Time-varying sparsity in dynamic regression models

Professor Jim Griffin
(joint work with Maria Kalli, Canterbury Christ Church University)

University of Kent

University of Kent

Regression models

Often we are interested in making forecasts or predictions (*e.g.* inflation).

Regression models

Often we are interested in making forecasts or predictions (*e.g.* inflation).

One generic method for prediction is regression where we assume that there are observed predictors which can be used to help. The model for predicting j periods ahead is

$$y_{t+j} = \alpha + \sum_{k=1}^{p} x_{t,k}\beta_k + \epsilon_t, \qquad t = 1, 2, \ldots, T$$

where

- y_t is the observation of the response variable at time t.
- $x_{t,k}$ is the value of the k-th predictor at time t.
- β_k is the coefficient for the k-th predictor.

University of Kent

Example: equity premium

We used the data set of Goyal and Welch (2008)

University of Kent

Example: equity premium

We used the data set of Goyal and Welch (2008)

- The response variable is the value weighted monthly return of the S & P 500 obtained from the CRSP database.

Example: equity premium

We used the data set of Goyal and Welch (2008)

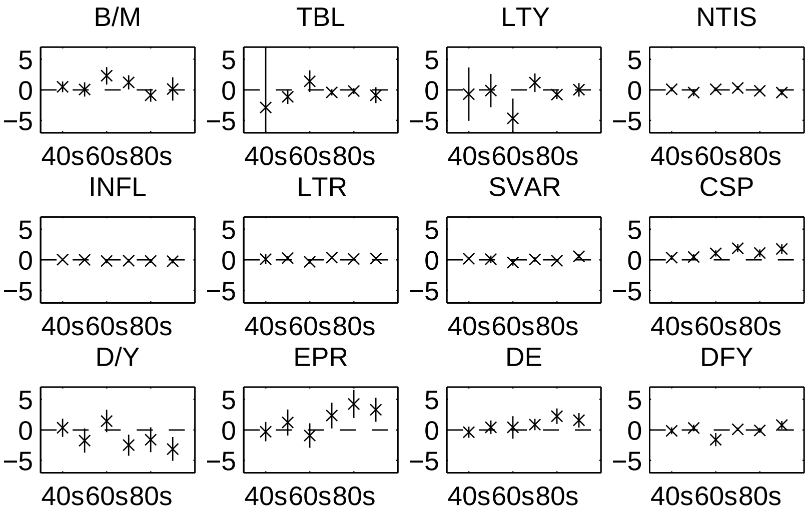
- The response variable is the value weighted monthly return of the S & P 500 obtained from the CRSP database.
- The sample period is May 1937 to December 2002.

Example: equity premium

We used the data set of Goyal and Welch (2008)

- The response variable is the value weighted monthly return of the S & P 500 obtained from the CRSP database.
- The sample period is May 1937 to December 2002.
- The set of twelve predictors includes variables relating to dividends, earnings, interest rates, bond yields and inflation.

Equity premium: regressions over decades



Comments

- The previous results show that the coefficient estimates are changing over time which suggests that the relationship between the predictors and the response is changing over time.

Comments

- The previous results show that the coefficient estimates are changing over time which suggests that the relationship between the predictors and the response is changing over time.
- The results also suggest that some predictors may not be important for predicting the response at all times or at some times.

Comments

- The previous results show that the coefficient estimates are changing over time which suggests that the relationship between the predictors and the response is changing over time.
- The results also suggest that some predictors may not be important for predicting the response at all times or at some times.
- This is an explanation of why "static" regression models, where predictor effects are assumed constant over time often produce poor out-of-sample forecasts or predictions when fitted to different time periods (see Fisher and Statman (2006), Paye and Timmermann (2006) and Dangl and Halling (2012)).

Dynamic regression model

One solution is the dynamic regression model.

Dynamic regression model

One solution is the dynamic regression model. The model assumes that

$$y_{t+j} = \alpha_t + \sum_{k=1}^{p} x_{t,k}\beta_{t,k} + \epsilon_t, \qquad t = 1, 2, \ldots, T$$

where

- y_t is the observation of the response variable at time t.
- $x_{t,k}$ is the value of the k-th predictor at time t.
- $\beta_{t,k}$ is the coefficient for the k-th predictor at time t.
- ϵ_t is the error at time t.

University of Kent

Estimation

The model cannot be directly estimated since there are $(p+1)T$ coefficients and only T observations.

University of Kent

Estimation

The model cannot be directly estimated since there are $(p+1)T$ coefficients and only T observations.

One standard solution assumes that $\beta_{1,1}, \ldots, \beta_{1,p}, \ldots, \beta_{T,1}, \ldots, \beta_{T,p}$ follows a stochastic process such as a random walk

$$\beta_{t,k} = \beta_{(t-1),k} + \nu_{t,k}$$

or vector autoregressive process

$$\beta_t = \Lambda \beta_{t-1} + \nu_t$$

where $\nu_{t,k}$ and ν_t are random disturbances.

University of Kent

Dynamic regression models with many regressors

Our focus is on how to construct a stochastic process for $\beta_{t,k}$ which corresponds to the findings from the results

Dynamic regression models with many regressors

Our focus is on how to construct a stochastic process for $\beta_{t,k}$ which corresponds to the findings from the results

1. Many values of $\beta_{t,k}$ will be close to zero.

University of Kent

Dynamic regression models with many regressors

Our focus is on how to construct a stochastic process for $\beta_{t,k}$ which corresponds to the findings from the results

1. Many values of $\beta_{t,k}$ will be close to zero.
2. $\beta_{t,k}$ will be close to zero at all times or at some times for some predictors.

Dynamic regression models with many regressors

Our focus is on how to construct a stochastic process for $\beta_{t,k}$ which corresponds to the findings from the results

1. Many values of $\beta_{t,k}$ will be close to zero.
2. $\beta_{t,k}$ will be close to zero at all times or at some times for some predictors.
3. Some coefficients will have values of $\beta_{t,k}$ which are away from zero for all or most times.

University of Kent

Bayesian regularization

Returning to the static regression model, the coefficients β need to be given a prior distribution.

Bayesian regularization

Returning to the static regression model, the coefficients β need to be given a prior distribution.

If there are many predictors, it is natural to assume that some (perhaps, many) coefficients will have values close to zero.

Bayesian regularization

Returning to the static regression model, the coefficients β need to be given a prior distribution.

If there are many predictors, it is natural to assume that some (perhaps, many) coefficients will have values close to zero.

The proportion of such predictors is often called the sparsity of the regression problem.

Bayesian regularization

Returning to the static regression model, the coefficients β need to be given a prior distribution.

If there are many predictors, it is natural to assume that some (perhaps, many) coefficients will have values close to zero.

The proportion of such predictors is often called the sparsity of the regression problem.

A prior distribution which can express different levels of sparsity is the normal-gamma, which has been suggested as a prior distribution in regression problems by Caron and Doucet (2008) and Griffin and Brown (2010).

Normal-gamma distribution

The normal-gamma prior can be written as

$$\beta_k|\psi_k \sim \mathrm{N}(0, \psi_k), \qquad \psi_k \sim \mathrm{Ga}(\lambda, \lambda/\mu)$$

where ψ_k will be called the relevance parameter for the k-th predictor.

Normal-gamma distribution

The normal-gamma prior can be written as

$$\beta_k|\psi_k \sim \mathrm{N}(0, \psi_k), \qquad \psi_k \sim \mathrm{Ga}(\lambda, \lambda/\mu)$$

where ψ_k will be called the relevance parameter for the k-th predictor.

The distribution of ψ_k expresses our prior belief about the relevance of the k-th predictor.

Normal-gamma distribution

The normal-gamma prior can be written as

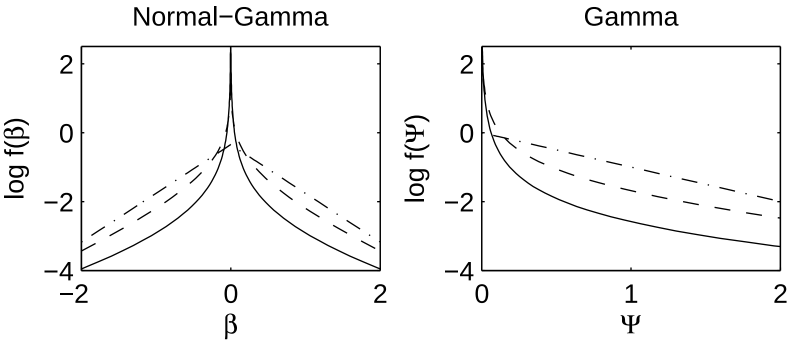
$$\beta_k|\psi_k \sim \text{N}(0, \psi_k), \qquad \psi_k \sim \text{Ga}(\lambda, \lambda/\mu)$$

where ψ_k will be called the relevance parameter for the k-th predictor.

The distribution of ψ_k expresses our prior belief about the relevance of the k-th predictor.

The choice of λ and μ determines how much mass is placed close to or away from zero.

Normal-gamma distribution



- $\lambda = 0.1$ (solid line)
- $\lambda = 0.333$ (dot-dashed line)
- $\lambda = 1$ (dashed line)

University of Kent

Normal-gamma autoregressive process

For dynamic regression models, we will construct a stochastic process for the coefficients whose marginal distribution is normal-gamma.

University of Kent

Normal-gamma autoregressive process

For dynamic regression models, we will construct a stochastic process for the coefficients whose marginal distribution is normal-gamma.

We write $\beta_{t,k} = \sqrt{\psi_{t,k}}\phi_{t,k}$ where

- $\phi_{1,k}, \ldots, \phi_{T,k}$ follows an AR(1) process with a standard normal stationary distribution and AR parameter φ_k.
- $\psi_{1,k}, \ldots, \psi_{T,k}$ follows an AR(1) process with a gamma marginal distribution with parameters λ_k and μ_k and AR parameter ρ_k (Pitt and Walker, 2005).

Normal-gamma autoregressive process

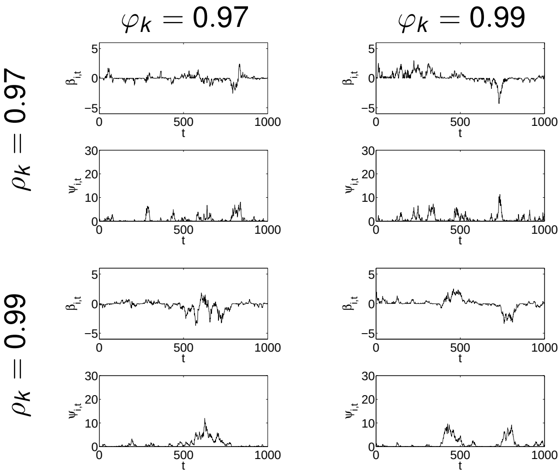
For dynamic regression models, we will construct a stochastic process for the coefficients whose marginal distribution is normal-gamma.

We write $\beta_{t,k} = \sqrt{\psi_{t,k}}\phi_{t,k}$ where

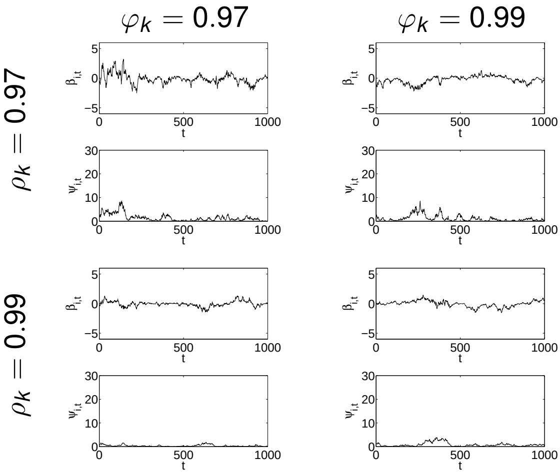
- $\phi_{1,k}, \ldots, \phi_{T,k}$ follows an AR(1) process with a standard normal stationary distribution and AR parameter φ_k.
- $\psi_{1,k}, \ldots, \psi_{T,k}$ follows an AR(1) process with a gamma marginal distribution with parameters λ_k and μ_k and AR parameter ρ_k (Pitt and Walker, 2005).

The processes $\phi_{1,k}, \ldots, \phi_{T,k}$ is independent of $\psi_{1,k}, \ldots, \psi_{T,k}$.

NGAR process ($\lambda_k = 0.2$)



NGAR process ($\lambda_k = 1$)



University of Kent

Comments

- λ_k controls the proportion of time that the process is close to zero.

Comments

- λ_k controls the proportion of time that the process is close to zero.
- μ_k controls the scale of the process.

Comments

- λ_k controls the proportion of time that the process is close to zero.
- μ_k controls the scale of the process.
- The parameters φ_k and ρ_k control the length of each period away from zero. Therefore, we choose priors for these parameters with most of their mass close to 1.

Further prior modelling

The NGAR process defines a process which spends some time close to zero and some time away from zero.

Further prior modelling

The NGAR process defines a process which spends some time close to zero and some time away from zero. However, “close” and “away” are relative to the parameter μ_k which controls scale.

Further prior modelling

The NGAR process defines a process which spends some time close to zero and some time away from zero. However, “close” and “away” are relative to the parameter μ_k which controls scale.

We define a second level of prior on μ_k so that

$$\mu_k \sim (\lambda^\star, \lambda^\star/\mu^\star).$$

Smaller values of $\lambda^\star$ suggests that more coefficients are close to zero at all times.

Further prior modelling

The NGAR process defines a process which spends some time close to zero and some time away from zero. However, "close" and "away" are relative to the parameter μ_k which controls scale.

We define a second level of prior on μ_k so that

$$\mu_k \sim (\lambda^\star, \lambda^\star/\mu^\star).$$

Smaller values of $\lambda^\star$ suggests that more coefficients are close to zero at all times.

The intercept α_t is assumed to follow a random walk.

University of Kent

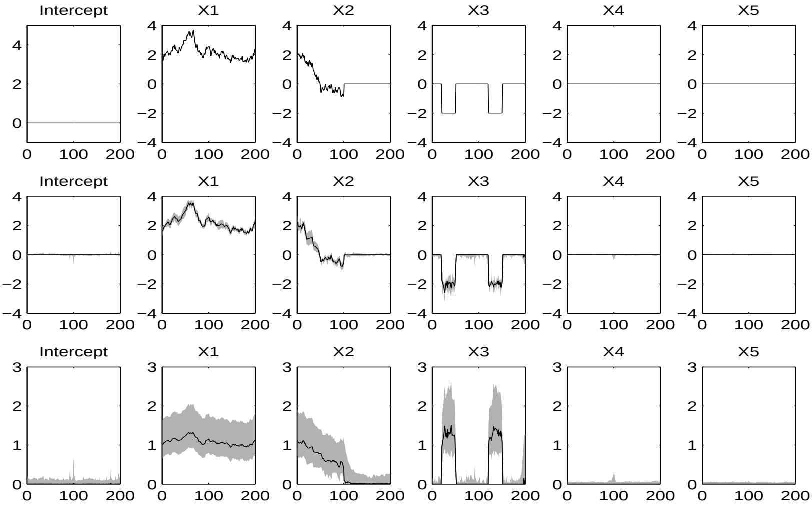
Stochastic volatility

The dynamic regression model is

$$y_{t+j} = \alpha_t + \sum_{k=1}^{p} x_{t,k}\beta_{t,k} + \epsilon_t, \qquad t = 1, 2, \ldots, T$$

which is completed by assuming that $\epsilon_t \sim \mathrm{N}(0, \sigma_t^2)$ where σ_t^2 is a given an AR(1) process with a gamma marginal distribution.

Simulated example



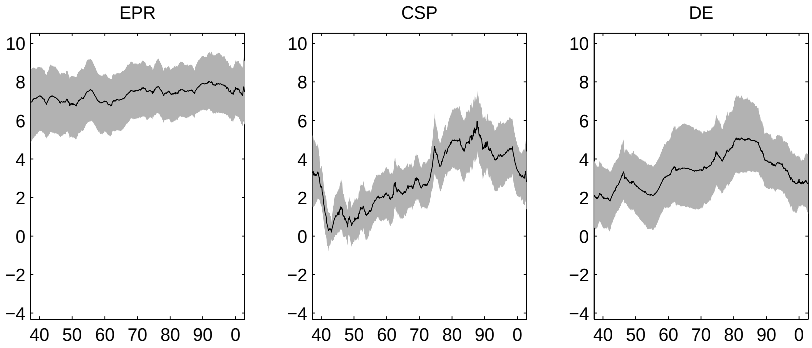
University of Kent

Equity premium

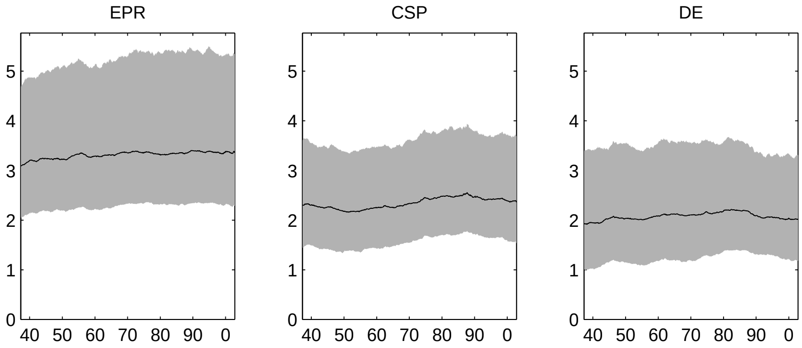
We used the data set of Goyal and Welch (2008)

- The response variable is the value weighted monthly return of the S & P 500 obtained from the CRSP database.
- The sample period is from May 1937 to December 2002.
- The set of twelve predictors includes variables relating to dividends, earnings, interest rates, bond yields and inflation.

Equity premium: coefficients



Equity premium: relevance



University of Kent

Inflation

- We constructed a data set using data series obtained from: FRED, the consumer survey database of the University of Michigan, the Federal Reserve Bank of Philadelphia, and the Institute of Supply Management.

Inflation

- We constructed a data set using data series obtained from: FRED, the consumer survey database of the University of Michigan, the Federal Reserve Bank of Philadelphia, and the Institute of Supply Management.
- We predict the gross domestic product (GDP) deflator.

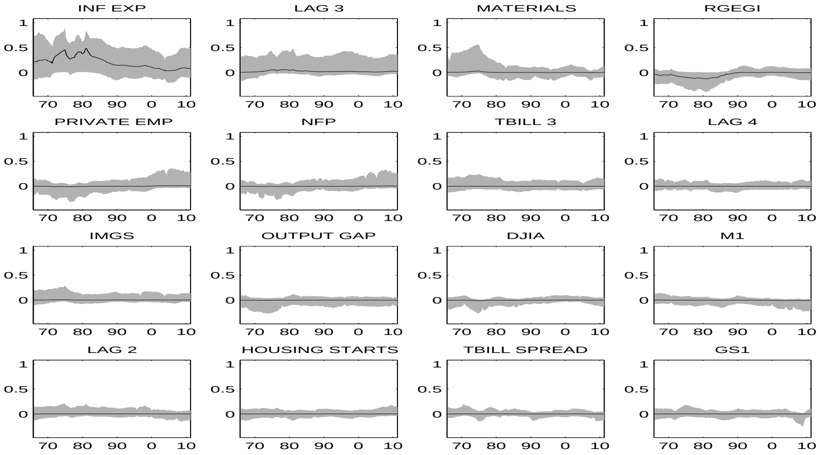
Inflation

- We constructed a data set using data series obtained from: FRED, the consumer survey database of the University of Michigan, the Federal Reserve Bank of Philadelphia, and the Institute of Supply Management.
- We predict the gross domestic product (GDP) deflator.
- The sample period is from Q2 of 1965 to Q1 of 2011.

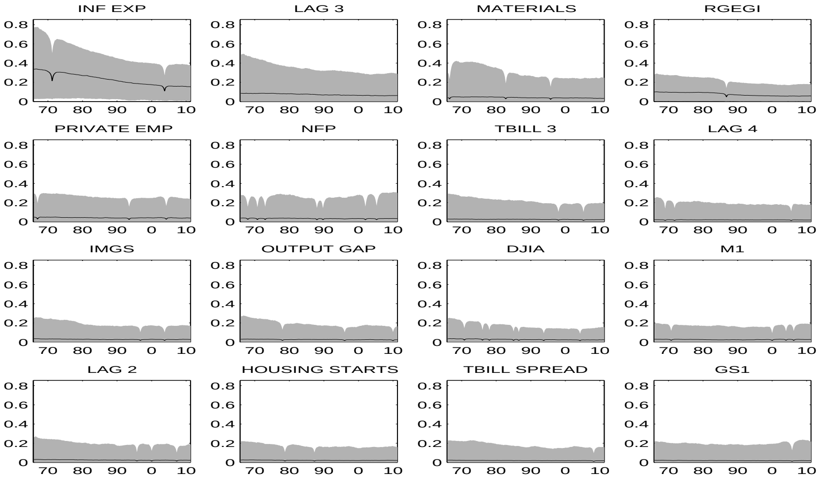
Inflation

- We constructed a data set using data series obtained from: FRED, the consumer survey database of the University of Michigan, the Federal Reserve Bank of Philadelphia, and the Institute of Supply Management.
- We predict the gross domestic product (GDP) deflator.
- The sample period is from Q2 of 1965 to Q1 of 2011.
- The data set includes 31 predictors, from activity and term structure variables to survey forecasts and previous lags.

GDP deflator: coefficients



GDP deflator: relevance



Out-of-sample prediction

We compare the DR model with the NGAR process prior to

- Time Varying Dimension (TVD) models (Chan *et al*, 2012),
- Dynamic Model Average (DMA) approach (Koop and Korobilis, 2011),
- Hierarchical shrinkage (HierShrink) (Belmonte *et al*, 2011).
- Rolling window Bayesian Model Averaging (BMA) using a *g*-prior for prediction.
- Random walk model of Atkeson and Ohanian (2001).

$$\text{RMSE} = \sqrt{\frac{1}{T-s}\sum_{t=s+1}^{T}\left(y_t - \mathrm{E}\left[y_t \mid y_1,\dots,y_{t-1},x_1,\dots x_t\right]\right)^2}$$

Out-of-sample prediction: RMSE

	Equity Premium	PCE Inflation	GDP Inflation
RW	1.100	0.635	**0.373**
NGAR	**0.977**	**0.611**	0.410
DMA	1.01	0.660	0.422
TVD1	2.193	2.688	2.688
TVD2	0.986	0.623	0.481
TVD3	0.992	0.628	0.500
HierShrink	1.547	1.131	2.556
gprior^1	2.822	0.796	0.660
gprior^2	1.648	0.712	0.588
gprior^3	1.282	0.681	0.516

The window lengths for the three g-priors were 100 (gprior^1), 200 (gprior^2) and 300 (gprior^3) for the equity premium data and 50 (gprior^1), 70 (gprior^2) and 90 (gprior^3) for the inflation data.

University of Kent

- The DR model with NGAR process prior is the best performing approach for two data sets (equity premium and PCE inflation) and the second best performing for the GDP inflation data (with only the random walk giving better predictions).
- In general, the approaches which allow the complexity of the regression model to change over time (NGAR, TVD and DMA) outperform the other approaches (HierShrink and rolling window g-prior). This illustrates the importance of allowing time-variation in the relevance of regression coefficients.

Summary

In this talk, I've presented a method which allows the effects of predictors in a regression model to vary over time and which allows their value to be very close to zero for some proportion of time.

University of Kent

Summary

In this talk, I've presented a method which allows the effects of predictors in a regression model to vary over time and which allows their value to be very close to zero for some proportion of time.

Bayesian inference about this model allows this proportion to adapt to the data.

Summary

In this talk, I've presented a method which allows the effects of predictors in a regression model to vary over time and which allows their value to be very close to zero for some proportion of time.

Bayesian inference about this model allows this proportion to adapt to the data.

The method also allows some predictor to have their coefficients close to zero at all times (effectively removing the predictor).

References

- Caron, F. and Doucet, A. (2008). Sparse Bayesian nonparametric regression, *Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland*.
- Dangl, T. and Halling, M. (2012). Predictive regressions with time varying coefficients, *Journal of Financial Economics*, **106**, 157–181.
- Fisher, K. and Statman, M. (2006). Market timing at home and abroad, *Journal of Investing*, **15**, 19–27.
- Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression models, *Bayesian Analysis*, **5**, 171–188.

- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21**, 1455–1508.
- Paye, B. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, **13**, 274–315.
- Pitt, M. K. and Walker, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications, *Journal of the American Statistical Association*, **100**, 554–564.

University of Kent