DYNAMIC CONDITIONAL BETA

Robert Engle[1]

June 13, 2012

I. Introduction

In empirical finance and in time series applied economics in general, the least squares model is the workhorse. In class there is much discussion of the assumptions of exogeneity, homoskedasticity and serial correlation. However in practice it may be unstable regression coefficients that are most troubling. Rarely is there a credible economic rationale for the assumption that the slope coefficients are time invariant.

Econometricians have developed a variety of statistical methodologies for dealing with time series regression models with time varying parameters. The three most common are rolling window estimates, interaction with trends, splines or economic variables, and state space models where the parameters are treated as a state variable to be estimated by some version of the Kalman Filter. Each approach makes very specific assumptions on the path of the unknown coefficients. The first approach specifies how fast the parameters can evolve, and by using least squares on each moving window, employs an inconsistent set of assumptions. The second specifies a family of deterministic paths for the coefficients that may have undesirable or inconsistent implications particularly when extrapolated. The third requires specifying a

stochastic process for the latent vector of parameters which may include unit roots and stochastic trends that are generally unmotivated and rarely based on any economic analysis.

There is no standardized approach that has become widely accepted. This paper will propose such an approach for a wide class of data generating processes. In addition, it will allow a test of the constancy of the parameter vector.

II. Dynamic Conditional Beta

Consider a vector of observables, $\left(y_t, x_{1,t}, ... x_{k,t} \right), t = 1, ... T$. The objective is to characterize the conditional distribution of y given x. In a time series context this requires finding $E_{t-1}\left(y_t \middle| x_t \right)$. In the very special case we generally assume, this may turn out to be linear in x with parameters that do not depend upon the past information set. Much more generally, this regression may be linear with parameters that do depend upon the past information set.

A natural formulation of this problem is in terms of the joint distribution of all the variables conditional on the past. Suppose

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} \middle| F_{t-1} \sim N\left(\mu_t, H_t \right) \tag{1}$$

where $\left(\mu_t, H_t \right)$ are measurable with respect to F_{t-1}. Then the desired conditional distribution is simply

$$y_t \middle| x_t, F_{t-1} \sim N\left(\mu_{y,t} + H_{yx,t} H_{xx,t}^{-1} \left(x_t - \mu_{x,t} \right), H_{yy,t} - H_{yx,t} H_{xx,t}^{-1} H_{xy,t} \right) \tag{2}$$

where subscripts represent natural partitions. Thus the time varying regression coefficients that are needed are

$$\beta_t = H_{xx,t}^{-1} H_{xy,t} \tag{3}$$

The betas from this expression will be called Dynamic Conditional Betas. These betas have been used in univariate settings by many authors beginning with Bollerslev, Engle and Wooldridge(1988) and including Engle(2009) and Brownlees and Engle(2011). This author does not recall seeing the multivariate expression in (3) before.

The expression however is very familiar if the time subscript is dropped. In this case we simply have ordinary least squares. If we use rolling regressions, then the matrices are changing and we can think of the regression coefficients as being approximations to (3). The assumption in (1) makes it clear exactly what the rolling covariance matrices should accomplish. These matrices should give conditional covariances of y with x and of x with x. Rolling windows are often used to compute "historical" volatilities and correlations which are viewed as forecasts. The time width of the window is a central feature of historical volatilities as it regulates the volatility of volatility. Naturally, the criticisms of historical volatility forecasts apply here as well. That is, a window of k days is the correct choice if the joint vector of returns on the next day is equally likely to come from any of the previous k days but could not be from returns more than k days in the past. Because of the unrealistic nature of this specification, exponential smoothing is a better alternative for most cases.

The normality assumption is obviously restrictive but even without normality, (2) can be interpreted as the conditional linear projection. The projection may be linear but with a non-normal error which will support estimation as in Brownlees and Engle(2011). The more difficult issue is that we do not observe directly either the vector of conditional means or conditional covariances. A model is required for each of these. For many financial applications where the data are returns, the mean is relatively unimportant and attention can be focused on the conditional covariance matrix. Think for a moment of the estimate of the beta of a stock or

portfolio, or multiple betas in factor models, style models of portfolios, or the pricing kernel, and many other examples. In these cases, a natural approach is to use a general covariance matrix estimator and then calculate the betas.

In the special case where one or more regression coefficients are time invariant, there is a restriction on the covariance matrix. If the ith coefficient is time invariant then

$$\iota_i ' H_{xx,t}^{-1} H_{xy,t} = \beta_i \tag{4}$$

where ι_i is the ith column of the identity matrix and the fact that the right hand side does not depend upon time is captured by the lack of a time subscript. When there is only one regressor, this can be more simply be seen as

$$\frac{H_{yx,t}}{H_{xx,t}} = \beta \tag{5}$$

so that the ratio of a covariance to the variance is constant.

III. Maximum Likelihood Estimation of DCB

Maximum likelihood estimation of the Dynamic Conditional Beta model requires specification of both the covariance matrix and the mean of the data vector. In general, the covariance matrix and mean of these data will include unknown parameters. Extending equation (1) to include a jx1 vector of unknown parameters in the mean and variance equations, we get

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} \Big| F_{t-1} \sim N\left(\mu_t(\theta), H_t(\theta)\right), \quad \underset{(jx1)}{\theta} \in \Theta \tag{6}$$

Clearly, this specification includes the wide range of multivariate volatility models that are used in practice such as VEC, BEKK, DCC and general classes of mean functions. In most cases, the model does not satisfy assumptions for weak exogeneity as in Engle Hendry and Richard(1985)

because the distribution of the x variables has information about the dynamic conditional betas. In special cases where the covariance matrix has no unknown parameters or these can be partitioned in special ways, weak exogeneity can be established.

The MLE of q can be computed for the data vector $z_t = (y_t, x_t')'$ as

$$\hat{q} = \arg\max -\frac{1}{2}\sum_{t=1}^{T} \log\left|H(q)_t\right| + (z_t - m(q)_t)'H(q)_t^{-1}(z_t - m(q)_t) \tag{7}$$

Then as b_t is a function of the unknown parameters and data, the MLE is simply

$$\hat{b}_t = H\left(\hat{q}\right)_{xx,t}^{-1} H\left(\hat{q}\right)_{xy,t} \tag{8}$$

A useful expression for the log likelihood can be found by changing variables. Convert the likelihood from (y_t, x_t') to (w_t, x_t') where $w_t = y_t - H_{yx,t}H_{xx,t}^{-1}x_t$. As the Jacobian of the transformation is 1, the likelihood simply has a recalculated covariance matrix which is now diagonal. Setting the means to zero makes the expression easier and gives:

$$L(y,x;q) = L(w,x;q) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log(h_{w,t}) + \frac{w_t^2}{h_{w,t}}\right) + \left[\frac{1}{2}\sum_{t=1}^{T} \log\left|H_{xx,t}\right| + x_t'H_{xx,t}^{-1}x_t\right] \tag{9}$$

The asymptotic standard errors can be developed in the same way. From (7), the limiting distribution of $\hat{q}$ can be established as

$$\sqrt{T}\left(\hat{q} - q_0\right) \otimes N\left(0, A^{-1}\right) \tag{10}$$

where A could be the information matrix if the normality assumption is valid or a robust covariance matrix of the sandwich form if it is not. See for example Bollerslev and Wooldridge(1992) for a proof. Applying the mean value theorem to (8)

$$\hat{b}_t - b_{0,t} = G_t\left(\bar{q}\right)\left(\hat{q} - q_0\right), \; G_t\left(\bar{q}\right) = \frac{\P\left[H(q)_{xx,t}^{-1} H(q)_{xy,t}\right]}{\P q'}\Bigg|_{q=\bar{q}} \tag{11}$$

where $\bar{q} \in (q_0, \hat{q})$ with possibly different values for different rows. Then

$$\sqrt{T}\left(\hat{b}_t - b_{0,t}\right) \otimes N\left(0, G_t A^{-1} G_t'\right) \tag{12}$$

and G_t can be estimated as $G_t(\hat{q})$.   Notice that this distribution applies at each point within the sample as the sample size goes to infinity. The large sample result delivers consistent estimates of the unknown fixed parameters q which then imply time varying standard errors for the betas.

IV. Maximum Likelihood when all betas are constant

In this classical case, the likelihood function can again be evaluated based on equation (1) where the specification includes the assumption that the betas are all constant. For convenience, we again assume that all means are zero. Now we write the log likelihood as the sum of the log of the conditional density of y given x plus the log of the marginal density of x. Ignoring irrelevant constants gives:

$$L(y, x; q) = \sum_{t=1}^{T} \left[\log f_{y|x}\left(y_t \middle| x_t, z_{t-1}, \ldots; q\right) + \log f_x\left(x_t \middle| z_{t-1}, \ldots; q\right) \right]$$

$$= -\frac{1}{2} \sum_{t=1}^{T} \left(\log(h_{u,t}) + \frac{(y_t - x_t'b)^2}{h_{u,t}} \right) - \frac{1}{2} \sum_{t=1}^{T} \left(\log\left|H(q)_{xx,t}\right| + x_t' H(q)_{xx,t}^{-1} x_t \right) \tag{13}$$

In this equation

$$u_t = y_t - x_t'b, \quad V_{t-1}(u_t) = h_{u,t} \tag{14}$$

If h_u is a general volatility process with no common parameters from H_{xx} then maximum likelihood estimators of beta are simply given by maximizing the first part of the likelihood. This is just least squares with a heteroskedastic error process.

The asymptotic distribution of beta is simply given by the conventional expression for heteroskedastic regression.

V. Maximum Likelihood when some parameters are constant

If it is known that a subset of the regression parameters are constant, then MLE can be adapted to estimate jointly the fixed and varying parameters of the model. Suppose there are k regressors and the first k_1 are time varying and the second k_2 are constant. In this case the dynamic conditional betas of the first k_1 elements will be the first k_1 elements of the expression in (4). Partitioning $x = (x_1 ', x_2 ')'$ and then defining $x_1^0 = (x_1 ',0)'$, the model can be expressed as

$$y_t = H_{yx,t} H_{xx,t}^{-1} x_{1,t}^0 + \beta_2 ' x_{2,t} + u_t \qquad (15)$$

Thus the conditional log likelihood can be evaluated by regressing y minus the first term on the fixed regressors and allowing heteroskedasticity in u. The result is:

$$L(y;x) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log(h_{u,t}) + \frac{u_t^2}{h_{u,t}}\right] - \left\{\frac{1}{2}\sum_{t=1}^{T}\left[\log\left|H(\theta)_{xx,t}\right| + x_t ' H(\theta)_{xx,t}^{-1} x_t\right]\right\} \qquad (16)$$

Notice that the second term in curly brackets is the same in (9) (13) and (16). In each case, the estimates of the covariance matrix of the x's will be identical so any comparison of log likelihoods can be done simply by comparing the marginal log likelihoods.

VI. Testing for a constant regression coefficient

In this section a set of approaches will be developed for testing that one or more parameters is constant. These hypotheses are basically non-nested although they more precisely are overlapping models. As a non-nested hypothesis testing problem, there are several standard approaches. The first developed initially by Cox(1961,1962) is a model selection approach

which simply asks which model has the highest value of a (penalized) likelihood. If there are two models being considered, there are two possible outcomes.

A second approach introduced by Vuong(1989) is to form a two tailed test of the hypothesis that the models are equally close to the true model. In this case, there are three possible outcomes, model 1, model 2, and that models are equivalent.

A third approach is artificial nesting as discussed in Cox(1961,1962) and advocated by Aitkinson(1970) An artificial model is constructed that nests both special cases. When this strategy is used to compare two models, there are generally four possible outcomes – model 1, model 2, both , neither. Artifical nesting of hypotheses is a powerful approach in this context as the model has a very simple form. For the multivariate case the natural artificial nesting consists of estimating the following equation:

$$y_t = x_t'\beta + x_t'\left(\beta_t \square \gamma\right) + \sqrt{h_{u,t}}\,u_t \tag{17}$$

where e indicates a Hadamard product or element by element multiplication. A Wald test of the hypothesis that the coefficient is constant corresponds to a t-test that gamma is zero. Similarly, a Wald test of the hypothesis that the coefficient is time varying corresponds to the t-test that beta is zero in the same regression. This testing approach leads as always to the possibility that both hypotheses or neither hypothesis could be rejected. If both are rejected, then it may be useful to consider the estimate of equation (17) as the accepted model. If neither is rejected, then it might be natural to test them jointly and consider whether x in any form is a useful explanatory variable.

VII. Overlapping Models and Econometric Solutions

Consider two data generating processes for the same vector of time series

$$z_t = (y_t', x_t')', \quad t = 1, ..., T \tag{18}$$

In the simplest case, there is only one endogenous variable and k conditioning variables. These can be expressed in terms of the joint distributions conditional on past information as:

$$z_t \big| F_{t-1} \sim f(z_t, \theta), \quad \theta \in \Theta$$
$$z_t \big| F_{t-1} \sim g(z_t, \phi), \quad \phi \in \Phi \tag{19}$$

Without loss of generality we can write each distribution conditional on the past as the product of a conditional and a marginal distribution each conditioned on the past.

$$f(z_t; q) = f_{y|x}(y_t | x_t; q) f_x(x_t; q)$$
$$g(z_t; f) = g_{y|x}(y_t | x_t; f) g_x(x_t; f) \tag{20}$$

The MLEs are given by

$$L_f(\hat{q}) = \sum_{t=1}^{T} \log f_{y|x}(y_t | x_t; \hat{q}) + \log f_x(x_t; \hat{q})$$
$$L_g(\hat{f}) = \sum_{t=1}^{T} \log g_{y|x}(y_t | x_t; \hat{f}) + \log g_x(x_t; \hat{f}) \tag{21}$$

Model f is said to be nested within g if for every $\theta_0 \in \Theta$ there is a $\phi^* \in \Phi$ such that $f(z, \theta_o) = g(z, \phi^*)$ so that events have the same probability whether $f(z, \theta_o)$ or $g(z, \phi^*)$ is the true DGP. A similar definition applies to the notion that g is nested within f. The models f and g are partially non-nested (Cox(1960),1961), Pesaran(1999)) or overlapping (Vuong(1989)) if there are some parameters in $Q \times F$ that lead to identical distributions. Vuong proposes a sequential procedure that tests for such parameters first and upon rejecting these cases, examines the purely non-nested case. The parameters θ and ϕ include parameters of the covariance matrix that must be estimated as well as parameters of the regression function such as the betas. The parameters of the covariance matrix could be the parameters of a multivariate GARCH

model if that is how the estimation is conducted. If a DCC approach is used, then the parameters would include both GARCH parameters and DCC parameters. From equation (3) and knowledge of a wide range of multivariate GARCH models, it is apparent that the only point in the parameter space where the fixed parameter and varying parameter are equivalent is the point where there is no heteroskedasticity in the observables. Even models like constant conditional correlation, BEKK, VEC and DECO have no parameters that will make the betas in (3) constant unless both $H_{xx,t}$ and $H_{xy,t}$ are themselves constant. Thus it is essential to determine whether there is heteroskedasticity in y and x to ensure that this is a non-nested problem.

As discussed above, several approaches to model building for non-nested models will be applied.

a) Model Selection

From the maximized likelihood in and a penalty function such as Schwarz or Aikaike, the selected model can be directly found. In most cases the likelihood will have the same specification of the conditioning variables, x, and consequently the maximized values will satisfy

$$\left\{ \sum_{t=1}^{T} \log\left(f_{x,t}\left(x_t; \hat{\theta} | F_{t-1} \right) \right) \right\} = \left\{ \sum_{t=1}^{T} \log\left(g_{x,t}\left(x_t; \hat{\gamma} | F_{t-1} \right) \right) \right\} \qquad (22)$$

Therefore model selection will only involve the first term in the log likelihood just as in a conventional regression. Thus model selection can be based on

$$LR_T = \sum_{t=1}^{T} \log\left(f_{y|x,t}\left(y_t; \hat{\theta} | x_t, F_{t-1} \right) \right) - \sum_{t=1}^{T} \log\left(g_{y|x,t}\left(y_t; \hat{\gamma} | x_t, F_{t-1} \right) \right) - \left(pen_f - pen_g \right) \qquad (23)$$

where $\left(pen_f - pen_g \right) = \left(k_f - k_g \right)$ for Aikaike Information Criterion (AIC) and

$= \left(k_f - k_g \right) * \log(T) / 2$ for Schwarz or Bayesian Information Criterion, BIC. Here k is the total

number of estimated parameters in each model. Clearly, as T becomes large, the penalty becomes irrelevant in model selection.

b) Testing Model Equality.

Following Vuong(1989) and Rivers and Vuong(2002), the model selection procedure can be given a testing framework. The criterion in (23) has a distribution under the null hypothesis that the two likelihoods are the same.   The hypothesis that the likelihoods are the same is interpreted by Vuong as the hypothesis that they are equally close in a Kulbach Leibler sense, to the true DGP. He shows that under this null, the LR criterion converges to a Gaussian distribution with a variance that can be estimated simply from the maximized likelihood ratio statistic. Letting

$$m_t = \log \left(\frac{f_{y|x,t}\left(y_t; \hat{\theta} | x_t, F_{t-1} \right)}{g_{y|x,t}\left(y_t; \hat{\gamma} | x_t, F_{t-1} \right)} \right) \tag{24}$$

then

$$\hat{\omega}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(m_t - LR_T / T \right)^2 \tag{25}$$

and it is shown in Theorem 5.1 that

$$T^{-1/2} LR_T / \hat{\omega} \xrightarrow{\ d\ } N(0,1) \tag{26}$$

This can be computed as the t-statistic on the intercept in a regression of m_t on a constant. Based on this limiting argument, three critical regions can be established – select model f, select model g, they are equivalent.

To apply this approach we substitute the log likelihood of the DCB model and the fixed coefficient GARCH model into (24). From (13) and (16) we get

$$m_t = L_{f,t} - L_{g,t} = -.5\left(\log\left(h_{f,t} / h_{g,t}\right) + e_{f,t}^2 - e_{g,t}^2\right) \qquad (27)$$

If a finite sample penalty for the number of parameters is adapted such as the Schwarz criterion, then the definition of m becomes

$$m_t = -.5\left(\log\left(h_{f,t} / h_{g,t}\right) + e_{f,t}^2 - e_{g,t}^2\right) - \left(k_f - k_g\right)\log(T)/T \qquad (28)$$

This series of penalized log differences is regressed on a constant and the robust t-statistic is reported. If it is significantly positive, then model f is preferred and if significantly negative, then model g is preferred. As m in equation (28) is likely to have autocorrelation, the t-statistic should be computed with a HAC standard error that is robust to autocorrelation and heteroskedasticity. See Rivers and Vuong(2002).

c) The third approach is based on artificial nesting and, as discussed above in (17), is simply an augmented regression model which allows for heteroskedasticity. From asymptotic standard errors, tests of both f and g null models can be constructed. The simplicity and power of this approach make it appear to be the most useful way to apply the DCB model. Although this regression has regressors that depend upon estimated parameters, this does not lead to the generated regressor problem of Pagan(1984). A theorem from Wooldridge(2002) however may be useful to show that parameter estimates are consistent and inference is supported.

The nested DCB model can be expressed as

$$y_t = x_t \beta_0 + \left(x_t \beta_t\left(\theta_0\right)\right)\gamma_0 + \varepsilon_t = \left(x_t, x_t \beta_t\left(\theta_0\right)\right)\phi_0 + \varepsilon_t \qquad (29)$$

When the parameter $\theta = \theta_0$. The model that is estimated however uses the estimated value of $\theta = \hat{\theta}$. Wooldridge(2002) establishes in section 12.4 the conditions for consistency and correct inference in this model. Consistency follows simply from the consistency of $\hat{\theta}$. To establish that inference is correct requires the condition that

$$E\left(\frac{\partial^2 L_t(\phi,\theta)}{\partial\phi\partial\theta}\right)\bigg|_{\phi=\phi_0,\theta=\theta_0} = 0 \qquad (30)$$

Under the standard likelihood assumptions and under the condition (10) this term in the Hessian can be expressed as

$$\frac{\partial^2 L_t(\phi,\theta)}{\partial\phi\partial\theta}\bigg|_{\theta=\theta_0,\phi=\phi_0} = \frac{\partial\left[\varepsilon_t\left(x_t,x_t\beta_t(\theta)\right)\right]}{\partial\theta} = \varepsilon_t\left(0,x_tG_t(\theta_0)\right) - \left(x_t,x_t\beta_t(\theta_0)\right)'x_tG_t(\theta_0)\gamma_0 \qquad (31)$$

Because G consists only of lagged variables, the first term has expected value zero. The second term will generally be non-zero except when $\gamma_0 = 0$. Thus inference on the nested DCB model can be done without correcting the standard errors when $\gamma_0 = 0$ which is the fixed beta null hypothesis. Thus the artificially nested model is perfectly appropriate for testing the null hypothesis of fixed betas and is consistent under the alternative, but inference under the alternative should be adjusted. An approach to this inference problem can be developed following Wooldridge but this has not yet been worked out.

VI. Asset Pricing and the DCB

Multifactor asset pricing theories begin with a pricing kernel and then derive cross sectional and time series implications that can be tested. This derivation is completely consistent with the Dynamic Conditional Beta as described above.

Let (r_t, f_t) be vectors of nx1 asset returns and kx1 pricing or risk factors respectively. The pricing factors are taken to be tradable with returns given by f_t. A pricing kernel can be specified as

$$m_t = a_t - b_t' f_t \qquad (32)$$

13

where the parameters a and b potentially change over time but the changes are not priced. From

the pricing kernel it is straight forward to derive the following expression:

$$
\begin{aligned}
1 &= E_{t-1}\left(m_t\left(r_t+1\right)\right), \quad 1+r_t^f = 1/E_{t-1}\left(m_t\right) \\
E_{t-1}\left(r_t - r_t^f\right) &= \left(1+r_t^f\right)Cov_{t-1}\left(r_t, f_t\right)b_t \\
&= b_{r,f,t}\left(1+r_t^f\right)Var_{t-1}\left(f_t\right)b_t \\
&= b_{r,f,t}E_{t-1}\left(f_t\right)
\end{aligned}
\tag{33}
$$

The last expression is based upon the assumption that the factors are priced with the same pricing

kernel. Notice that in this derivation

$$
b_{r,f,t} = Cov_{t-1}\left(r_t, f_t\right)\left[V_{t-1}\left(f_t\right)\right]^{-1}
\tag{34}
$$

which is exactly the formula for DCB as in (3). This nxk matrix has the betas from regressions

of asset returns on factor returns and these betas are conditional on past information. The themes

of asset pricing come through completely in that the expected return on an asset is linear in beta

and depends upon the risk premium embedded in the factor.

In Bali and Engle(2012) this empirical model is challenged to explain cross sectional

returns in a one factor dynamic model. It performs very well providing new evidence of the

power of beta when it is estimated dynamically. Furthermore, they find strong evidence of

hedging demand in the conditional ICAPM. In Bali and Engle(2010) a similar model is

examined from primarily a time series point of view and again evidence is found for hedging

demands. It is found that the most useful second factor is a volatility factor.

VII. Asset Pricing for Industry Portfolios

To illustrate the usefulness of the DCB in an asset pricing context, the 12 industry

portfolios will be examined in the context of the Fama French 3 factor model. Data are from

Ken French's web site and cover the period 1963-2011 which is over 12,000 daily observations per return series. For each industry the following model is estimated:

$$r_t^j - r_t^f = a_j + b_{j,m}\left(r_t^m - r_t^f\right) + b_{j,hml}r_t^{hml} + b_{j,smb}r_t^{smb} + \sqrt{h_t^j}\,e_t^j \tag{35}$$

The model is estimated under a variety of assumptions:

   a) OLS with constant coefficients and robust standard errors

   b) GJR-GARCH with constant coefficients

   c) DCB with regression coefficients from (3) and covariance matrices estimated with DCC for the factors and bivariate DCC for the correlations between the dependent variable and each of the factors. The GJR-GARCH model is used for each series.

   d) NESTED DCB includes each factor both with a constant coefficient and the time varying coefficient.

Table 1 shows the test criteria for choosing between constant betas and time varying betas for each of the 12 industries. The model selection criterion simply chooses the model with the highest value of the penalized likelihood. In this case, seven industries are better fit with constant betas and five with time varying betas. The Vuong test examines whether these differences are significant. These are t-ratios and consequently at conventional levels, there are four industries with significant improvement using purely time varying betas and five with significantly better performance with constant betas.

When the nested model is applied there are t-statistics for each beta corresponding to pure time variation and pure fixed beta. The t-statistics in Table 2 indicate that almost all the coefficients are significantly different from zero. This means that the null hypothesis of zero fixed beta as well as the null hypothesis of no time variation are both rejected almost every time. There is clear evidence that time varying betas add to the explanatory power of this regression.

15

Interestingly, the coefficients on the time varying betas are typically less than one in this application indicating the advantage of some shrinkage.

From an asset pricing perspective, it is interesting to learn if the DCB model estimates are more consistent with asset pricing theory than fixed parameter estimates. The multifactor pricing model predicts that the alpha of these regressions should be zero. Table 3 presents t-statistics for each of these models from the full time series.

Table 1

Testing for Constant Betas

NAMEN	SCHWARZ_DCB	SCHWARZ_GARCH	VUONGTEST
buseq	1.617244	1.706647	6.308966
chems	1.118553	1.131116	1.286670
durbl	1.899514	1.893769	-0.316227
enrgy	2.039066	2.076260	2.775102
hlth	1.522259	1.433991	-5.669738
manuf	0.605026	0.348881	-8.384324
money	0.950317	0.875159	-3.513272
nodur	0.686547	0.873242	14.30646
other	0.781455	0.620909	-6.510039
shops	1.191182	1.172615	-1.317494
telcm	1.813173	1.787497	-2.081003
utils	0.915316	1.011304	6.572574

Table 2

Tests of Nested DCB Model

NAMEN	FIX1	FIX2	FIX3	DCB1	DCB2	DCB3
buseq	31.2	-11.1	-3.24	38.4	36.7	30.2
chems	23.3	-2.79	-10.6	25.9	30.0	10.9
durbl	32.0	16.7	0.92	26.8	26.1	26.7
enrgy	32.7	6.59	-12.8	33.1	41.0	16.4
hlth	32.7	-21.7	-16.7	22.9	26.0	18.8
manuf	88.0	16.7	19.6	20.5	23.1	11.2
money	52.6	24.7	19.0	53.6	47.0	35.5
nodur	17.0	-2.12	-3.10	40.9	37.0	49.6
other	42.4	2.67	54.8	24.0	45.5	21.4
shops	20.8	-6.86	7.23	26.6	28.0	15.4
telcm	23.0	6.87	-16.5	26.0	27.8	12.1
utils	24.2	11.3	2.26	50.8	46.7	33.2

Table 3

Testing that Alpha is Zero in the Three Factor FF Model

NAMEN	TSTAT_DCB	TSTAT_GARCH	TSTAT_NEST	TSTAT_OLS
buseq	-0.19	1.63	0.54	2.46
chems	-2.03	0.76	-0.72	0.87
durbl	-4.00	-2.70	-3.52	-2.52
enrgy	-1.33	2.16	0.24	0.85
hlth	1.93	5.04	3.88	3.42
manuf	-3.19	-1.82	-2.66	-0.35
money	-2.51	-2.01	-2.86	-3.10
nodur	-0.65	4.15	0.38	3.74
other	-1.70	-2.60	-2.59	-2.75
shops	0.32	2.08	0.66	1.76
telcm	-0.27	0.69	0.01	-0.14
utils	-4.13	-2.55	-3.86	-0.39

It is clear from this table that many of the alphas are significant for particular industries regardless of how the model is estimated. From the NESTED DCB model, 5 industries have significantly negative alphas and one, health, has a significantly positive alpha. This is slightly better than for the fixed parameter GARCH where there are 4 significantly negative and 4 significantly positive alphas.

A few pictures reveal differences across sectors in the time series patterns of their betas over the last six years. From these figures, it is clear that typically the beta on hml is the most variable.   Business Equipment has a negative beta on hml so trades like a growth stock. Energy appears to be a value stock in 2006 and early 2007 but then switches to neutral. Manufacturing loads very little on either hml or smb and Money begins neutral and then rises to be a strong value stock as the market declines during the financial crisis. The market beta rises for the Money sector during the financial crisis but not nearly as much as it does in a one factor model.
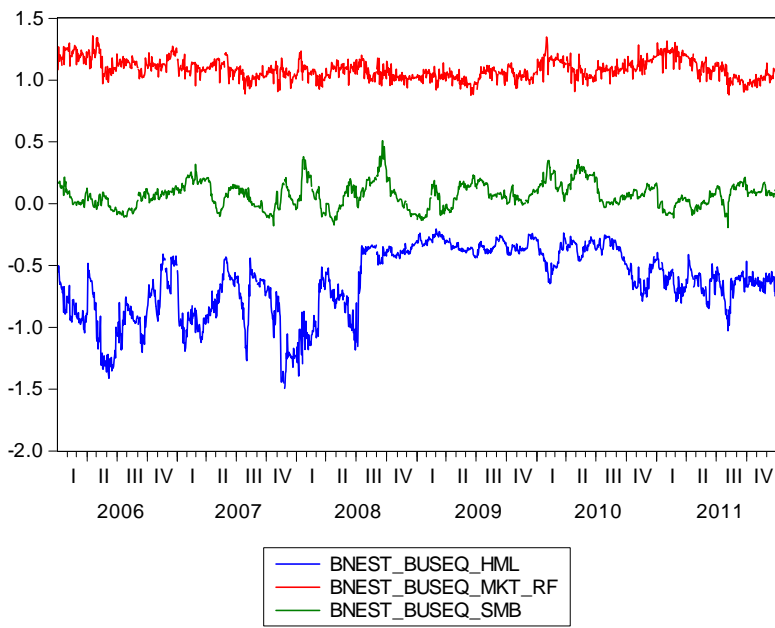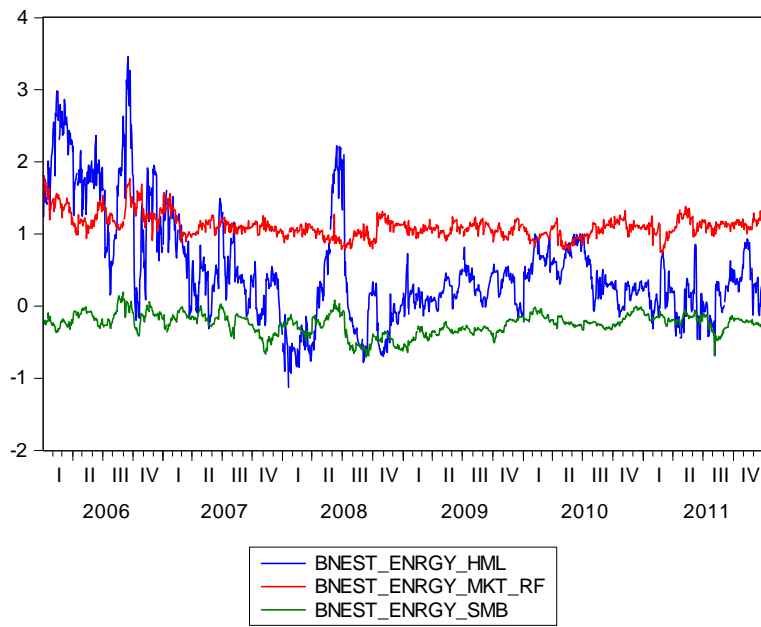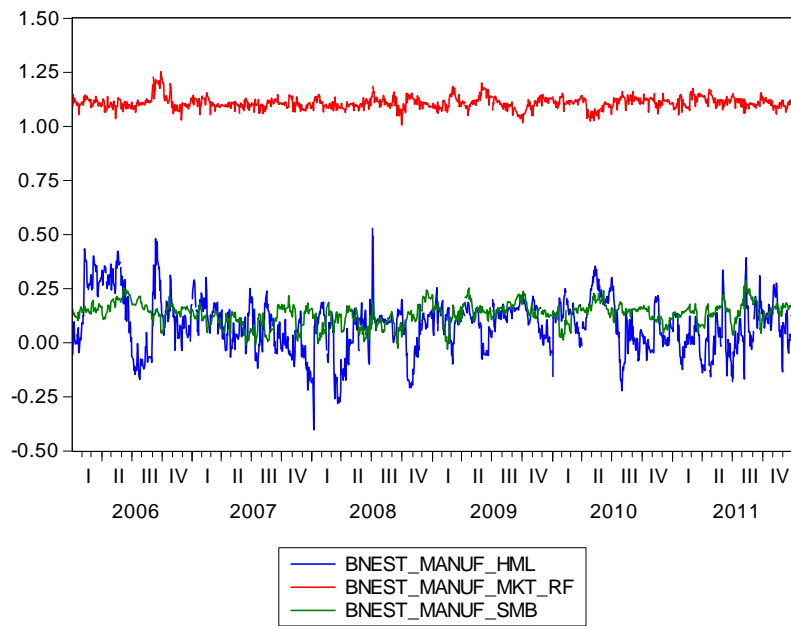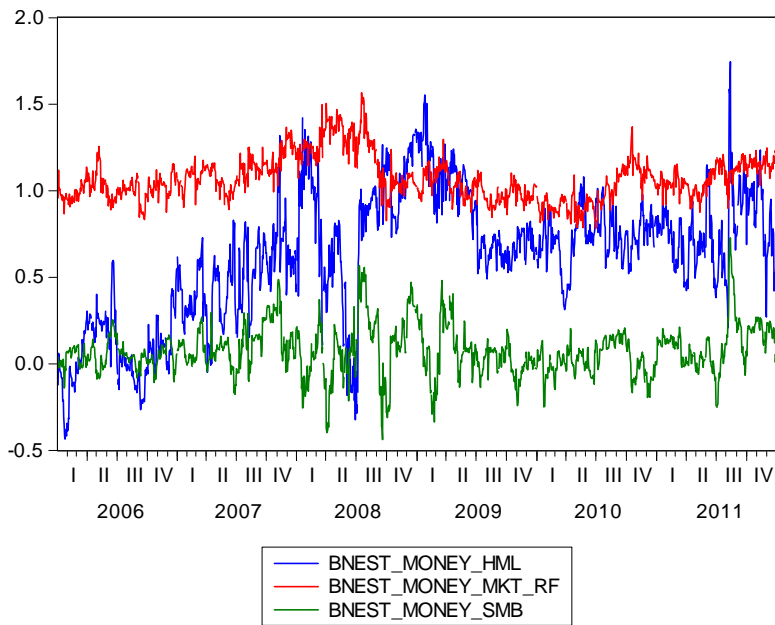
Figure 1

Figure 2

Figure 3

Figure 4

20

VII. Global Systemic Risk with DCB

A central question in the analysis of systemic risk is the following: "how much capital would a firm need to raise in order to carry on its business, if we have another financial crisis?" This is a question of systemic risk because the only source of capital in a financial crisis is likely to be the taxpayer. The taxpayer will consider the consequences of bailing out this institution or letting it fail. In a financial crisis, this institution will not be alone in requesting capital. The bigger the total request the more severe the situation as the real economy cannot function without a viable financial sector. Prudent regulators will foresee this event and require a sufficient cushion that the firms do not need to raise capital, or at least not as much, in a crisis. Financial firms will also foresee this event however their costs are limited to their salaries and ownership positions and therefore do not include the costs imposed on the rest of society. This systemic risk externality makes it clear that risks rationally taken by financial firms may be greater than are socially optimal. This argument is developed in more detail in Acharya, Pedersen, Phillipon and Richardson(2010).

To measure the capital shortfall, Brownlees and Engle(2010) propose a time series approach which essentially estimates the beta of a firm equity on a broad market index. They estimate a bivariate volatility model between the return on the broad market and the equity return on the firm being analyzed. This model allows volatilities and correlations to evolve and the process is used to simulate the probability of severely negative outcomes over an extended period. Assuming that firms can only operate if capital is a non-trivial proportion of their total liabilities, we define SRISK as the capital that would be needed to achieve a market cap that is 8% of the book value of assets in the event of another crisis. The input to SRISK is size, leverage and risk. Each is important and a firm that wants to reduce its SRISK, can operate on any of

these characteristics.   A crisis is taken to be a 40% drop in global equity values over six months.

SRISK is computed for US firms weekly and published on the internet at

http://vlab.stern.nyu.edu .

In extending this analysis to international markets and to multi-factor systemic risk assessments, the model is naturally generalized to a multivariate volatility model. The data are considered to be the equity return on one firm, the equity return on a global index of equity returns, and perhaps other risk factors such as the equity return a regional index, mortgage returns, sovereign cds spreads, funding illiquidity measures, or other factors. In addition, the use of daily data for assets that are priced in different time zones means that closing prices are not measured at the same time and consequently there may be some important effect from lagged factor returns.

To adapt the DCB model to this setting requires first adjusting for the timing of returns. In the figure below, closing prices in New York and in a foreign market are plotted on a time line. It is clear that the foreign market will close before New York on day t-1 and again on day t. Consequently the company return will correlate with NY returns on the same day and also on the day before. Thus it appears that there is a lagged effect of NY returns on foreign returns, but this is merely a consequence of non-synchronous trading.

By using an ETF traded in NY which is index to a global equity portfolio the model can be expressed simply as

$$R_{i,t} - R_t^f = a + b R_{m,t} + g R_{m,t-1} + e_t^i \qquad (36)$$

To estimate this model we wish to allow the betas to vary over time and the DCB is a natural model. Because of the lag however, it is important to reformulate equation (1) by allowing an additional lag. Otherwise the last dependent variable will be in the information set.
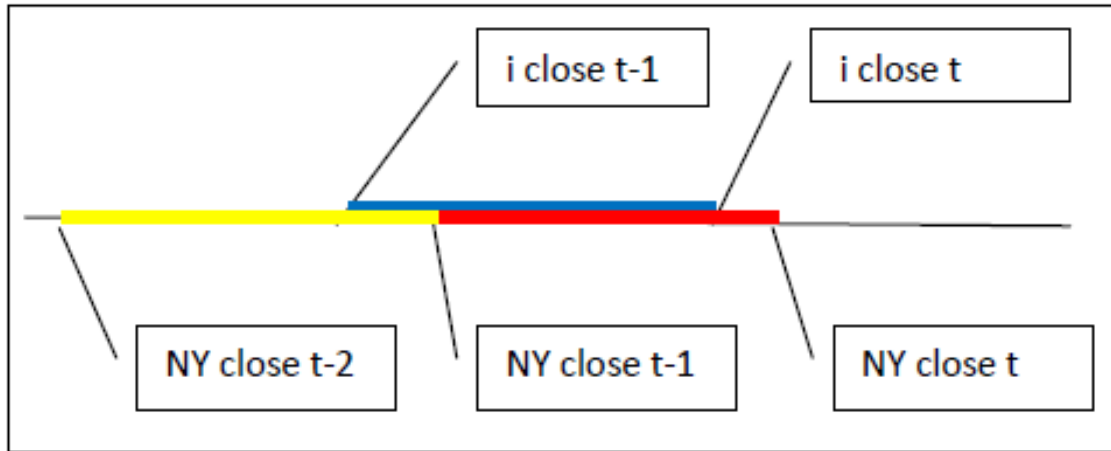
Figure 1

Assuming the means are zero for ease of notation

$$\begin{pmatrix} R_{i,t} \\ R_{m,t} \\ R_{m,t-1} \end{pmatrix} \Big| F_{t-2} \sim N(0, H_t) \tag{37}$$

Thus the DCB equation becomes

$$R_{i,t} = \beta_{i,t} R_{m,t} + \gamma_{i,t} R_{m,t-1} + u_{i,t} \tag{38}$$

From this expression it is clear that $E_{t-2}(u_t) = 0$ but this means that u potentially has a first order moving average term. In fact, this must be the case if the returns of asset i measured in its own country time is to be a Martingale Difference. From (38) $E(R_{i,t}R_{i,t-1})$ will be non-zero unless u has an MA(1) representation with a negative coefficient just sufficient to offset the autocorrelation induced by the non-synchronous data. As in almost all asset return equations, there is heteroskedasticity so u must have a GARCH –MA(1) representation.

23

The Nested DCB model is a natural approach to specifying this equation. It has proven useful in the previous example and will be used here.

$$R_{i,t} = \left(\phi_1 \beta_{i,t} + \phi_2\right) R_{m,t} + \left(\phi_3 \gamma_{i,t} + \phi_4\right) R_{m,t-1} + u_t \tag{39}$$

For each lag, there is both a fixed and a time varying component.

If global equity returns are serially independent, then an even easier expression is available since the covariance between these two factors is zero. The expression for these coefficients can be written as

$$\begin{pmatrix} \beta_{i,t} \\ \gamma_{i,t} \end{pmatrix} = \begin{pmatrix} H_{R_{m,t},R_{m,t}} & H_{R_{m,t},R_{m,t-1}} \\ H_{R_{m,t},R_{m,t-1}} & H_{R_{m,t-1},R_{m,t-1}} \end{pmatrix}^{-1} \begin{pmatrix} H_{R_{m,t},R_{i,t}} \\ H_{R_{m,t-1}R_{i,t}} \end{pmatrix} \tag{40}$$

which simplifies to

$$\beta_{i,t} = \frac{E\left(R_{i,t}R_{m,t} \mid F_{t-2}\right)}{E\left(R_{m,t}^2 \mid F_{t-2}\right)}, \quad \gamma_{i,t} = \frac{E\left(R_{i,t}R_{m,t-1} \mid F_{t-2}\right)}{E\left(R_{m,t-1}^2 \mid F_{t-2}\right)} \tag{41}$$

The expected return of firm i when the market is in decline is called the Marginal Expected Shortfall or MES and is defined as

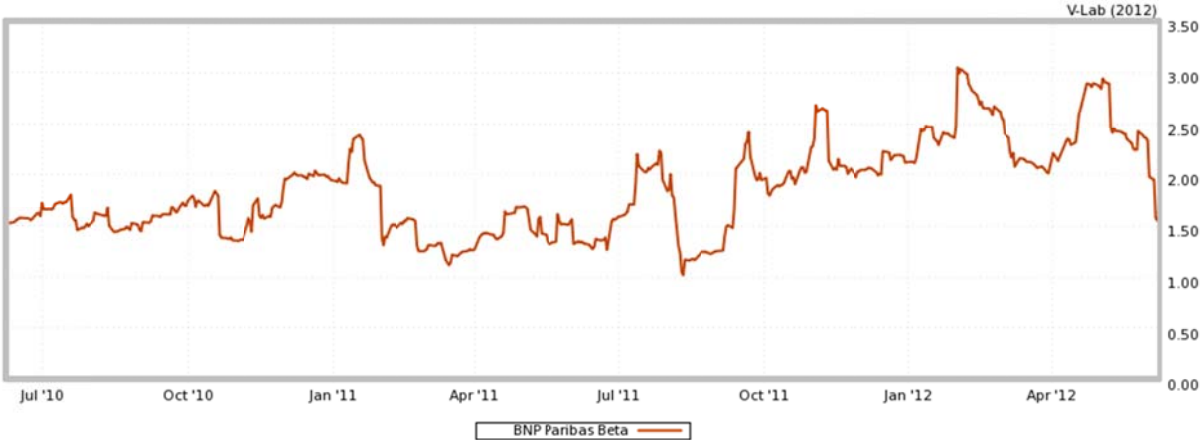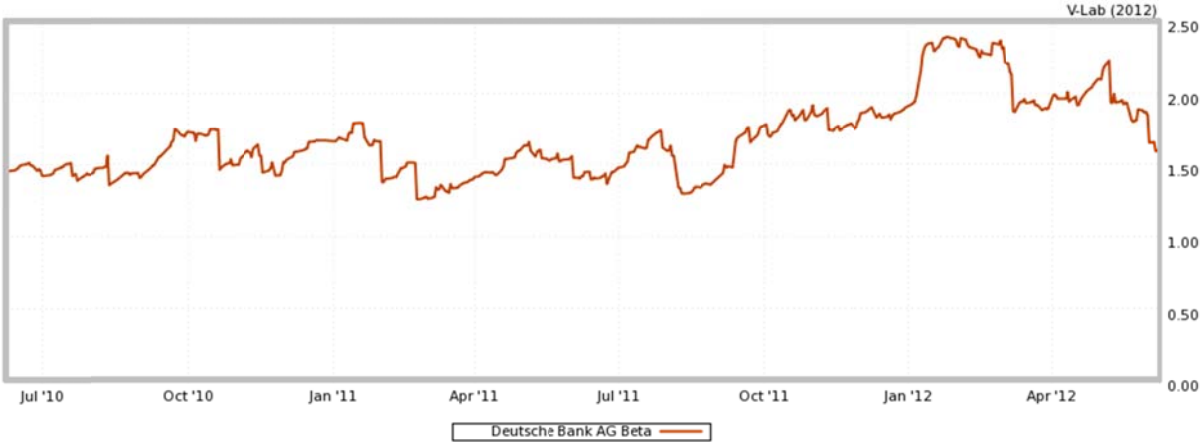$$MES_{i,t} = -E_{t-1}\left(R_{i,t} \mid R_{m,t} < c\right) \tag{42}$$

In the asynchronous trading context with DCB, the natural generalization is to consider the loss on two days after the global market declines on one day. The answer is approximately the sum of beta and gamma from equation (38) times the Expected Shortfall of the global market.
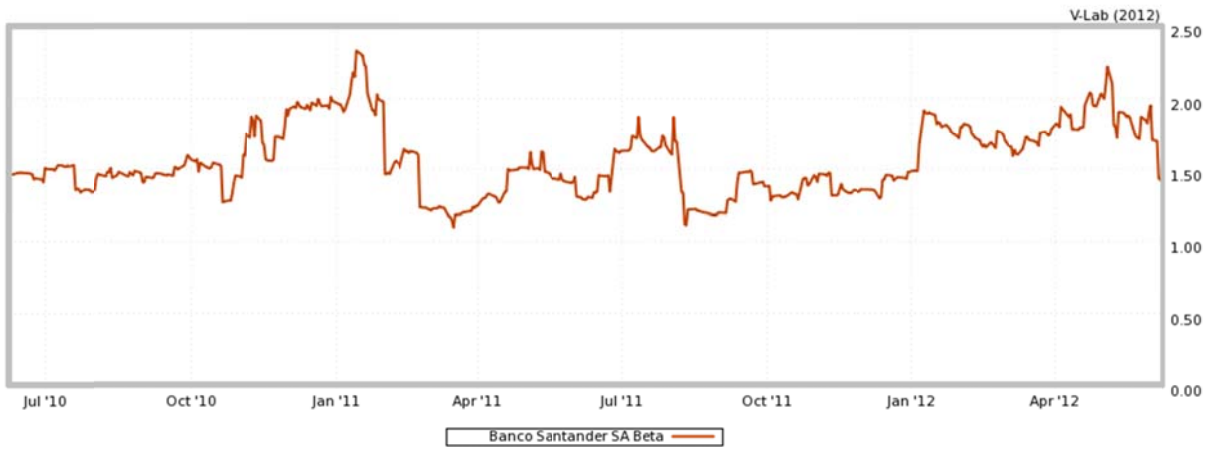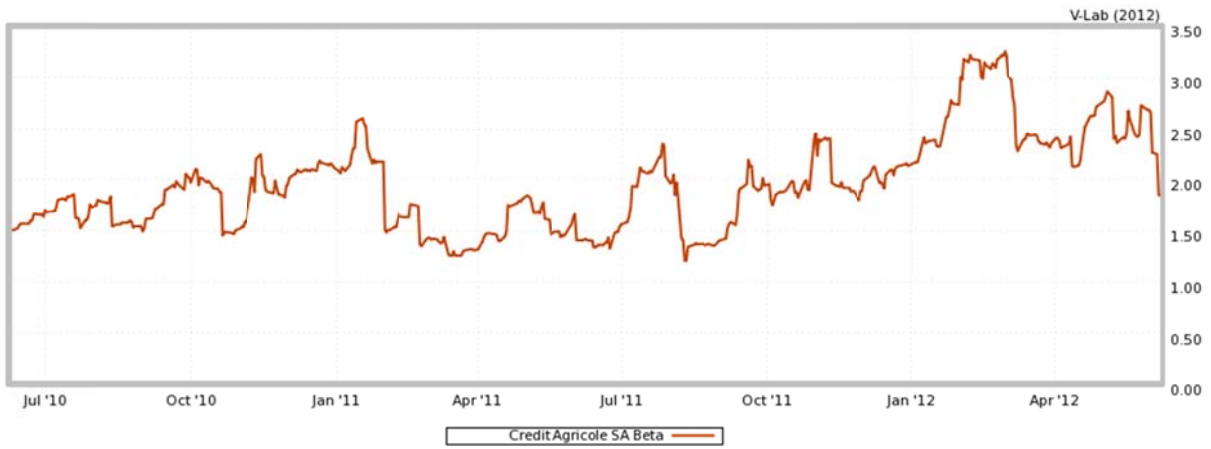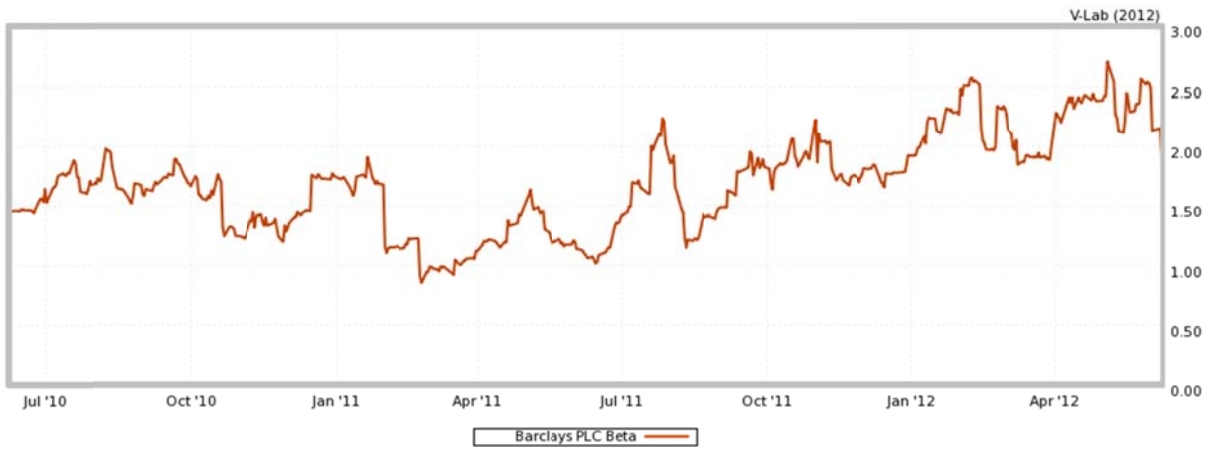
$$\begin{aligned} MES &= -E_{t-2}\left(R_{i,t} + R_{i,t+1} \mid R_{m,t} < c\right) \\ &= -E_{t-2}\left(\beta_{i,t+1}R_{m,t+1} + \gamma_{i,t+1}R_{m,t} + \beta_{i,t}R_{m,t} + \gamma_{i,t}R_{m,t-1} \mid R_{m,t} < c\right) \\ &= -E_{t-2}\left(\left(\gamma_{i,t+1} + \beta_{i,t}\right)E_{t-2}\left(R_{m,t} \mid R_{m,t} < c\right)\right) \\ &\approx -\left(\gamma_{i,t} + \beta_{i,t}\right)E_{t-1}\left(R_{m,t} \mid R_{m,t} < c\right) \end{aligned} \tag{43}$$

The last approximation is based on the near unit root property of all these coefficients when examined on a daily basis.

This statistical model is now operating in V-LAB where it computes SRISK for 1200 global financial institutions weekly and ranks them. A few plots of the betas of European banks will serve to illustrate the effectiveness of the DCB model. In each case the beta is the sum of the current and the lagged beta.

For example, the beta of Deutsche Bank was about 1.50 in the summer of 2011 but it rose to almost 2.50 as the sovereign financial crisis became the dominant news in global equity markets. Since April this risk has decreased.

Deutsche Bank AG Beta

BNP Paribas Beta

Barclays PLC Beta

Credit Agricole SA Beta

Banco Santander SA Beta

In many respects these show the same patterns. Rising betas starting in August 2011 and declining betas over the last few weeks in May and June. The decline in beta will reduce the SRISK but in many cases this effect is counteracted by declining equity values. For example, the 10 bank on the European list is Santander and its SRISK is given by

Banco Santander SA SRISK

Similarly, number 20 is BBVA with SRISK

Banco Bilbao Vizcaya Argentari SRISK

These banks are rising in the chart as others are dropping.

IX Conclusion

This paper has introduced a new method to estimate time series regressions that allow for time variation in the regression coefficients called Dynamic Conditional Beta or DCB. The

approach makes clear that the regression coefficient should be based on the predicted

covariances of endogenous and exogenous variables as well as potentially the predicted means.

Estimation of the model is discussed in a likelihood context with some betas varying and others

constant. Testing the constancy of beta is a non-nested hypothesis and several approaches are

suggested and implemented. The most attractive appears to be artificial nesting which motivates

the NESTED DCB model where both constant and time varying coefficients are introduced.

     The model is applied in two contexts, multi-factor asset pricing and systemic risk

assessment. The Fama French three factor model is applied to industry portfolios and the fixed

beta assumption is tested. There is strong evidence that the betas are time varying. From plots it

is apparent that the hml beta is the most volatile. In the global systemic risk context a dynamic

two factor model is estimated which allows foreign equities to respond to current as well as

lagged global prices which is expected because of non-synchronous markets. As a result, betas

for global financial institutions can be followed over time. The big European banks have had

rising betas over the fall of 2011 as the sovereign debt crisis grows in strength. This is an input

to the NYU Stern Systemic Risk Ranking displayed on V-LAB which documents the serious

nature of the current level of systemic risk.

REFERENCES

Acharya, V. V., Pedersen, L., Philippon, T., and Richardson, M., 2010, "Measuring Systemic Risk", Technical report, Department of Finance, NYU Stern School of Business.

Acharya, V. V., Pedersen, L. H., Philippon, T, and Richardson, M. P., 2010, "Measuring Systemic Risk", in *Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance*, Acharya, V. V., Cooley, T., Richardson, M, and Walter, I. (Eds), John Wiley & Sons.

Acharya V., R. Engle and M. Richardson, (2012) "Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks" *forthcoming Papers and Proceedings of American Economic Review*

Atkinson, A. 1970. A method for discriminating between models. Journal of the Royal Statistical Society B 32, 323–53.

Bali, Turan G.; Engle, Robert F., (2010), "The intertemporal capital asset pricing model with dynamic conditional correlations"*, Journal of Monetary Economics*, May, Vol. 57 Issue 4, p377-390, 14p;

Bali, Turan, Robert Engle, and Yi Tang,(2012) "Dynamic Conditional Beta is Alive and Well in the Cross-Section of Daily Stock Returns", manuscript, Georgetown

Bollerslev, T., R. Engle and J. Wooldridge(1988), "A Capital Asset Pricing Model with Time-Varying Covariances", *Journal of Political Economy,* Volume 96, Issue 1, (Feb), pp116-131

Bollerslev, Tim and Jeffrey Wooldridge(1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances", *Econometric Reviews,* Volume 11,No.2, pp 143-172

Brownlees, C and R. Engle(2011) "Volatility, Correlation and Tails for Systemic Risk Measurement", NYU Department of Finance Working Paper

Burns, Patrick, Robert Engle and Joseph Mezrich, (1998) "Correlations and Volatilities of Asynchronous Data", *Journal of Derivaties,* Volume 5, Issue 4, pp7-18

Cox, D. R. (1961). Tests of separate families of hypothesis. Proceedings of the Forth Berkeley Symposium on Mathematical Statistics and Probability 1, 105–23.

Cox, D. R. (1962). Further results on tests of separate families of hypothesis. Journal of the Royal Statistical Society B 24, 406–24.

Engle, Robert F. (2002) "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business and Economic Statistics,* Volume 20, pp339-350

Engle, Robert (2009) *Anticipating Correlations: A New Paradigm for Risk Management,* Princeton University Press

Engle, Robert, David Hendry and J.F. Richard(1983), "Exogeneity", *Econometrica,* Volume 51, Number 2, pp277-304

Fama, Eugene and Kenneth French,(2004) "The Capital Asset Pricing Model: Theory and Evidence", *Journal of Economic Perspectives,* Volume 18, Number 3, pp25-46

Ken French web site:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Pagan, Adrian, (1984) "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review,* Vol 25, No1, pp 221-247

Rivers, D. and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. Econometrics Journal 5, 1–39.

Vuong, Quang,(1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica,* Vol 57, No.2 March, pp 307-333

Wooldridge, Jeffry,(2002) *Econometric Analysis of Cross Section and Panel Data",* MIT Press