DYNAMIC CONDITIONAL BETA

Robert Engle[1]

February 27, 2014

ABSTRACT

Dynamic Conditional Beta (DCB) is an approach to estimating regressions with time varying parameters. The conditional covariance matrices of the exogenous and dependent variable for each time period are used to formulate the dynamic beta. Joint estimation of the covariance matrices and other regression parameters is developed. Tests of the hypothesis that betas are constant are non-nested tests and several approaches are developed including a novel nested model. The methodology is applied to industry multifactor asset pricing and to global systemic risk estimation with non-synchronous prices.

Keywords: GARCH, DCC, Time Varying Parameters, Multivariate GARCH, Non-Nested Tests, Multi-factor Asset Pricing, Systemic Risk, SRISK

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

[1] Robert Engle is Director of the Volatility Institute of NYU's Stern School of Business and Professor of Finance. He appreciates comments and suggestions from many people on this paper, particularly participants at Oxford's Launch of EMoD, SoFiE Singapore, and the University of Chicago Booth Seminar. Particular thanks go to Viral Acharya, Turan Bali, Niel Ericsson, Gene Fama, Eric Ghysels, Jim Hamilton, David Hendry, Alain Monfort, Matt Richardson, Jeff Wooldridge and to Rob Capellini, Hahn Le, and Emil Siriwadane.

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 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, $\left(y_t, \mathrm{x}_t'\right)', t = 1,...T$ where $x_t' = \left(x_{1,t},...x_{k,t}\right)$. The objective is to characterize the conditional distribution of y given x. Although we will think of y_t as a scalar time series, most of the formulae apply directly if it is also a vector. In a time series context this requires finding $f_{t-1}\left(y_t \middle| x_t\right)$ which is the conditional distribution of y given x and the past of y and x. The regression is the mean of this distribution denoted $E_{t-1}\left(y_t \middle| x_t\right)$. 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 or of course may be a general non-linear function of x and the past.

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} \Big| \mathbb{F}_{t-1} \sim N\left(\mu_t, H_t\right) \tag{1}$$

where $\left(\mu_t, H_t\right)$ are measurable with respect to $\mathbb{F}_{t-1}$, the information set adapted to the past of $\left(y_t, x_t\right)$. Then the desired conditional distribution is simply

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

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.

This expression for DCB 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 can be generalized. For example, if the joint distribution in (1) is in the elliptical family of distributions, then (2) holds with the appropriate elliptical distribution. Even without elliptical distributions, (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.

It is important to recognize that equation (3) is a natural extension of estimators that have long been used in time series. Probably the first application was Bollerslev, Engle and Wooldridge(1988) where the beta was computed as the ratio of the conditional covariance to the conditional variance. One of the leading applications of multivariate GARCH models has always been estimation of hedge ratios. This is discussed in surveys such as Bollerslev, Chou and Kroner(1994), Bollerslev, Engle and Nelson(2004), Bera and Higgins(1993) and Lien and Tse(2002). Specific hedging models are proposed by Kroner and Sultan(1993), Cecchetti, Cumby and Figlewski(1998), Bera, Garcia and Roh(1997) and Braun, Nelson and Sunier(1995).

More recently, hedging performance is used as a test of the multivariate garch specification in Engle(9), and Lien, Tse and Tsui(2002) among others. Recent applications to asset pricing and systemic risk are in Bali, Engle and Tang(2012) and Brownlees and Engle(2012). When data are available at a higher frequency than the analysis, estimates of realized variances, realized covariances and realized betas can be constructed. These realized betas are time varying and can be directly forecast and modeled. See for example Andersen, Bollerslev, Diebold and Wu(2005)(2006) and Patten and Verardo(2012).

All of these examples would be considered as Dynamic Conditional Betas by the description above. Almost all of these applications involve only bivariate models where there is just one covariance to be estimated and the hedge ratio is simply the ratio of the conditional covariance to the conditional variance. In no case is there a statistical test of the null hypothesis that the beta is truly constant. There is no discussion of maximum likelihood estimation in these more general contexts.

This paper will extend the analysis of dynamic conditional betas into a multivariate context and then explicitly consider the estimation and testing of the model from a classical perspective. The tools developed here would have been useful for many of these prior studies.

III. Maximum Likelihood Estimation of DCB

A. All betas are varying

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} \Big| \mathbb{F}_{t-1} \sim N\left(\mu_t\left(\theta_0\right), H_t\left(\theta_0\right)\right), \quad \underset{(jx1)}{\theta_0} \in \Theta \tag{4}$$

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 θ can be computed for the data vector $z_t = \left(y_t, x_t '\right)'$. Letting $\ddot{z} \equiv \left\{z_1, ..., z_T\right\}$,

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, L\left(\theta;\ddot{z}\right), \quad L\left(\theta;\ddot{z}\right) = -\frac{1}{2}\sum_{t=1}^{T} \log\left|H\left(\theta\right)_t\right| + \left(z_t - \mu\left(\theta\right)_t\right)' H\left(\theta\right)_t^{-1}\left(z_t - \mu\left(\theta\right)_t\right) \tag{5}$$

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

$$\hat{\beta}_t = H\left(\hat{\theta}\right)_{xx,t}^{-1} H\left(\hat{\theta}\right)_{xy,t} \tag{6}$$

The limiting distribution of such estimators can be established under certain conditions. In Bollerslev and Wooldridge(1992), Theorem 2.1 establishes conditions for the limiting distribution of $\hat{\theta}$, to be

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to N\left(0, A^{-1}\right) \tag{7}$$

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. This theorem requires correct specification of the first two moments at the true parameter vector, and a variety of regularity conditions. The theorem requires $\theta_0 \in$ interior Θ and identifiably unique. $\left(\mu, H\right)$ must be twice continuous differentiability of for all possible values of z, the log likelihood must satisfy a uniform weak law of large numbers, the score vector must satisfy a central limit theorem and the Hessian must satisfy a law of large numbers and converge to a non-singular matrix. Some of these conditions can be weakened , particularly in the context of stationary ergodic random variables.

Define $\beta_{0,t}$ as the true, conditional beta at time t from (6) when $\hat{\theta} = \theta_0$. Applying the mean value theorem to (6) at time t,

$$\hat{\beta}_t - \beta_{0,t} = G_t\left(\overline{\theta}_t\right)\left(\hat{\theta} - \theta_0\right),\ G_t\left(\overline{\theta}_t\right) = \left.\frac{\partial\left[H\left(\theta\right)_{xx,t}^{-1} H\left(\theta\right)_{xy,t}\right]}{\partial\theta'}\right|_{\theta=\overline{\theta}_t} \tag{8}$$

where $\overline{\theta}_t \in \left(\theta_0, \hat{\theta}\right)$ with possibly different values for different rows and different t. Then

$$\sqrt{T}\left(\hat{\beta}_t - \beta_{0,t}\right) \xrightarrow[T\to\infty]{D} N\left(0, G_t A^{-1} G_t'\right) \tag{9}$$

and G_t can be estimated as $G_t\left(\hat{\theta}\right)$. 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 θ which then imply time varying standard errors for the betas.

A useful expression for the log likelihood can be found by changing variables. Convert the likelihood from $\left(y_t, x_t'\right)$ *to* $\left(w_t, x_t'\right)$ *where* $w_t = y_t - \mu_{y,t} - H_{yx,t} H_{xx,t}^{-1}\left(x_t - \mu_{x,t}\right)$. Here *w* can be interpreted as the residuals from a DCB regression and $h_{w,t}$ is the conditional variance of the residuals. As the Jacobian of the transformation is 1, the likelihood simply has a recalculated covariance matrix which is now diagonal giving:

$$L\left(\theta; \ddot{y}, \ddot{x}\right) = L\left(\theta; \ddot{w}, \ddot{x}\right) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log(h_{w,t}) + \frac{w_t^2}{h_{w,t}}\right) - \left\{\frac{1}{2}\sum_{t=1}^{T}\log\left|H\left(\theta\right)_{xx,t}\right| + x_t' H\left(\theta\right)_{xx,t}^{-1} x_t\right\} \tag{10}$$

B. 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:

$$\begin{aligned} L(\theta;\ddot{y},\ddot{x}) &= \sum_{t=1}^{T}\left[\log f_{y|x}\left(y_t \middle| x_t, z_{t-1},...;\theta\right) + \log f_x\left(x_t \middle| z_{t-1},...;\theta\right)\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(h_{u,t}\right) + \frac{\left(y_t - x_t{}'\beta\right)^2}{h_{u,t}}\right) - \left\{\frac{1}{2}\sum_{t=1}^{T}\left[\log\left|H(\theta)_{xx,t}\right| + x_t{}'H(\theta)_{xx,t}^{-1}x_t\right]\right\} \end{aligned} \tag{11}$$

In this equation

$$u_t = y_t - x_t{}'\beta, \;\; V_{t-1}\left(u_t\right) = h_{u,t} \tag{12}$$

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 as in Engle(1982) and more generally Bollerslev and Wooldridge(1992).

C. Some betas are constant and some are varying

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 (6). Partitioning $x=(x_1',x_2')'$ and then defining $x_1^0=(x_1',0)'$, the model can be expressed as

$$y_t = H_{yx,t}H_{xx,t}^{-1}x_{1,t}^0 + \beta_2' x_{2,t} + v_t \tag{13}$$

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

$$L(\theta;\ddot{y},\ddot{x}) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log\left(h_{v,t}\right)+\frac{u_t^2}{h_{v,t}}\right]-\left\{\frac{1}{2}\sum_{t=1}^{T}\left[\log\left|H(\theta)_{xx,t}\right|+x_t' H(\theta)_{xx,t}^{-1} x_t\right]\right\} \tag{14}$$

Notice that the second term in curly brackets has the same form in (10) (11) and (14). For many specifications of the volatility process, the estimates of the covariance matrix of *x* will be identical for these three specifications, so any comparison of log likelihoods can be done simply by comparing the marginal log likelihoods. The test becomes a test of which residual vector, *(w,u,v)* has the highest likelihood. Interestingly, the regression models with time varying parameters do not necessarily have the best fit. Remember, the coefficients are the predicted coefficients for each data point.

IV. Testing for constant regression coefficients

In this section a set of approaches will be developed for testing that one or more parameters are constant. These hypotheses are basically non-nested although they more precisely are overlapping hypotheses. The overlapping occurs because if there is no conditional variance or covariance in the data set, the null and alternative hypotheses become identical. Thus there are points in the parameter space where the hypotheses are not testable.

There are several standard approaches to testing non-nested hypotheses. The first developed initially by Cox(1961,1962) is a model selection approach which simply asks which model has the highest value of a (possibly 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 one variable case the natural artificial nesting consists of estimating the following equation:

$$y_t = x_t\beta + x_t\beta_t\gamma + \sqrt{h_{u,t}}u_t \tag{15}$$

The explanatory variable x enters twice – once by itself and once multiplied by the predicted beta. 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.

V. Overlapping Hypotheses and Econometric Solutions

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

$$z_t = \left(y_t', x_t'\right)', \;\; t = 1,...,T \tag{16}$$

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:

$$\begin{aligned} z_t \left| \mathbb{F}_{t-1} \right. &\sim f_t\left(z_t, \theta\right), \;\; \theta \in \Theta \\ z_t \left| \mathbb{F}_{t-1} \right. &\sim g_t\left(z_t, \phi\right), \;\; \phi \in \Phi \end{aligned} \tag{17}$$

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.

$$\begin{aligned} f_t\left(z_t;\theta\right) &= f_{y|x,t}\left(y_t \left| x_t;\theta\right.\right) f_{x,t}\left(x_t;\theta\right) \\ g_t\left(z_t;\phi\right) &= g_{y|x,t}\left(y_t \left| x_t;\phi\right.\right) g_{x,t}\left(x_t;\phi\right) \end{aligned} \tag{18}$$

The maximized log likelihoods of these models are given by

$$\begin{aligned} L_f\left(\hat{\theta}\right) &= \sum_{t=1}^{T} \log f_{y|x,t}\left(y_t \left| x_t;\hat{\theta}\right.\right) + \log f_{x,t}\left(x_t;\hat{\theta}\right) \\ L_g\left(\hat{\phi}\right) &= \sum_{t=1}^{T} \log g_{y|x,t}\left(y_t \left| x_t;\hat{\phi}\right.\right) + \log g_{x,t}\left(x_t;\hat{\phi}\right) \end{aligned} \tag{19}$$

Model f is said to be nested within g if for every $\theta_0 \in \Theta$ there is a $\phi^* \in \Phi$ such that $f\left(z,\theta_o\right) = g\left(z,\phi^*\right)$ so that events have the same probability whether $f\left(z,\theta_o\right)$ or $g\left(z,\phi^*\right)$ 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 $\left(\Theta \otimes \Phi\right)$ 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 conditional 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 (19) and a penalty function such as Schwarz or Aikaike, the selected model can be directly found. The criterion function is the penalized likelihood ratio given by

$$LR_T = \left\{ \sum_{t=1}^{T} \log\left(f_{z,t}\left(z_t;\hat{\theta}\right)\right)\right\} - \left\{ \sum_{t=1}^{T} \log\left(g_{z,t}\left(z_t;\hat{\gamma}\right)\right)\right\} - \left(pen_f - pen_g \right) \tag{20}$$

A positive value of LR selects *f* and a negative value selects g. For the Aikaike Information Criterion (AIC), $\left(pen_f - pen_g \right) = \left(k_f - k_g \right)$ and for the Schwarz or Bayesian Information Criterion, (BIC) it is given by $\left(k_f - k_g \right) * \log\left(T \right) / 2$ Here k is the total number of estimated parameters in each model. Clearly, as T becomes large, the penalty becomes irrelevant in model selection.

In many cases the likelihood will have the same estimate of the likelihood of the conditioning variables, *x*, and consequently the maximized values will satisfy

$$\left\{\sum_{t=1}^{T}\log\left(f_{x,t}\left(x_t;\hat{\theta}\right)\right)\right\}=\left\{\sum_{t=1}^{T}\log\left(g_{x,t}\left(x_t;\hat{\gamma}\right)\right)\right\} \tag{21}$$

This will happen if the dynamic covariance matrix of the x variables is estimated the same way regardless of the specification of the dynamics of the regression coefficients. 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

$$LR_T=\sum_{t=1}^{T}\log\left(f_{y|x,t}\left(y_t;\hat{\theta}\,|x_t\right)\right)-\sum_{t=1}^{T}\log\left(g_{y|x,t}\left(y_t;\hat{\gamma}\,|x_t\right)\right)-\left(pen_f-pen_g\right) \tag{22}$$

which can be expressed in terms of the residuals as in equations (10), (11) and (14). Model selection between DCB and constant betas can be based on the penalized log likelihood

$$LR_T=\sum_{t=1}^{T}\left(\log\left(h_{w,t}\right)-\frac{w_t^2}{h_{w,t}}\right)-\sum_{t=1}^{T}\left(\log\left(h_{u,t}\right)-\frac{u_t^2}{h_{u,t}}\right)-\left(pen_w-pen_u\right) \tag{23}$$

The penalties will depend upon the number of parameters estimated in each model. Since the parameters in the x matrix are the same and the parameters in the error process are the same, the only difference is the number of fixed betas that are estimated.

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 (22) 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

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

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 (11) and (14) we get

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

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

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

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 (15), 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, which leads to the generated regressor problem of Pagan(1984). A theorem from Wooldridge(2002) however shows that parameter estimates are consistent and some inference is supported.

The nested DCB model can be expressed as

$$y_t = x_t\beta + x_t\beta_t\left(\theta_0\right)\gamma + \varepsilon_t = \left(x_t, x_t\beta_t\left(\theta_0\right)\right)\phi + \varepsilon_t,\ \phi = \left(\beta,\gamma\right)' \tag{29}$$

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 at $\phi = \phi_0$, requires the condition that

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

Under the standard likelihood assumptions and under the condition (7) this term in the Hessian can be expressed as

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

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

$$m_t = a_t - b_t' f_t \tag{32}$$

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:

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

The last expression is based upon the assumption that the factors are priced with the same pricing kernel. Notice that in this derivation

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

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, Engle and Tang(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 a pooled time series cross section 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:

$$r_t^j - r_t^f = \alpha_j + \beta_{j,m}\left(r_t^m - r_t^f\right) + \beta_{j,hml} r_t^{hml} + \beta_{j,smb} r_t^{smb} + \sqrt{h_t^j}\,\varepsilon_t^j \qquad (35)$$

The DCC model is used to estimate the correlations between all variables. For each equation there are 4 variables, the three factors and an endogenous variable. Because the factors are the same for each equation, it is natural to use the same covariance matrix for them. This is accomplished by estimating a 3x3 DCC once and using it for all equations. If a 4x4 DCC specification is implemented for each sector, then the DCC parameters from the factors can be applied to the endogenous variables directly. This is in the spirit of the Composite Likelihood approach to estimating vast covariance matrices of Engle et al (2014). .

The model is therefore estimated under three assumptions:

a) OLS with constant coefficients and robust standard errors

b) GJR-GARCH with constant coefficients

c) DCB with regression coefficients from (3). Correlation matrices estimated with trivariate DCC for the factors. Bivariate DCC is used for the correlations between the dependent variable and each of the factors using the parameters estimated from the trivariate factor correlation matrix. The GJR-GARCH model is used for each series.[2]

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, all twelve industries are better fit with time varying betas. The Vuong test examines whether these differences are significant. These are t-ratios and consequently at conventional levels, all twelve industries show significant improvement using purely time varying 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.

[2] Notice that this makes the factor covariance matrix the same for each industry and insures that the joint correlation matrix of all the endogenous and exogenous variables is positive definite. When the correlations with the endogenous variables were estimated individually without constraints, several had persistence exceeding one. The Schwarz criterion uniformly preferred the model with restricted covariance to the model with unrestricted covariances.

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

NAME	SCHWARZ_DCB	SCHWARZ_GARCH	VUONGTEST
Buseq	1.597089	1.706647	9.726441
Chems	1.097934	1.131116	3.333880
Durbl	1.835669	1.893769	6.559011
Enrgy	1.973194	2.076260	9.118627
Hlth	1.365149	1.433991	6.551288
Manuf	0.291902	0.348881	4.329228
Money	0.716225	0.875159	11.95807
Nodur	0.669520	0.873242	15.94370
Other	0.559703	0.620909	3.983843
Shops	1.113731	1.172615	5.679295
Telcm	1.738182	1.787497	5.564302
Utils	0.894953	1.011304	8.349354

Schwarz information criteria are computed for DCB and for constant beta. These are the negative of the log likelihood so smaller is better. The Vuong test from equation (28) is a t-test where positive values favor DCB.

Table 2

Tests of Nested DCB Model

NAME	FIX1	FIX2	FIX3	DCB1	DCB2	DCB3
Buseq	22.2	-10.0	-3.25	36.0	38.0	30.8
Chems	16.8	-2.43	-9.45	25.7	26.9	9.16
Durbl	19.0	9.61	1.05	22.6	28.8	25.5
Enrgy	18.7	0.55	-7.28	29.0	46.8	16.2
Hlth	14.7	-10.1	-11.0	33.6	34.9	15.7
Manuf	18.5	8.33	10.4	21.9	33.0	19.1
Money	16.0	6.76	10.8	52.6	42.4	32.9
Nodur	13.6	-2.10	-3.09	40.0	35.2	45.4
Other	24.7	3.28	14.5	25.3	38.9	37.1
Shops	18.6	-6.08	2.37	32.3	27.2	17.5
Telcm	15.5	5.15	-10.5	26.2	27.4	12.4
Utils	17.7	7.25	3.00	43.2	40.5	34.8

In the nested model of equation (17) the t-statistics of the factor with constant beta are given first and the t-statistics of the time varying betas are given second. Factor 1 is mkt_rf, Factor 2 is HML and Factor 3 is SMB.

Table 3

Testing that Alpha is Zero in the Three Factor FF Model

NAMEN	TSTAT_DCB	TSTAT_NEST	TSTAT_GARCH	TSTAT_OLS
Buseq	0.54	0.98	1.63	2.46
Chems	-0.67	-0.12	0.76	0.87
Durbl	-2.60	-2.81	-2.70	-2.52
Enrgy	0.06	0.54	2.16	0.85
Hlth	2.81	3.83	5.04	3.42
Manuf	-3.27	-3.06	-1.82	-0.35
Money	-2.69	-2.76	-2.01	-3.10
Nodur	1.19	1.71	4.15	3.74
Other	-1.47	-1.91	-2.60	-2.75
Shops	0.65	1.19	2.08	1.76
Telcm	-0.88	-0.38	0.69	-0.14
Utils	-3.17	-3.23	-2.55	-0.39

T-statistics are presented on the intercept of a factor model estimated with DCB, Nested DCB, constant coefficient regression with GARCH residuals, and constant coefficients with robust standard errors.

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, 4 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. Because there are 12 thousand daily betas, quarterly averages are plotted. From these figures, it is clear that typically the beta on HML is the most variable. Business Equipment has a negative beta on HML in the 90s reaching a beta of -2 and then returning to about -1. This sector therefore trades like a growth stock with time varying intensity. Energy appears to be a value stock from 1982-90 and 2003-2007 but was a growth stock in 1980-82. Money becomes a growth stock in 2004-2007 and then with a dramatic reversal becomes a value stock as its equity values decline during the financial crisis. The market beta also rises for the Money sector during the financial crisis but not nearly as much as it does in a one factor model. The beta on SMB becomes negative after 1990 indicating that the sector moves like a large cap stock. All the nested betas for all sectors are shown in the figures in the appendix.

These estimates are useful for evaluating the style of a fixed asset portfolio since this style will change over time as the betas change. Furthermore, an investor who wants to take a position in a particular style such as growth or value, could use these estimates to rotate industries in and out of his portfolio.

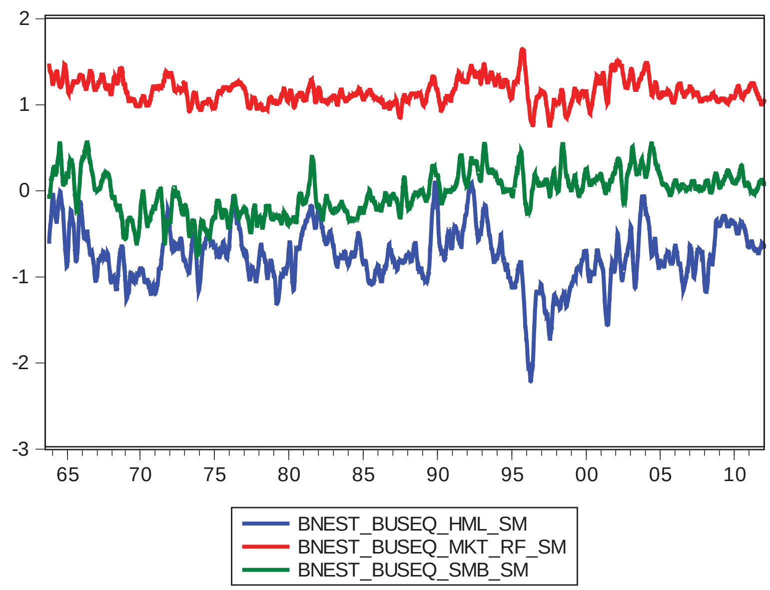

Figure 1

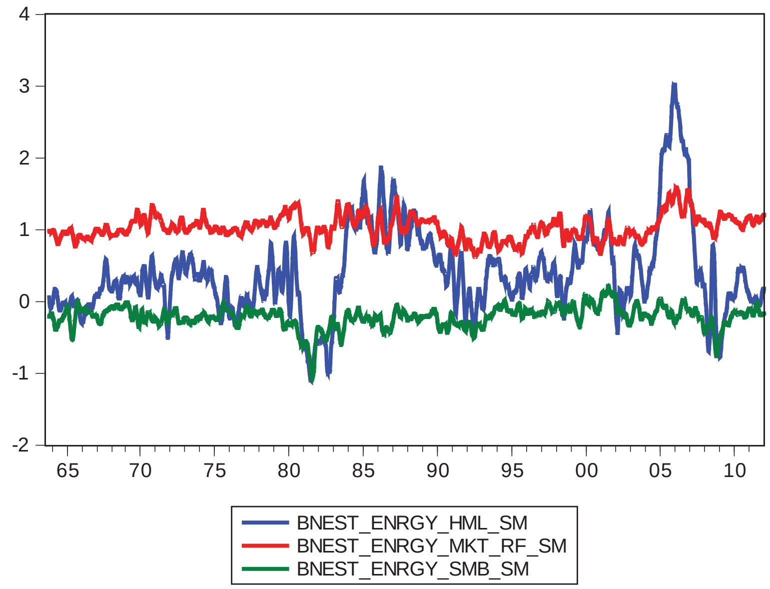

Figure 2

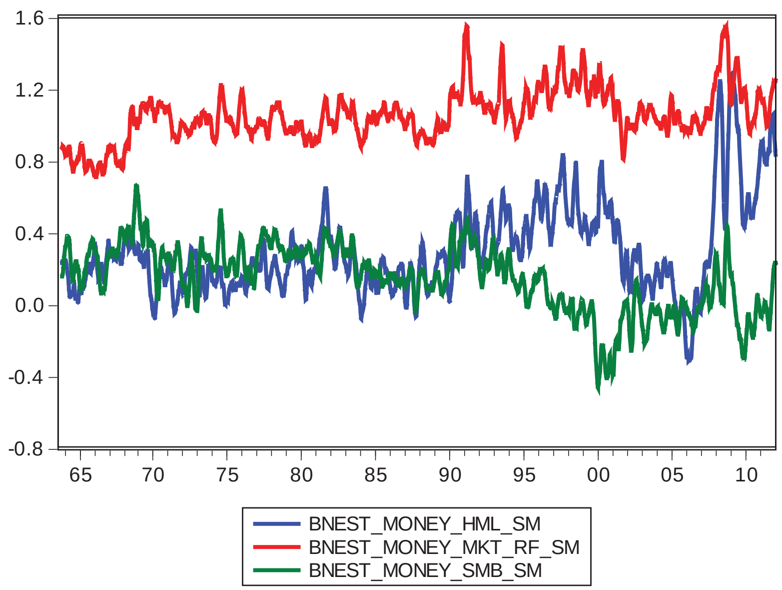

Figure 3

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 weekly for firms from about 70 countries and published on the internet at http://vlab.stern.nyu.edu .

In extending this analysis to international markets, 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 some lags. 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 indexed to a global equity portfolio, the model can be expressed simply as

$$R_{i,t} - R_t^f = \alpha + \beta R_{m,t} + \gamma R_{m,t-1} + e_t^i \tag{36}$$

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 lagged dependent variable will be in the information set.

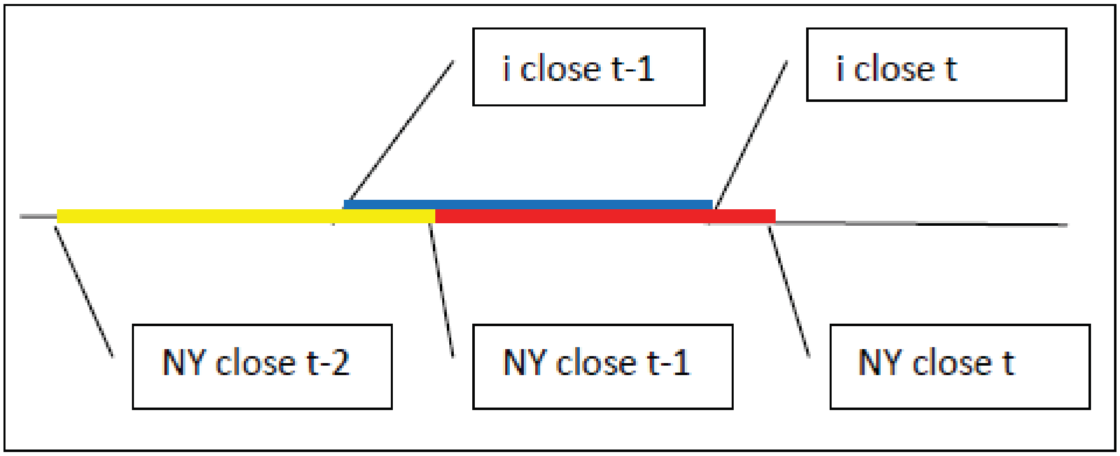

Figure 1

Figure 4.

Assuming the means are zero for ease of notation, the covariance of returns can be expressed as

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

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 $E_{t-2}\left(u_t\right) = 0$ 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) $E\left(R_{i,t} R_{i,t-1}\right)$ 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.

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

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

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

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

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.

$$\begin{aligned} MES &= -E_{t-2}\left(R_{i,t} + R_{i,t+1} \middle| R_{m,t} < c\right) \\ &= -E_{t-2}\left(\beta_{i,t+1} R_{m,t+1} + \gamma_{i,t+1} R_{m,t} + \beta_{i,t} R_{m,t} + \gamma_{i,t} R_{m,t-1} \middle| R_{m,t} < c\right) \\ &= -E_{t-2}\left(\left(\gamma_{i,t+1} + \beta_{i,t}\right) E_{t-2}\left(R_{m,t} \middle| R_{m,t} < c\right)\right) \\ &\simeq -\left(\gamma_{i,t} + \beta_{i,t}\right) E_{t-1}\left(R_{m,t} \middle| R_{m,t} < c\right) \end{aligned} \tag{43}$$

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. Plots of the betas of the four highest SRISK 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.

The beta of Deutsche Bank was about 1.50 in the summer of 2011 but it rose to more than 2.50 as the sovereign financial crisis became the dominant news in global equity markets. Since 2013 this risk has decreased. Credit Agricole reached betas of 3.5 during the sovereign debt crisis but these have now returned to around 1.5. Similar observations apply to BNP Paribas and to Barclays. Interestingly, only Barclays showed a rise in beta during the financial crisis of 2007-2009. The SRISK of all European firms is calculated from these betas and from the leverage and market cap of each firm. Adding these together indicates the systemic risk facing Europe over time. Clearly it has declined substantially as shown in Figure 9, both from the financial crisis of 7-9 and from the sovereign financial crisis that is primarily a European crisis.

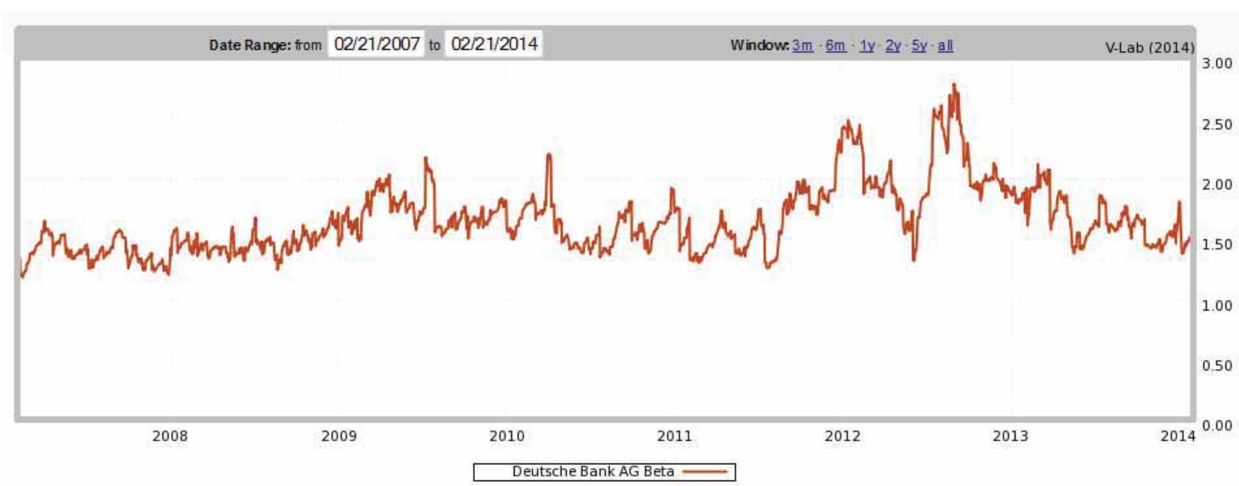

Figure 5

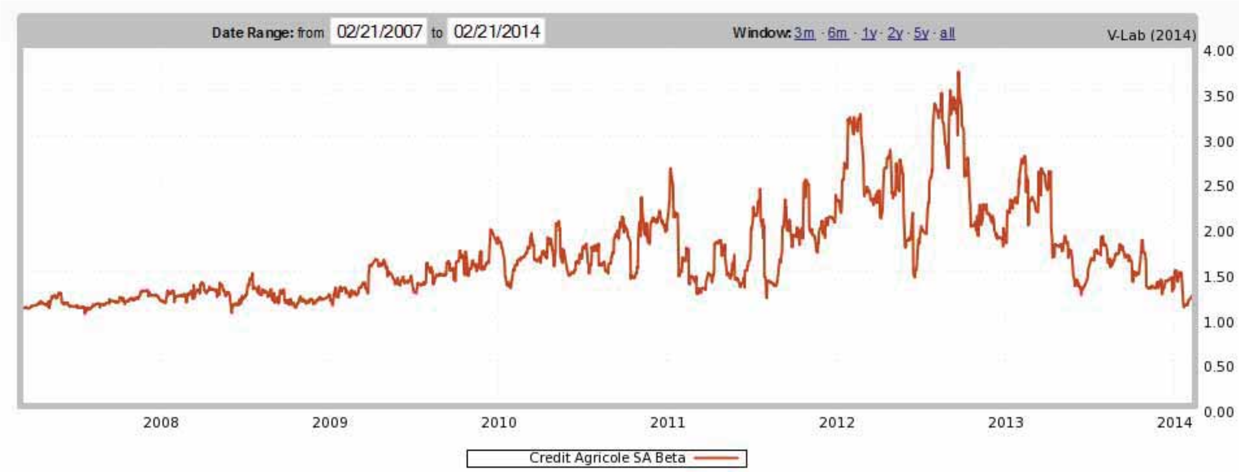

Figure 6

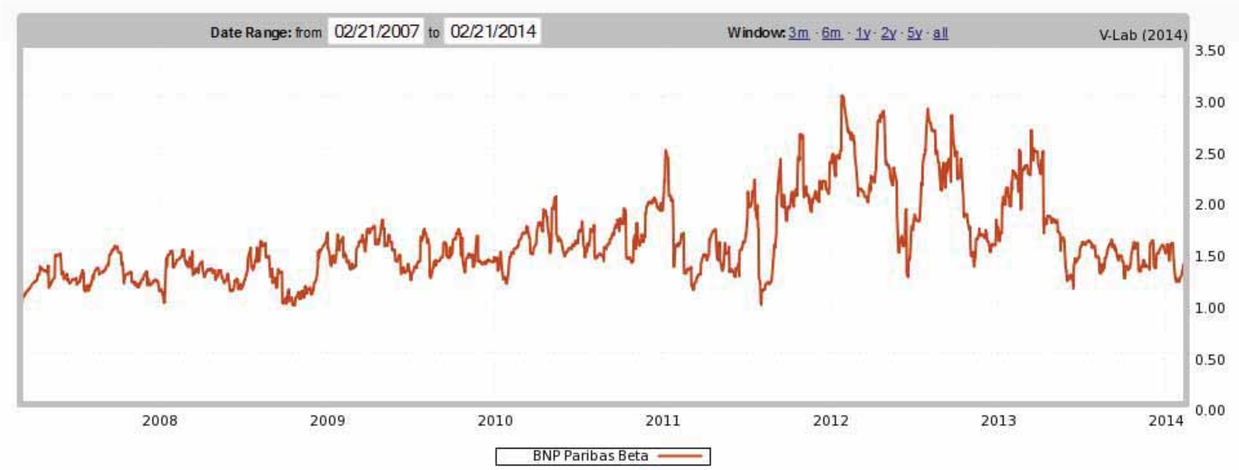

Figure 7

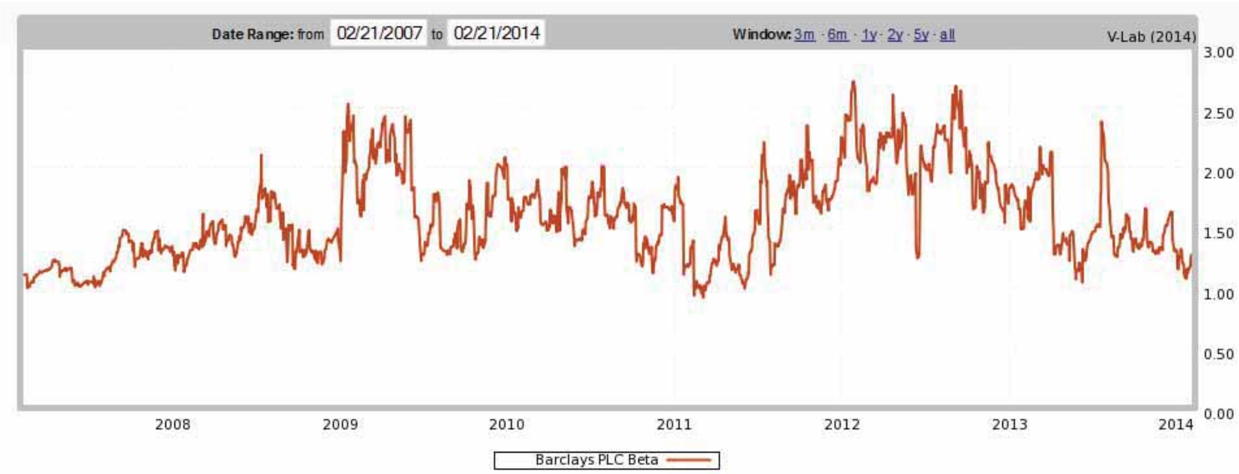

Figure 8

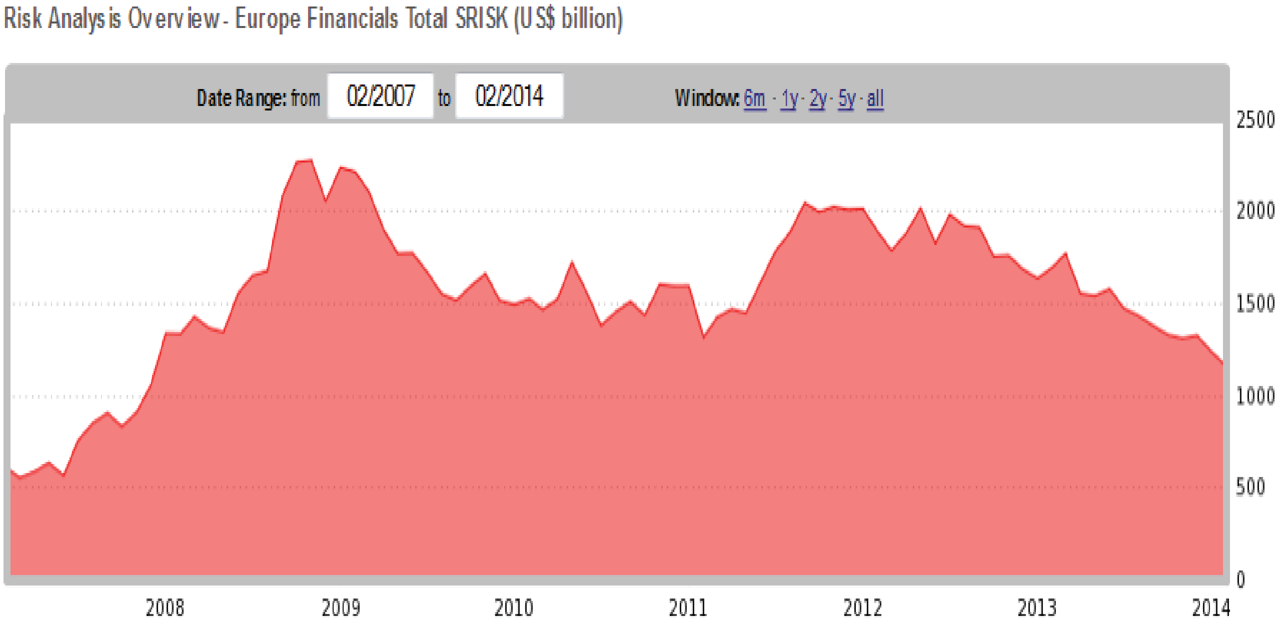

Figure 9

Total SRISK in US $ of European Financials

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 systemic risk in 2012 and its improvement in 2013.

REFERENCES

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

Acharya V.V., R.F. Engle, and M.P. Richardson. 2012. Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks," *Papers and Proceedings of American Economic Review* 102(3): 59-64.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and J. Wu. 2006. "Realized Beta: Persistence and Predictability." In T. Fomby and D. Terrell (eds.), *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series in Honor of R.F. Engle and C.W.J. Granger*, vol. B. Oxford, UK: Elsevier Ltd.

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

Bali, T.G. and R.F. Engle. 2010. The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations. *Journal of Monetary Economics* 57(4): 377-390.

Bali, T.G., R.F. Engle, and Y. Tang. 2012. "Dynamic Conditional Beta is Alive and Well in the Cross-Section of Daily Stock Returns." Manuscript, Georgetown University.

Ball, R., and S.P. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25: 51-74.

Barndorff-Nielson, O., P.R. Hansen, A. Lunde, and N. Shephard. 2011. Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. *Journal of Econometrics* 162: 149-169.

Bera, A.K., and M.L. Higgins. 1993. ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys* 7: 305-366.

Bera, A.K., and M.L. Higgins. 1998. "A Survey of ARCH Models." In R. Jarrow (ed.), *Volatility: New Techniques for Pricing Derivatives and Managing Financial Portfolios.* Risk Publications.

Bera, A.K., P. Garcia, and J-S. Roh. 1998. Estimation of Time-Varying Hedge Ratios for Corn and Soybeans: BGARCH and Random Coefficient Approaches. *Sankhya* 59:346-368.

Bollerslev, T., R.F. Engle, and J. Wooldridge. 1988. A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy* 96(1): 116-131.

Bollerslev, T., and J. Wooldridge. 1992. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. *Econometric Reviews 11(2):* 143-172.

Bollerslev, T., R.Y. Chou, and K.F. Kroner. 1992. ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics* 52(1): 5-59.

Bollerslev, T., R.F. Engle, and D.B. Nelson. 1994. "ARCH Models." In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. IV. Amsterdam: North-Holland.

Bollerslev, T., F.X. Diebold, and Wu, J. 2005. “Realized Beta: Persistence and Predictability.” In T. Fomby (ed.), *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series*, vol. B. Amsterdam: Elsevier Science B.V.

Braun, P.A., D.B. Nelson, and A.M. Sunier. 1995. Good News, Bad News, Volatility, and Betas. *The Journal of Finance* 50: 1575-1603.

Brownlees, C., and R.F. Engle. 2011. “Volatility, Correlation and Tails for Systemic Risk Measurement.” Working Paper, Department of Finance, New York University.

Burns, P., R.F. Engle, and J. Mezrich. 1998. Correlations and Volatilities of Asynchronous Data. *Journal of Derivatives* 5(4): 7-18.

Cecchetti, S.G., R.E. Cumby, and S. Figlewski. 1988. “Estimation of the Optimal Futures Hedge.” *Review of Economics and Statistics* 70: 623-630.

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, R.F., D. Hendry, and J.F. Richard. 1983. Exogeneity. *Econometrica* 51(2): 277-304.

Engle, R.F. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics* 20: 339-350.

Engle, R.F. 2009. *Anticipating Correlations: A New Paradigm for Risk Management.* Princeton, NJ: Princeton University Press.

Engle, R.F., C. Pakel, N. Shephard, and K. Sheppard. 2014. “Fitting Vast Dimensional Time-Varying Covariance Models.” Manuscript.

Fama, E., and K. French. 2004. The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives* 18(3): 25-46.

French, K.R. 2014. “Current Research Returns.” http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Kroner, K., and J. Sultan. 1993. Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures. *Journal of Finance and Quantitative Analysis* 28(4): 535-551.

Lien, D., Y.K. Tse, and A.K.C. Tsui. 2000. Evaluating the Hedging Performance of the Constant-Correlation GARCH Model. *Applied Financial Economics* 12(11): 791-798.

Lien, D., and Y.K. Tse. 2002. Some Recent Developments in Futures Hedging. *Journal of Economic Surveys* 16(3): 357-396.

Pagan, A. 1984. Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review* 25(1): 221-247.

Patton, A.J., and M. Verardo. 2012. Does Beta Move with News?: Firm-Specific Information Flows and Learning About Profitability. *Review of Financial Studies* 25(9): 2789-2839.

Rivers, D., and Q. Vuong. 2002. Model Selection Tests for Nonlinear Dynamic Models. *Econometrics Journal* 5: 1–39.

Vuong, Q. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* 57(2): 307-333.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

APPENDIX

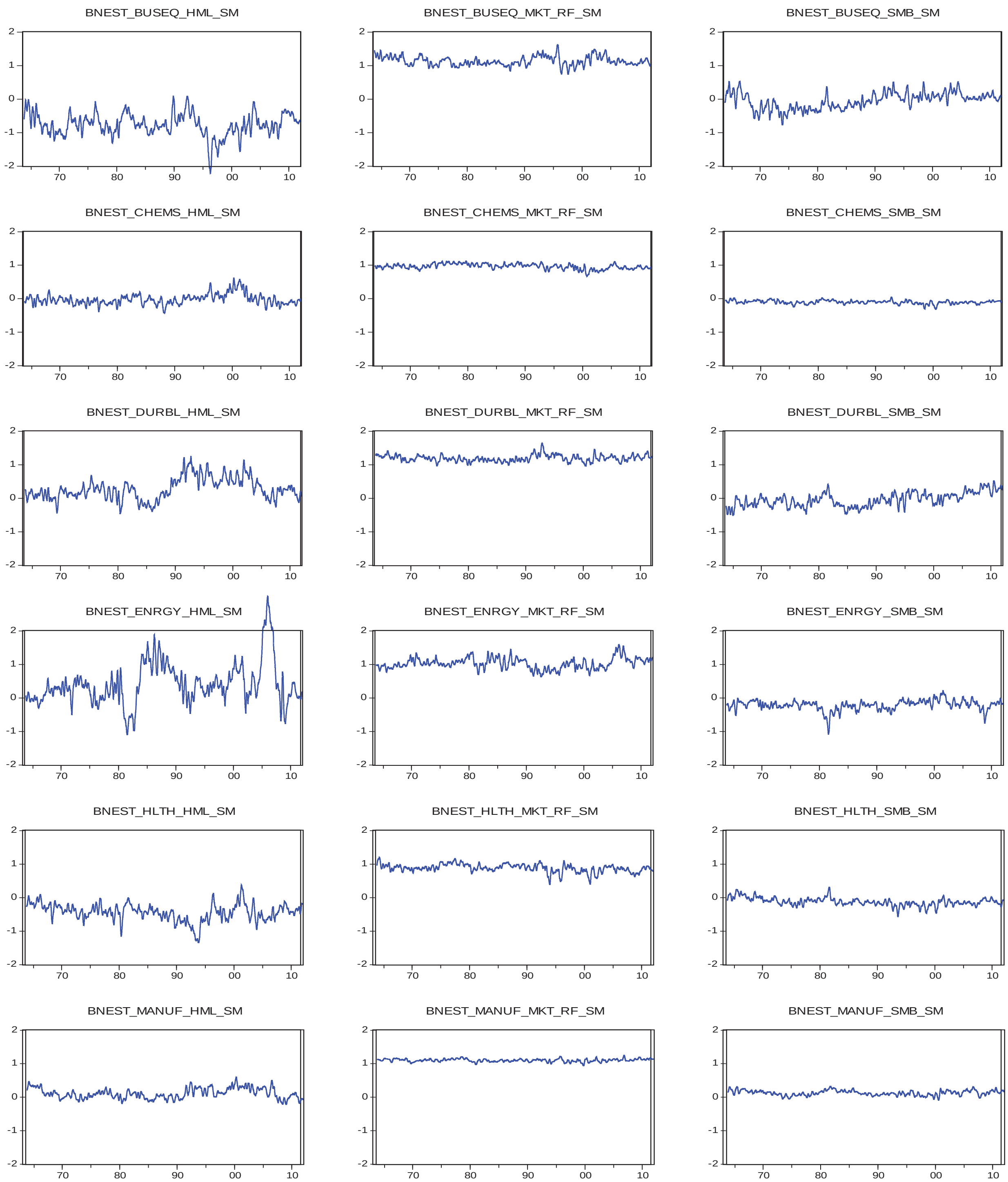

Figure A.1
Nested Betas for Three Factors by Sector

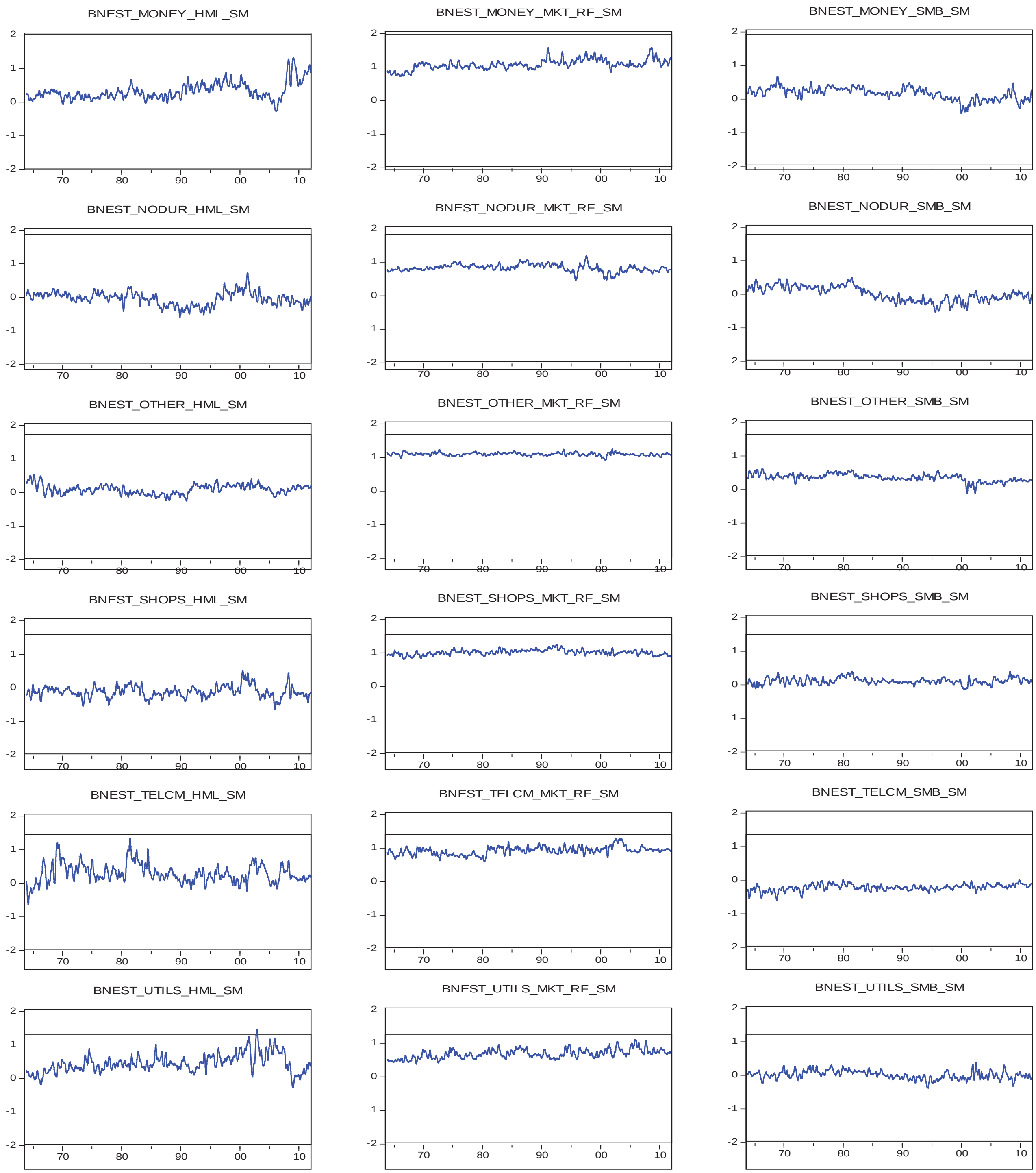

Figure A.2
Nested Betas for Three Factors by Sector