

Modelling Input Varying Correlations Between Multiple Responses

Andrew Gordon Wilson and Zoubin Ghahramani

University of Cambridge,
Cambridge, UK

agw38@cam.ac.uk zoubin@eng.cam.ac.uk

Abstract. We introduced a generalised Wishart process (GWP) for modelling input dependent covariance matrices $\Sigma(x)$, allowing one to model input varying correlations and uncertainties between multiple response variables. The GWP can naturally scale to thousands of response variables, as opposed to competing *multivariate volatility* models which are typically intractable for greater than 5 response variables. The GWP can also naturally capture a rich class of covariance dynamics – periodicity, Brownian motion, smoothness, ... – through a covariance kernel.

1 Introduction

Modelling covariances between random variables is fundamental in statistics. For convenience, covariances between multiple responses are usually assumed to be constant. However, accounting for how these covariances depend on inputs (e.g. time) can greatly improve statistical inferences. For example, to predict the expression level of a gene at a particular time, it helps to consider the expression levels of correlated genes, and how these correlations depend on time.

Modelling of dependent covariances between multiple responses is largely uncharted territory. The small number of existing models for dependent covariances are mostly found in the econometrics literature, and are referred to as *multivariate volatility models*. In econometrics, a good estimate of a time varying covariance matrix $\Sigma(t) = \text{cov}[\mathbf{r}(t)]$ for a vector of returns $\mathbf{r}(t)$ is useful for estimating the risk of a particular portfolio. Multivariate volatility models are also used to understand *contagion*: the transmission of a financial shock from one entity to another (Bae et al., 2003). However, it is generally useful – in econometrics, machine learning, or otherwise – to know input dependent uncertainty, and the dynamic correlations between multiple entities.

Despite their importance, conventional multivariate volatility models suffer from tractability issues and a lack of generality. MGARCH (Bollerslev et al., 1988; Silvennoinen and Teräsvirta, 2009), multivariate stochastic volatility (Harvey et al., 1994; Asai et al., 2006), and the original Wishart process (Bru, 1991; Gouriéroux et al., 2009), are typically highly parametrised, parameters are often difficult to interpret or estimate (given the constraint $\Sigma(t)$ must be positive definite), are typically intractable for more than 5 response variables, and are

restricted to Brownian motion or Markovian covariance dynamics (Silvennoinen and Teräsvirta, 2009; Gouriéroux, 1997; Gouriéroux et al., 2009).

Modelling of dependent covariances is beautifully suited to a Bayesian non-parametric approach. We introduced the Bayesian nonparametric *generalised Wishart process* (GWP) prior (Wilson and Ghahramani, 2010, 2011) over input dependent matrices $\Sigma(x)$, where $x \in \mathcal{X}$ is an arbitrary input variable. The generalised Wishart process volatility model has the following desirable properties:

1. The GWP is tractable for up to at least 1000×1000 covariance matrices.
2. The small number of free parameters give information about the underlying source of volatility, like whether there is periodicity (and if so what the period would be), and how far into the past one should look for good forecasts.
3. The input variable can be any arbitrary $x \in \mathcal{X}$ just as easily as it can represent time (useful for spatially varying dependencies, and for including covariates like interest rates in time series models).
4. The dynamics of $\Sigma(x)$ can easily be specified as periodic, smooth, Brownian motion, etc., through a kernel function.
5. Missing data are handled easily, and there is prior support for any (uncountably infinite) sequence of covariance matrices $\{\Sigma(x_1), \dots, \Sigma(x_n)\}$.

2 Construction

The Wishart distribution is a distribution over positive definite matrices. Given a $p \times \nu$ matrix A with entries $A_{ij} \sim \mathcal{N}(0, 1)$, and a lower triangular matrix of constants L , the product $LAA^\top L^\top$ has a Wishart distribution:

$$LAA^\top L^\top \sim \mathcal{W}_p(\nu, LL^\top). \quad (1)$$

To turn the Wishart distribution into a generalised Wishart process (in its simplest form), one replaces the Gaussian random variables with *Gaussian processes* (Rasmussen and Williams, 2006). We let the matrix A be a function of inputs x , by filling each entry with a Gaussian process: $A_{ij}(x) \sim \mathcal{GP}(0, k)$. We let

$$\Sigma(x) = LA(x)A(x)^\top L^\top. \quad (2)$$

At any given x , the matrix $A(x)$ is a matrix of Gaussian random variables, since a Gaussian process function evaluated at any input location is simply a Gaussian random variable. Therefore at any x , $\Sigma(x)$ has a Wishart marginal distribution. $\Sigma(x)$ is a collection of positive definite matrices, indexed by x , and dynamics controlled by the covariance kernel k . $\Sigma(x)$ has a generalised Wishart process prior, and we write $\Sigma(x) \sim \mathcal{GW}\mathcal{P}(\nu, L, k)$. The parameters are easily interpretable. L controls the prior expectation of $\Sigma(x)$ at any x : $\mathbb{E}[\Sigma(x)] = \nu LL^\top$. The greater ν the greater our confidence in this prior expectation. The covariance kernel controls how the entries of $\Sigma(x)$ vary with x : $\text{cov}(\Sigma_{ij}(x), \Sigma_{ij}(x')) \propto k(x, x')^2$. A single draw from a GWP prior over 2×2 covariance matrices is illustrated in Figure 1. Given vector valued observations $\mathbf{r}(x)$ (e.g. a vector of stock returns indexed by x), we can efficiently infer a posterior over the generalised Wishart process using Elliptical Slice Sampling (Murray et al., 2010), which is a recent MCMC technique designed to sample from posteriors with Gaussian priors.

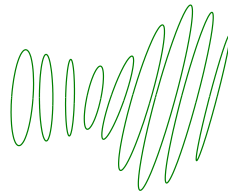


Fig. 1. A draw from a generalised Wishart process (GWP). Each ellipse is a 2×2 covariance matrix indexed by time, which increases from left to right. The rotation indicates the correlation between the two variables, and the axes scale with the diagonals of the matrix. Like a draw from a Gaussian process is a collection of function values indexed by time, a draw from a GWP is a collection of matrices indexed by time.

3 Results

We generated a 2×2 time varying covariance matrix $\Sigma_p(t)$ with periodic components, simulating data at 291 time steps from a Gaussian:

$$\mathbf{y}(t) \sim \mathcal{N}(\mathbf{0}, \Sigma_p(t)). \quad (3)$$

Periodicity is especially common to financial and climate data, where daily trends repeat themselves. For example, the intraday volatility on equity indices and currency exchanges has a periodic covariance structure. Andersen and Bollerslev (1997) discuss the lack of – and critical need for – models that account for this periodicity. With a GWP, we can simply use a periodic kernel function, whereas in previous Wishart process volatility models (Bru, 1991; Gouriéroux et al., 2009), we are stuck with a Markovian covariance structure. Figure 2 shows the results. We also performed step ahead forecasts of $\Sigma(t)$ on financial data with promising results, elucidated in Wilson and Ghahramani (2010, 2011). The recent *Gaussian process regression network* (GPRN) (Wilson et al., 2011, 2012) uses a GWP noise model, and extends the multi-task Gaussian process framework to handle *input dependent* signal and noise correlations between multiple responses. The GPRN has strong predictive performance and scalability on many real datasets, including a gene expression dataset with 1000 response variables.

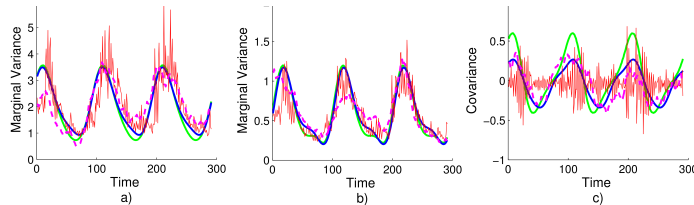


Fig. 2. Reconstructing the historical $\Sigma_p(t)$ for the periodic data set. We show the truth (green), and GWP (blue), WP (dashed magenta), and MGARCH (thin red) predictions. a) and b) are the diagonal elements of $\Sigma_p(t)$, c) is the covariance.

References

- Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3):115–158.
- Asai, M., McAleer, M., and Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, 25(2):145–175.
- Bae, K., Karolyi, G., and Stulz, R. (2003). A new approach to measuring financial contagion. *Review of Financial Studies*, 16(3):717.
- Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 96(1):116–131.
- Bru, M. (1991). Wishart processes. *Journal of Theoretical Probability*, 4(4):725–751.
- Gelfand, A., Schmidt, A., Banerjee, S., and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312.
- Gouriéroux, C. (1997). *ARCH models and financial applications*. Springer Verlag.
- Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.
- Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.
- Murray, I., Adams, R. P., and MacKay, D. J. (2010). Elliptical Slice Sampling. *JMLR: W&CP*, 9:541–548.
- Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for Machine Learning*. The MIT Press.
- Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. *Handbook of Financial Time Series*, pages 201–229.
- Wilson, A., Knowles, D., and Ghahramani, Z. (2011). Gaussian process regression networks. *Arxiv preprint arXiv:1110.4411*.
- Wilson, A., Knowles, D., and Ghahramani, Z. (2012). Gaussian process regression networks. *International Conference on Machine Learning*.
- Wilson, A. G. and Ghahramani, Z. (2010). Generalised Wishart Processes. *Arxiv preprint arXiv:1101.0240*.
- Wilson, A. G. and Ghahramani, Z. (2011). Generalised Wishart Processes. In *UAI*.