Gaussian Process: A Smooth and Flexible Approach to Estimating Index Complementarities in Organizational Economics

Kieron J. Meagher*
Australian National University

Rodney W. Strachan
University of Queensland

March 14, 2014

Abstract

A common question in organizational economics is how does a bundle or index of managerial practices or characteristics impact on firm or employee level outcomes. The presence of complementarities is of particular interest but we argue should not be restricted to particular functional forms like the multiplicative 'interaction' of slope coefficients. As an alternative we propose the use of Gaussian processes to estimate a smooth non-linear function of the management practices index/bundle.

1 Introduction

At the heart of organizational economics is an interest in how the management of an organization impacts on the organization's performance and/or on the employees of the organization. At an empirical level 'management' is a compound concept made up of many process, practices and structures. Much theorizing in both Economics and Management predicts that the impact of

*Corresponding author: Research School of Economics, Australian National University, Canberra, ACT 0200, Australia. *e-mail*: kieron.meagher@anu.edu.au

improved management practices should not be constant across the distribution of firms but rather that the impacts of small improvements should be modest for most firms and that the big payoff comes from getting the whole bundle of practices right (see [4] for survey and overview of this extensive literature).

This strong complementarity from getting the right bundle of management practices has also been identified in the very narrowly focused industry studies of the insider econometrics approach (see [6]). By focusing on very narrow industries all plants will be using approximately the same technology and the same set of specific management practices are readily identifiable as being utilized or not through a process of plant visits and observation by researchers. As an empirical strategy in this literature, researchers have identified the complementarity from using a bundle of practices with a dummy variable.

The insider econometrics approach is very time intensive and is not scalable to national level studies of a broad range of industries. In their seminal paper [3] report the results of a telephone survey based approach to measuring management practices using 18 dimensions, with 5 point scaling on each dimension. The 18 dimensions are collapsed to a single management index through arithmetic averaging and standardization (z-scoring). They establish that when measured in this way, management practices have a large and statistically significant impact on output of manufacturing firms in the USA, UK, France and Germany. The result is established primarily through the OLS estimation of an augmented Cobb-Douglas production function.[1,2]

The problem we consider here is how to investigate empirically the presence of complementarities from having the right bundle of practices when using a management index. Assigning a dummy variable to firms with a good bundle of practices is rather ad hoc and it would be much more desirable to let the data show where any such non-linearity/convexity in the index is present. Simple interactions of the slope coefficients is infeasible

[1] The Cobb-Douglas framework is standard in the literature because it has proved to be well behaved in practice and approximates more flexible functional forms (see [10]). The results in [3] are robust to considering return on assets and Tobin's Q as dependent variables and to alternative estimation approaches like [7]

[2] The Bloom and Van Reenen approach has had significant policy impact and has been extended to manufacturing firms in over twenty countries across Europe, Asia and the Americas (see [2]) and to healthcare, retail and schools. Details can be found at the World Management Survey website: http://worldmanagementsurvey.org.

here due to the high dimensionality: there are thousands of such interactions (153 pairs, 816 triples, and so on). Furthermore theory and the empirical findings from the insider econometrics studies both suggest that the complementarity need not be a constant proportional effect. In this paper we outline the proposal from Meagher and Strachan (2013) to use a Gaussian process approach ([9]), in which each firm has its own management function, to investigate complementarities in a very flexible and non-linear way. The discussion is framed within the context of the total factor productivity since this is the focus of much of the literature, including [3]. However the Gaussian process approach could be used in a variety of organizational economics settings, for example to examine the impact of leadership on the employee experience (see for example [1]).

2 The Model

2.1 Example: Cobb-Douglas and TFP

One of the most common ways to think about a firm's performance in economics is in terms of total factor productivity (see [10] for an excellent survey). The Gaussian Process is not constrained to this setting but as a concrete example to frame the issues we will briefly discuss the key issues in the context of estimating the impact of management on total factor productivity.

Total factor productivity (TFP), is most commonly calculated from the constant and residuals of a log-log regression of sales, S, on inputs.[3] The log-log regression estimates a Cobb-Douglas production function

$$S_i = A_i L_i^{\beta_1} K_i^{\beta_2} G_i^{\beta_3}$$

as

$$\ln S_i = \ln A + \beta_1 \ln L + \beta_2 \ln K_i + \beta_3 \ln G_i + \varepsilon_i \tag{1}$$

where $A_i = Ae^{\varepsilon_i}$ is total factor productivity, L is labor, K is capital and G is intermediate inputs. Thus the constant plus the residual from this regression, which we will denote by $a_i = \ln A + \varepsilon_i$, represents log TFP as the (log of) sales

[3]Theoretically output is the prefered dependent variable but is frequently not available from firm level financial records. Price deflators, if available, can be applied to sales but have their own associated problems.

unexplained by these factors but which may be explained by other factors such as management practices.

Recognizing the potential importance of management practices, [3] pioneer the extension of this analysis in a very natural way, by augmenting the standard log-log specification with the management index, m_i, specifically

$$\ln S_i = \ln A + \beta_1 \ln L + \beta_2 \ln K_i + \beta_3 \ln G_i + \beta_4 m_i + \varepsilon_i.$$

The management score is taken by simply averaging a firm's score across the 18 categories and is then normalized as a z-score by subtracting the sample mean and dividing by the sample standard deviation. That is, the z - statistic computed as an affine function of the management index;

$$z_i = \frac{m_i - \overline{m}}{s_m}$$

where $\overline{m}$ is the sample mean of m_i and s_m is the estimated sample standard deviation.

This linear estimation setting of the log-log model is simple but it precludes a serious investigation of complementarities. The standard solution in applied econometrics would be to plot the raw data to 'eyeball' any nonlinearity and allow some further, specific non-linear terms in m_i in the estimation. In the following section we discuss an alternative approach from Meagher and Strachan (2013) which is something of a middle road, allowing more flexibility than the linear model while providing more structure and inferential opportunities than the graphical analysis. Specifically the approach will allow the estimation of $A_i(m_i)$, a smooth function of management for each firm.

2.2 A Gaussian Process Approach

Meagher and Strachan (2013) propose a likely form for the relationship between the response variable of interest (output) and the management index in which they impose no structure beyond smoothness. The response variable of interest, such as log sales, we denote by y_i. As we are interested in modelling the response of y_i to the management index score, z_i, without making any strong assumptions on the form of this response, we begin by modelling the response in a very flexible but smooth form using a Gaussian process. Using a Gaussian process is sometimes called a nonparametric estimate as no functional form is assumed for the mean of y_i.

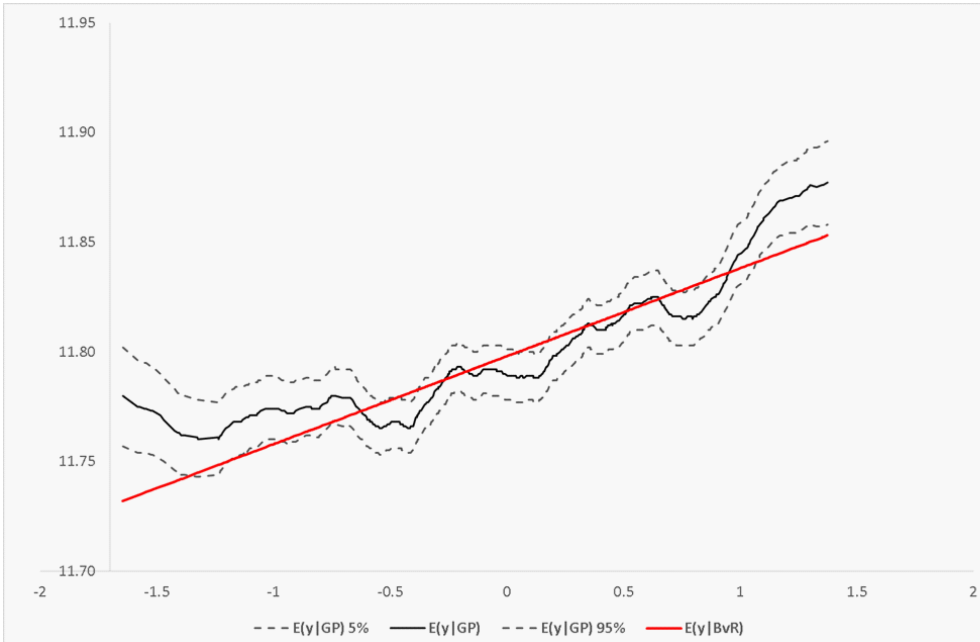

Figure 1: Plot of the Gaussian Processes (GP) and Linear OLS as in Bloom and Van Reenen (2007), BvR (2007). For the GP, the plot shows $E(y_i|GP)$ and the lower and upper bounds of the 95% credible interval denoted as $E(y_i|GP) - lower$ and $E(y_i|GP) - upper$ respectively. On the $x - axis$ is the z - statistic computed from the management index. Reproduced from Meagher and Strachen (2013), Figure 2.

The model estimated by a Gaussian process is

$$\begin{aligned} y_i &= \mu_i + x_i\beta + \varepsilon_i \\ \varepsilon_i &\sim N\left(0, \sigma^2\right). \end{aligned}$$

The coefficient $\mu_i = \mu(z_i)$ varies smoothly with z_i but is not otherwise given a specific functional form. The smoothness is achieved via the correlation structure among the μ_i. To describe the process more explicitly, collect the N different μ_i into an $N \times 1$ vector $\mu = (\mu_1, \mu_2, \ldots, \mu_N)'$ and give μ the prior distribution $\mu \sim N(0, R)$ where R is an $N \times N$ matrix of correlations. The strength of the correlation between any two points i and j is determined by the distance, $d_{ij} = \|z_i - z_j\|$, between the two points z_i and z_j. There are many ways to specify the correlation structure in R, but for our purposes we take a squared exponential form such that the correlations are given by

$$\rho_{ij} = \exp\left\{-\frac{\alpha}{2}h\left(d_{ij}\right)\right\}$$

where $h_i = h\left(\|z_i - z_j\|\right) > 0$ is a monotonically increasing function of the distance d_{ij}. By construction the resulting function will be continuous.

Meagher and Strachan (2013) presents the posterior average of the prediction

$$E\left(y_i|z_i, x_i = \overline{x}, \mu, \beta, \sigma^2\right) = \mu_i + \overline{x}\beta$$

obtained using the Gaussian process on the World Management Survey data from Bloom and Van Reenen (2007). In Bayesian analysis, the parameters are treated as random and unknown, so functions of the parameters, such as $E\left(y_i|z_i, x_i = \overline{x}, \mu, \beta, \sigma^2\right)$, are also random and have their own distributions (see Meagher and Strachan, 2013 for the estimated 95% credible interval). The results from Meagher and Strachan (2013) are reproduced here in Figure 1. Their results show that the Gaussian process approach picks up an important non-linearity in the relationship between management and TFP which is consistent with the literature on complementarities.

3 Conclusion

Motivated by theory, we show how to estimate the possibility of a non-linear relationship between management and output. From a very flexible Gaussian process model, each firm is allowed to have its own output related function of management (with these functions varying smoothly between close firms). Meagher and Strachan (2013) apply this process to estimating the impact of management on TFP using the World Management Survey data from [3] and find strong evidence of non-linearities/complementarities for the USA, UK, France and Germany.

References

[1] Boedker C., Vidgen R., Meagher K., Cogin J., Mouritsen J., and Runnalls J. M. (2011) Leadership, Culture and Management Practices of High Performing Workplaces in Australia: High Performing Workplace Index. The Society for Knowledge Economics, October, 2011. ISBN: 978-0-9775436-3-2.

[2] Bloom, N., Sadun, R., and Van Reenen, J. (2012). The organization of firms across countries. Quarterly Journal of Economics, 127(4):1663–1705.

[3] Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. The Quarterly Journal of Economics, 122(4):1351–1408.

[4] Brynjolfsson, E. and Milgrom, P. (2013). Complementarity in organizations. In Gibbons, R. and Roberts, J., editors, The Handbook of Organizational Economics, pages 11–55. Princeton University Press.

[5] Gibbons, R. and Henderson, R. (2013). What do managers do? In Gibbons, R. and Roberts, J., editors, The Handbook of Organizational Economics, pages 680–731. Princeton University Press.

[6] Ichniowski C. and K. Shaw (2013). Insider Econometrics. In Gibbons, R. and Roberts, J., editors, The Handbook of Organizational Economics, pages 263–314. Princeton University Press.

[7] Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. Econometrica, 64(6):pp. 1263–1297.

[8] Meagher, K and R. Strachan (2013). Evidence on the Non-Linear Impact of Management, presented NBER Organizational Economics Working Group, Boston, December 2013.

[9] Rasmussen, C.E.; Williams, C.K.I (2006). Gaussian Processes for Machine Learning. The MIT Press.

[10] Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49(2):326–365.