:T})$.
- Evaluate the Metropolis-Hastings step at $\theta^* = (\theta_{<j}[i], \theta_j^*, \theta_{>j}[i-1])$

Learning the model

PGAS for Bayesian Learning of SSMs - Lindsten et Al(2014)

Sampling $x_{1:T}$ - Particle Filtering

- Compute recursively the weights characterizing the distribution we want to explore. (Sequential Importance Sampling)
- Sample the particles history giving higher probability to particles trajectories with higher weights. (Particle Filtering)
- Set the reference trajectory as $x_{1:T}[i-1]$. (Gibbs)
- The ancestor sampling for the reference particle is done using updated weights after the resampling of the remaining particles. (Ancestor Sampling)

Hence the name Particle Gibbs Ancestor Sampling (PGAS).

A Comparison to Related Models

Additive Hybrid Model

Our Economic model:

$$x_t = Ax_{t-1} + \nu_t$$
$$y_t = C_* + Cx_t + u_t$$

Ireland(2004): No longer iid u_t, but allows $u_t = Du_{t-1} + \xi_t$, and normally distributed error terms.

Model Discrepancy Approach: $u_t = b(x_t) + \epsilon_t$, with b being a GP, and with normally distributed error terms.

Economic Model and Data

Economic Model

- Matrices A, C, C_* are functions of economic parameters $(\beta, \delta, \eta, \theta, \rho, A, \gamma)$.
- x_t: unobserved log-deviations of *capital*, and the *technology shock*.
- y_t:log-deviations of *economic output* (Y_t), *consumption* (C_t) and *number of hours worked* (H_t).

Data

When comparing our performance to Ireland(2004), we used the data supplied by Ireland himself, which he took from the FRED database (St. Louis Fed).

Economic Model and Data

Data

For the remaining comparisons, we used the following series from FRED:

- Y_t - Real GDP, billions of chained 2012 dollars, quarterly, seasonally adjusted annual rate (GDPC1).
- C_t - Real Personal Consumption expenditures, billions of chained 2012 dollars, quarterly, seasonally adjusted annual rate (PCECC96)
- H_t - Hours of wage and salary workers on non-farm payrolls: private sector, billions of hours, quarterly, seasonally adjusted annual rate (PRSCQ).

All series were converted to per capita terms by dividing by civilian, non-institutional population level (16 and above).

GP Hybrid Model

Covariance Function, Priors

Covariance Function

- We construct a Multi-Output Matèrn Kernel, dependent on a matrix M and parameter ℓ, and we allowed rough processes(non-differentiable).
- M: how different components of GP correlate each other.
- ℓ: range/length-scale parameter. The more that farther data are correlated, the higher ℓ is.

Priors

- The structural parameters' priors were chosen based on their allowed/generally expected values(Betas, Gammas, Inverse-Gammas).
- For M, we chose an Inv-Wishart prior.

GP Hybrid Model

Proposal Distribution

Proposal for MH Step

We block θ as $(\theta_{\text{Econ}}, \theta_{q,\sigma}, \ell, M)$:

- For θ_{Econ} and $\theta_{q,\sigma}$, we chose 2 Multivariate-T Distributions
- For ℓ, a Univariate-T Distribution.
- For M, we chose an Inverse-Wishart Distribution

The means of the proposals were created using a Random Walk step.

Empirical Results

A Summary

Estimation and 1-Lag Forecast for an expanding window upto 15 data points

- Traces show mixing could be improved, but chains are still usable.
- Economic parameters estimations were not able to be 'corrected' into having a closer interpretation to economic theory.
- Amount of data used does not seem to be very informative.
- *Bias term* components for *output* and *consumption* seem to be correlated, but not *hours worked*.
- The GP SSM corrects models' prediction for *output* and *consumption*, but not for *hours worked*.
- The computational time performance for the GP SSM estimation algorithm needs to be improved.

Empirical Results

Continued

Comparing Predictive Performance with Ireland's model

Using Ireland's *original* data:

- when doing a 1-lag forecast, with a rolling window of 10 data points, the GP got a lower RMSE.

Venues for future research

- Reduce the computational burden.
- Assess possible identification issues.
- Use more informative priors to compensate for possible insufficient amount of data used in estimation.

Hilbert Approximation of a Multidimensional GP

Using Spectral methods as in Svensson et Al(2016)

Multi-Dimensional Approximation of GP as in Svensson et Al(2016)

If f follows a GP, then $f(x_t) \approx W\phi(x_t)$

- ϕ vector with the eigenfunctions which characterize the Laplace operator on a compact set of the input space.
- $W \mid \Sigma \sim \text{Matrix-Normal}(0, \Sigma, V)$, where V is dependent on the spectral density of the GP's Cov function.
- Σ has Inv-Wishart distribution and characterises the rows cov of W. (conjugate prior).

Doing UQ with the approximation

Using Spectral methods as in Svensson et Al(2016)

$$
\begin{aligned}
x_t &= g(x_{t-1}) + \eta_t \\
y_t &= f(x_t) + b(x_t) + \epsilon_t
\end{aligned}
$$

- We do $b(x_t) \approx W\phi(x_t)$.
- PGAS adapted to this new approach will give us a posterior sample of W, which can be used to predict $b(x_*)$.
- The improvement in computational performance comes from sampling the states, where we no longer have to evaluate the likelihood when computing the weights.

The bias corrected predictions for y_* at x_* are drawn from $p(y_* \mid x_*, y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N} N(y_* \mid f[i](x_*) + W[i]\phi(x_*), \Sigma[i])$

Specifying the Model

Covariance Function, Priors and Proposal

- We use the Spectral Density corresponding to the Matern Kernel
- We use the same prior and proposals for the M-H step, with the difference that now we have no matrix M to estimate.

Degree and Input Size

- We used a 12 degree approximation for each component of the output.
- Using the Full GP state samples as a soft reference, we decided to choose $[-7, 7]$ as the compact interval for each dimension of the input.

Empirical Results

A Summary

- Traces show the mixing for certain parameters have decreased, particularly for those of economic model.
- The RMSE, as expected, is greater than those of the Full GP hybrid model, but still smaller than those reported by Ireland(2004).
- The RMSE results are highly dependent on the L_i and m_i.
- Computational performance time improvement of 20%. The more data, and the greater the output dim, the greater the improvement when compared to the full GP.

The Great Recession

Objectives

- Compare forecasting performance to Negro, Schorfheide(2013).
- Assess impact of tighter Cov functions and priors on predictive distribution and identifiability.

Results

- The 20% credible interval (40th and 60th percentile interval of the posterior predictive distribution) captures almost all of the variation.
- Tighter Cov functions and priors lead to tightening of the predictive distribution.

Conclusions

- Given that we are using observational data, it is very likely that the *bias term* captures also some measurement error, besides model discrepancy.
- Considering the uncertainty bounds from The Great Recession exercise, even in the case of tighter priors, it seems that the economic model is not very useful in predicting the economic dynamics.
- It may be a stepping stone into bridging the econometric time-series modelling perspective and the theoretical Macroeconomic DSGE models.

References

M. C. Kennedy, A. O'Hagan(2001)
Bayesian Calibration of Computer Models
Journal of Royal Statistical Society B Vol. 63, Issue 3, pp. 425–464.

Ireland, P. (2004)
A Method for Taking Models to the Data
Journal of Economic Dynamics and Control Vol. 28, Issue 26, pp. 1205-1226

Lindsten, F., Jordan, M.I., Schön, T.S (2014)
Particle Gibbs with Ancestor Sampling
Journal of Machine Learning Research Vol. 15, pp. 2145-2184

References(cont.)

Lindsten, F., Jordan, M.I., Schön, T.S (2014)
Particle Gibbs with Ancestor Sampling
Journal of Machine Learning Research Vol. 15, pp. 2145-2184

Frigola-Alcalde, R., Lindsten, F., Schön, T.S, Rasmussen, C. (2013)
Bayesian Inference and Learning in Gaussian Process State-Space Models with Particle MCMC
Advances in Neural Information Processing Systems Vol. 26, pp. 3156-3164

Svensson, A., Solin, A., Särkkä, S., Schön, T.(2016)
Computationally Efficient Bayesian Learning of Gaussian Process State Space Models
*ar*Xiv:1506.02267 [stat.CO]

Negro, M. Del and Schorfheide, F. (2013)
DSGE Model-Based Forecasting
Handbook of Economic Forecasting Ed. by G. Elliott and A. Timmermann. Vol. 2. Cambridge University Press. Chap. 2, pp. 57–140.