# MATH 426: Final Project

Christian Conroy and Francesca Mandel

## Introduction

Data from the World Bank World Development Indicators and the IMF International Financial Statistics databases has been collected into a panel of financial and macroeconomic indicators from 1980-2015. Of particular interest is how market capitalization, stock price volatility, and current account balance change over time in developed and emerging market countries. While market capitalization, stock price volatility, and current account balances are ultimately factored into whether a country is categorized as an emerging or developed market, we want to understand if developed and emerging economies are experiencing different levels of growth or decline in these longitudinal factors across time. The remainder of this analysis report is organized as follows: section 2 describes the data collected and the analysis conducted, section 3 reports the results of the planned analysis, and section 4 discusses the results in context, comparing between emerging and developed markets and characterizing changes over time.

## Methods

Annual measurements of macroeconomic indicators were taken on 46 countries from 1980 to 2015. *Market capitalization* is measured in US dollars, *stock price volatility* is measured on a scale based on the variance of returns for a given stock price index across a 360-day average, and *current account balance* is measured in US dollars.[1] Countries were grouped by market classification (23 emerging and 23 developed), which remains constant over the time period of interest. The data is balanced; however, there is missingness, particularly in earlier years. To account for the missing observations, we conducted covariate-based imputation to complete our data.

Initial exploratory analysis looked at mean profile plots comparing emerging and developed markets for each macroeconomic indicator to understand the general trends of the data. Then, to best understand how the three variables of interest change with time as well as with market classification, three separate models were built, one for each variable. Generalized estimating equations (GEE) were selected for each of the three models. Consideration was given to linear mixed effects models, but GEE models were preferred based on smaller standard errors and a focus on population-level inferences. Furthermore, given the large number of measurement occurrences, an unstructured covariance matrix could not be fit for any of the models, and assumptions of independent or exchangeable covariance patterns could not be supported. Thus, analysis was limited to an autoregressive (AR1) model. For that reason, the benefit of obtaining the empirical variance estimator despite possible misspecification of the covariance made GEE models more appropriate for the goals of our analysis.

Model 1 is a generalized estimating equation characterizing *market capitalization* in terms of time treated linearly. The linear treatment of time was selected based on comparison to higher order parametric forms and by analyzing a model with smoothed effects. The model with only the main effect of time was determined to be best by a joint test of the coefficients for time and the interaction between time and classification (p=0.33). Model 1 was fit with an AR1 covariance pattern.

Model 2 characterizes *stock price volatility* in terms of a generalized additive mixed model with a random intercept and slope as well as a smoothed effect. This method allows for estimating the degree of smoothness of terms in the generalized additive model as variances of the wiggly components of the smooth terms treated as random effects. Such a model tests $\hat{S}$ instead of $\hat{\beta}$ and treats the coefficients in the spline term as random to provide the best linear unbiased prediction (BLUP). We specify the model:

$$Y_{ij} = \beta_0 + \beta_1(Classification) + S_0(Time) + S_1(Time) + b_{1j} + b_{2j}(Classification) + \epsilon_{ij}$$

Given the fluctuations in *stock price volatility* over time for both emerging and developed markets, fitting a linear model using either LME or GEE did not accurately capture the trends in the data.

Model 3 characterizes *current account balance* in terms of time treated linearly and the main effect of market classification. The linear treatment of time was selected based on comparison to higher order parametric forms and by analyzing a model with smoothed effects. The model with the main effects of time and classification was determined to be best by a Wald test (p=0.4). Model 3 was fit with an AR1 covariance pattern.

## Results

The mean profile plot for *market capitalization* in Figure 1 comparing emerging and developed markets shows similar baseline values for the two groups. After 1995, developed markets show greater increases than emerging markets.

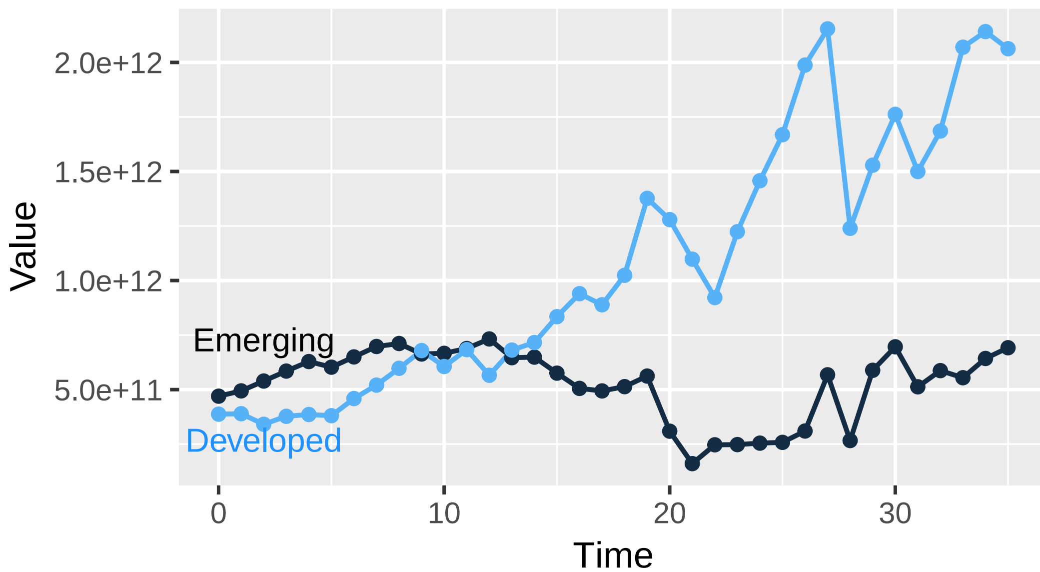


Figure 1: Mean profile response of *market capitalization* for emerging and developed markets

Model 1, summarized in Table 1, shows a marginally significant main effect of time (p=.091). A Wald test indicated that classification and the interaction between classification and time were not significant (p=0.33), suggesting that there is not a significant difference in the rate of change in *market capitalization* over time between emerging and developed countries nor in their baseline values in 1980. Overall, *market capitalization* increases by 26.7 billion dollars for all countries over a one-year span. The estimated AR1 correlation matrix suggests a correlation of 0.951 for measurements in adjacent years. See Figure 5 in appendix for further representation of correlation between measurement occurrences.

Table 1: Model 1 - Market Capitalization

| | Estimate | Empirical SE | p-value |
|---|---|---|---|
| Intercept | 3.86e+11 | 8.43e+10 | <.001 |
| Time | 2.67e+10 | 1.58e+10 | .091 |

The mean profile plot for *stock price volatility* in Figure 2 comparing emerging and developed markets shows lower levels of volatility in developed markets at most time points. However, both groups show several periods of increasing and decreasing volatility across the 35-year span.

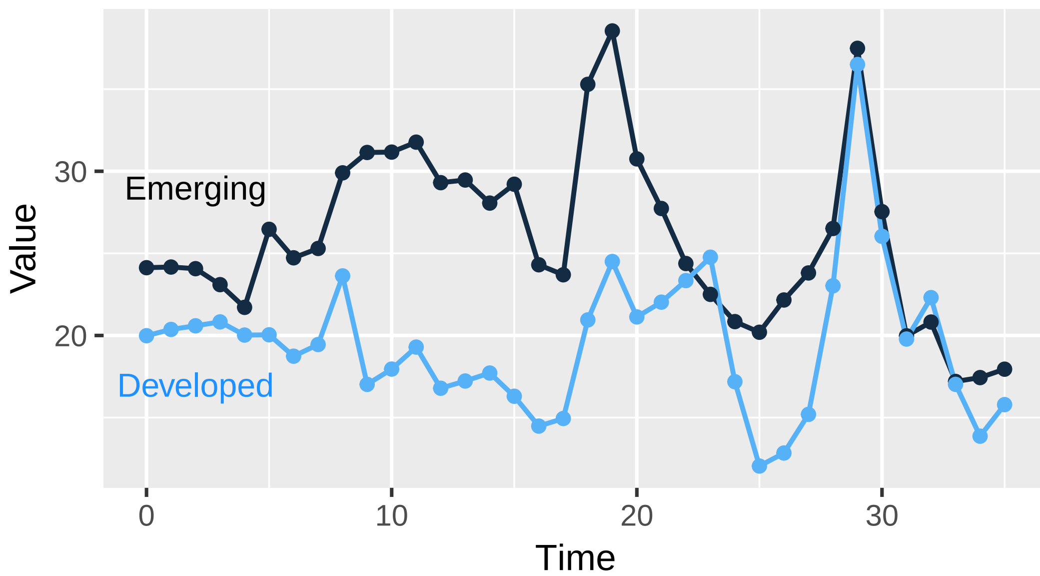


Figure 2: Mean profile response of *stock price volatility* for emerging and developed markets

Model 2, summarized in Table 3 in appendix, shows a significant difference in *stock price volatility* at baseline in 1980 (p<.001), with emerging markets starting at 25.91 and developed markets starting at 19.55. While there is no sufficient way to empirically assess the significance of the difference in rates of change for *stock price volatility* between emerging and developed markets over time, we can make several inferences from the plot of the generalized additive model with integrated smoothness estimation (See Figure 3).

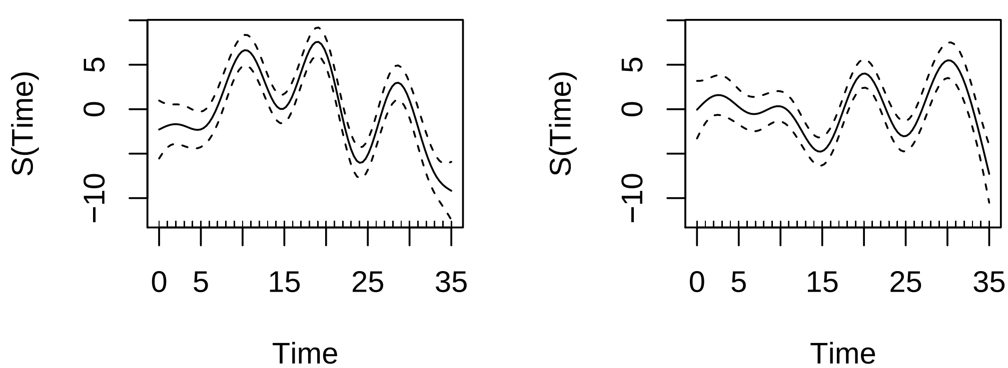


Figure 3: Smooth effects of time (Left: emerging markets; Right: developed markets)

Firstly, there appears to be a significant difference in the rate of change between emerging and developed market countries from 1980-1990, with emerging markets showing larger fluctuations and developed markets remaining relatively centered around 0. Secondly, as the mean response trajectories in Figure 2 showed initially, the stock price volatility for emerging and developed market countries begins to exhibit greater parallelism starting around 1995. Lastly, while the rates of change somewhat align from 2005-2015, emerging market countries demonstrate greater volatility in the negative and developed market countries demonstrate greater volatility in the positive.

The mean profile plot for *current account balance* in Figure 4 comparing emerging and developed markets shows a general increase for both groups across time. Developed markets generally show higher balances than emerging markets.

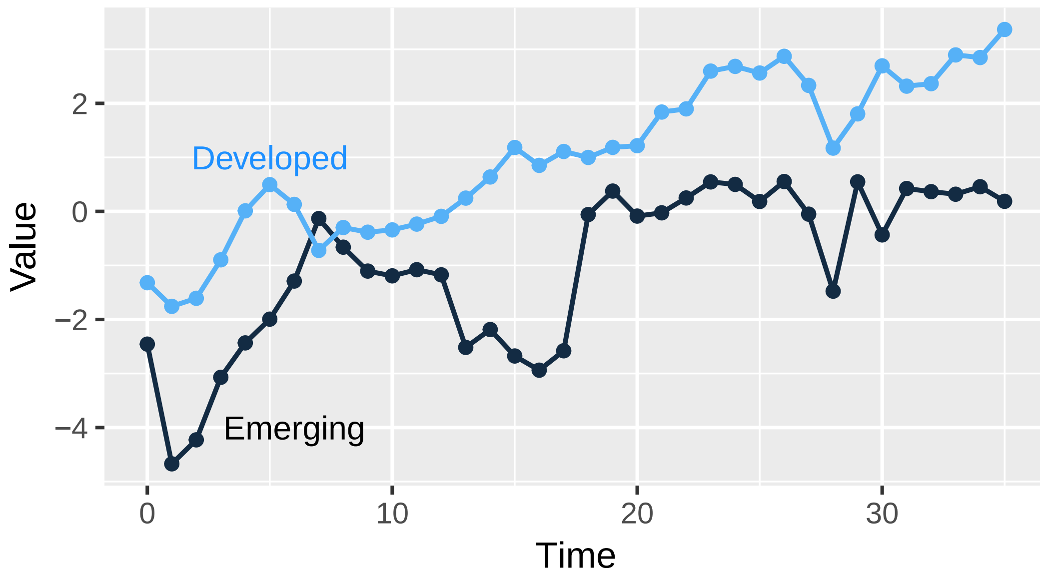


Figure 4: Mean profile response of *current account balance* for emerging and developed markets

Model 3, summarized in Table 2, shows a significant main effect of time ($p<.001$) and a significant main effect of classification ($p=.008$). A Wald test showed that the interaction between time and classification is not significant ($p=0.4$). This suggests that the rate of change in account balance is not different between emerging and developed markets. However, the model shows that developed markets start at a higher baseline balance in 1980 (-0.87) than emerging markets (-2.93). Furthermore, account balance in all countries increases over time; over a one-year span, account balance increases by 0.1078. The estimated AR1 correlation matrix suggests a correlation of 0.903 for measurements in adjacent years. See Figure 6 in appendix for further representation of correlation between measurement occurrences.

Table 2: Model 3 - Current Account Balance

| | Estimate | Empirical SE | p-value |
|---|---|---|---|
| Intercept | -2.928 | 0.535 | <.001 |
| Time | 0.108 | 0.029 | <.001 |
| Classification | 2.063 | 0.776 | 0.008 |

## Discussion

Developed and emerging markets experience similar levels of growth in both market capitalization and current account balance. However, the two groups exhibited different baseline account balances in 1980. While there is not a significant difference between the rates of change in current account balances for emerging versus developed countries, the positive coefficient for time is in line with the historical narrative of emerging market countries in particular moving from account deficits to surpluses between the 1980s and 1990s.[2] Given the rapid expansion of emerging markets economies that become increasingly integrated with developed market economies through the 2000s, it makes sense that developed and emerging market economies show relatively parallel trajectories for market capitalization and current account balance.

As interdependence between emerging and developed countries has increased, market shocks have spread between developed and emerging market countries. The results from our analysis of the differences in the rates of change for stock price volatility between emerging and developed market countries align with this narrative, with the general additive mixed model showing an increasing parallelism between the stock price volatility rates for the two groups over time.[3]

# References

[1] World Development Indicators, The World Bank; International Finance Statistics, International Monetary Fund
[2] Turner, Philip. "Financial globalisation and emerging market capital flows." Participants in the meeting. 2008.
[3] Abou-Zaid, Ahmed S. "Volatility spillover effects in emerging MENA stock markets." Review of Applied Economics 7.1-2 (2011).

## Appendix

Table 3: Model 2 - Stock Price Volatility

| | Estimate | Empirical SE | p-value |
|---|---|---|---|
| Intercept | 3.86e+11 | 8.43e+10 | <.001 |
| Time | 2.67e+10 | 1.58e+10 | .091 |

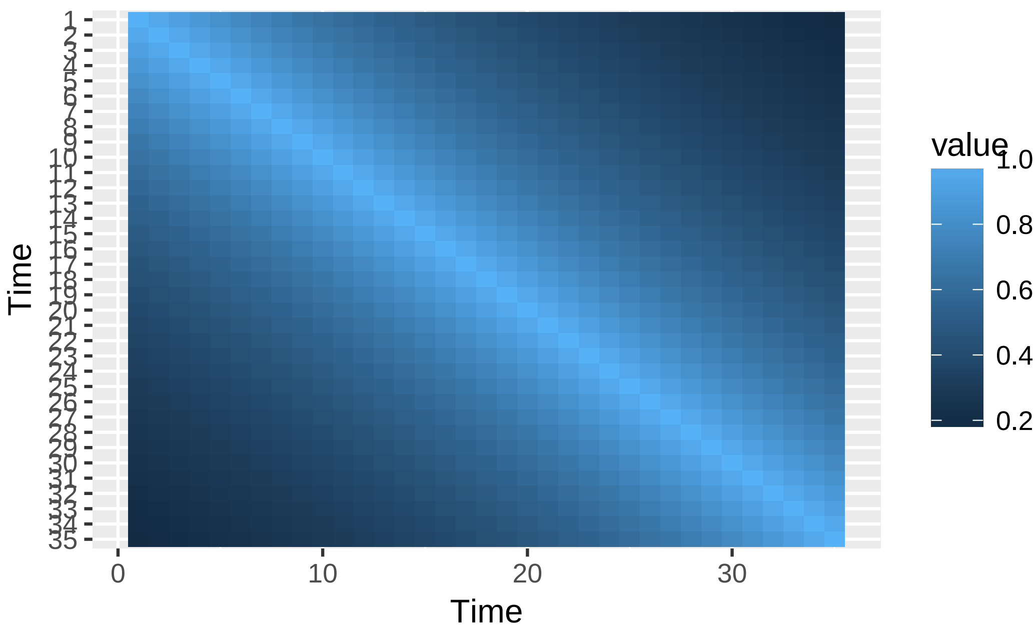


Figure 5: Heat map of correlation matrix for *market capitalization*

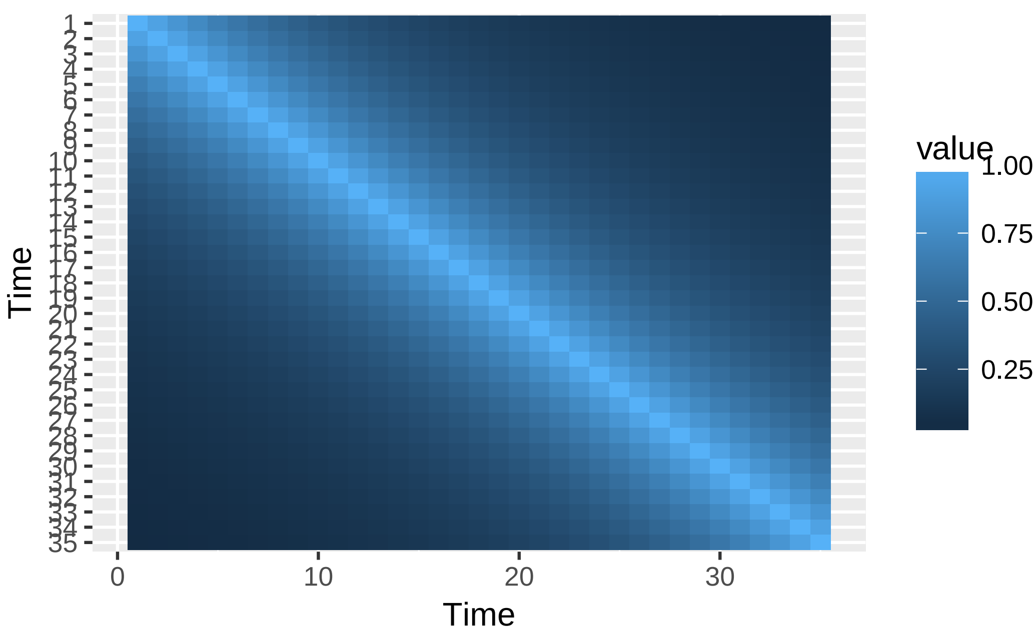


Figure 6: Heat map of correlation matrix for *current account balance*

Code

```
> ##### subset data #####
> macLong <- read.table('MacroLong.txt', header = TRUE)
> mcap <- macLong[which(
+   macLong$Indicator=='Market capitalization of listed domestic companies (current US$)'),]
> colnames(mcap) <- c('Country', 'Country_Code', 'Indicator', 'Indicator_Code', 'Year',
+                     'Value', 'Classification')
> mcap$Time <- rep(0:35, length(unique(mcap$Country)))
> cab <- macLong[which(
+   macLong$Indicator=='Current account balance (% of GDP)'),]
> colnames(cab) <- c('Country', 'Country_Code', 'Indicator', 'Indicator_Code', 'Year',
+                    'Value', 'Classification')
> cab$Time <- rep(0:35, length(unique(cab$Country)))
> spv <- macLong[which(
+   macLong$Indicator=='Stock price volatility'),]
> colnames(spv) <- c('Country', 'Country_Code', 'Indicator', 'Indicator_Code', 'Year',
+                    'Value', 'Classification')
> spv$Time <- rep(0:35, length(unique(spv$Country)))
> ## imputation
> pred        <- gam(Value ~ Classification + s(Time), data=mcap, family=gaussian)
> mcap$Value <- ifelse(is.na(mcap$Value), predict(pred), mcap$Value)
> pred.spv        <- gam(Value ~ Classification + s(Time), data=spv, family=gaussian)
> spv$Value <- ifelse(is.na(spv$Value), predict(pred.spv), spv$Value)
> pred.cab        <- gam(Value ~ Classification + s(Time), data=cab, family=gaussian)
> cab$Value <- ifelse(is.na(cab$Value), predict(pred.cab), cab$Value)
> ## Market Capitalization
> test.model <- gamm(Value ~ Classification + s(Time),
+               random = list(Country = ~ 1), data = mcap)
> plot(test.model$gam)
> # GEE
> model1 <- geeglm(Value ~ Time*Classification, family = gaussian(link = 'identity'),
+                  corstr = 'ar1', id=Country, data = mcap)
> summary(model1)
> model2 <- geeglm(Value ~ Time+Classification, family = gaussian(link = 'identity'),
+                  corstr = 'ar1', id=Country, data = mcap)
> summary(model2)
> anova(model2, model1)
> model3 <- geeglm(Value ~ Time, family = gaussian(link = 'identity'),
+                  corstr = 'ar1', id=Country, data = mcap)
> summary(model3)
> anova(model3, model1)
> ## LME
> modelL1 <- lme(Value ~ Time*Classification, random = ~ 1 + Time | Country, data = mcap)
> summary(modelL1)
> modelL2 <- lme(Value ~ Time*Classification, random = ~ 1 + Time | Country, data = mcap,
+                correlation = corAR1(form = ~1|Country))
> summary(modelL2)
> # correlation matrix
> rho <- 0.951
> cov <- diag(35)
> cov <- rho^abs(row(cov)-col(cov))
> library(reshape2)
> meltcov <- melt(cov)
> heatmap <- ggplot(data = meltcov, aes(x=Var1, y=ordered(Var2, levels = rev(sort(unique(Var2)))),
+                                       fill=value)) + geom_tile()
> heatmap + labs(x="Time",y="Time")
> ## Stock Price Volatility
```

```
> # Assessing and running spline model
> model1 <- gamm(Value ~ Classification + s(Time, by=factor(Classification)),
+                random = list(Country = ~ 1 + Time), data = spv)
> par(mfrow=c(1,2))
> plot(model1$gam)
> summary(model1$gam)
> summary(model1$lme)
> # LME
> model2 <- lme(Value ~ Classification*Time, random = ~ 1 + Time | Country, data = spv)
> summary(model2)
> model3 <- lme(Value ~ Classification*Time, random = ~ 1 | Country, data = spv)
> anova(model2, model3)
> model4 <- lme(Value ~ Classification*Time, random = ~ 1 + Time | Country, data = spv,
+               correlation = corAR1(form = ~1 | Country))
> summary(model4)
> # GEE
> model5 <- geeglm(Value ~ Classification*Time, family = gaussian(link = "identity"),
+                  corstr = 'ar1', id = Country, data = spv)
> summary(model5)
> spv.model2 <- geeglm(Value ~ Time+Classification, family = gaussian(link = 'identity'),
+                      corstr = 'ar1', id=Country, data = spv)
> summary(spv.model2)
> anova(spv.model2, spv.model1)
> spv.model3 <- geeglm(Value ~ Time, family = gaussian(link = 'identity'),
+                      corstr = 'ar1', id=Country, data = spv)
> summary(spv.model3)
> anova(spv.model3, spv.model2)
> ## Current Account Balance
> model6 <- gamm(Value ~ Classification + s(Time), random = list(Country = ~ 1), data = cab)
> plot(model6$gam)
> # LME
> model7 <- lme(Value ~ Classification*Time, random = ~ 1 + Time | Country, data = cab)
> summary(model7)
> model8 <- lme(Value ~ Classification*Time, random = ~ 1 | Country, data = cab)
> model9 <- lme(Value ~ Classification*Time, random = ~ 1 + Time | Country, data = cab,
+               correlation = corAR1(form = ~1 | Country))
> summary(model9)
> # GEE
> cab.model1 <- geeglm(Value ~ Time*Classification, family = gaussian(link = 'identity'),
+                  corstr = 'ar1', id=Country, data = cab)
> summary(cab.model1)
> cab.model2 <- geeglm(Value ~ Time+Classification, family = gaussian(link = 'identity'),
+                  corstr = 'ar1', id=Country, data = cab)
> summary(cab.model2)
> anova(cab.model2, cab.model1)
> cab.model3 <- geeglm(Value ~ Time, family = gaussian(link = 'identity'),
+                  corstr = 'ar1', id=Country, data = cab)
> summary(cab.model3)
> anova(cab.model3, cab.model2)
> #correlation
> rho <- 0.903
> cov <- diag(35)
> cov <- rho^abs(row(cov)-col(cov))
> meltcov <- melt(cov)
> heatmap <- ggplot(data = meltcov, aes(x=Var1, y=ordered(Var2, levels = rev(sort(unique(Var2))))),
+                                        fill=value)) + geom_tile()
> heatmap + labs(x="Time",y="Time")
```