The time-varying GARCH-in-mean model

Gustavo Fruet Dias[a,1]

[a] *CREATES and Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.*
E-mail address: gdias@econ.au.dk

Abstract

I propose an estimation strategy for the stochastic time-varying risk premium parameter in the context of a time-varying GARCH-in-mean (TVGARCH-in-mean) model. A Monte Carlo study shows that the proposed algorithm has good finite sample properties. Using monthly excess returns on the CRSP index, I document that the risk premium parameter is indeed time-varying and shows high degree of persistence.

JEL classification numbers: C13, C15, C22, G12

Keywords: risk-return tradeoff, time-varying coefficients, iterative estimators, GARCH-type models

[1] I thank George Kapetanios, Cristina Scherrer and an anonymous referee for constructive suggestions. I acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

Preprint submitted to Economics Letters *June 2, 2017*

1. Introduction

Asset pricing theories suggest that riskier assets should demand higher expected returns. Using Merton's (1973) theoretical framework, the conditional expectation of the market excess returns reads

$$\mathbb{E}\left(r_{t+1}^{m} \mid \mathcal{F}_t\right) - r_t^f = \lambda_t Var\left(r_{t+1}^{m} \mid \mathcal{F}_t\right), \tag{1}$$

where r_{t+1}^{m} and r_t^f are the returns on the market portfolio and risk-free asset, $\mathcal{F}_t$ is the market-wide information available at time t, and λ_t is the coefficient of relative risk aversion defined as the elasticity of marginal value with respect to wealth. Most studies assume the risk-return trade-off is constant over time and linear in the variance, which is usually associated with the reasons behind mixed empirical evidences when estimating the risk-return trade-off (Linton and Perron (2003), Brandt and Wang (2010), Christensen, Dahl, and Iglesias (2012), among others). To address this issue, I adopt the time-varying GARCH-in-mean (TVGARCH-in-mean) model in the spirit of Anyfantaki and Demos (2016) which allows λ_t to be a time-varying stochastic process and put forward a feasible estimation strategy for λ_t (see references in Anyfantaki and Demos (2016) for variants of the TVGARCH-in-mean models). Specifically, I combine Giraitis, Kapetanios, and Yates's (2013) time-varying kernel least squares estimator with Linton and Perron's (2003) semiparametric iterative approach to estimate the time-varying risk premium coefficient. A Monte Carlo study shows that the proposed algorithm has good finite sample properties. Using the excess returns of the Center for Research on Security Prices (CRSP) index, I document that the risk premium parameter is indeed time-varying, alternating positive (statistically significant) and nonsignificant values over time.

2. The time-varying GARCH-in-mean

The generic TVGARCH-in-mean(p,q) is defined as:

$$r_t = \lambda_t \sigma_t + \epsilon_t, \tag{2}$$

$$\epsilon_t = \sigma_t \eta_t, \tag{3}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \tag{4}$$

$$\epsilon_t^2 = \psi_0 + u_t + \sum_{i=1}^{\infty} \psi_i u_{t-i}, \tag{5}$$

where η_t is an independent and identically distributed (iid) zero mean process with unit variance; σ_t is a latent conditional standard deviation; (5) is the $MA(\infty)$ representation of the conditional variance equation; $u_t = \epsilon_t^2 - \sigma_t^2$ is a martingale difference sequence process; $\phi = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)'$ collects the free parameters in (4); and $\psi_i := \varrho_i(\phi)$ $i = 1, 2, ...$ are deterministic functions of the elements in ϕ. Similarly as in Giraitis, Kapetanios, and Yates (2013), the time-varying risk premium parameters are assumed to evolve smoothly over time, so that it satisfies a local stability condition in the form of $\sup_{s:\|s-t\|\leq h} \| \lambda_t + \lambda_s \|_2 = O_p\left(h/t\right)$.

Estimating the free parameters in (2) and (4) by maximum-likelihood is not a feasible alternative, as the class of TVGARCH-in-mean(p,q) models involves two unobserved processes: λ_t and ϵ_t. Anyfantaki and Demos (2016) address this issue in the context of the time-varying EGARCH(1,1)-in-mean model. Specifically, their work differs from mine in two ways. First, they parameterize the conditional variance as an EGARCH(1,1) model and, most importantly, λ_t as a stationary AR(1) process. By contrast, λ_t in (2) is assumed to satisfy $\sup_{s:\|s-t\|\leq h} \| \lambda_t + \lambda_s \|_2 = O_p\left(h/t\right)$, which encompasses the case of the driftless random walk process considered in Chou, Engle, and Kane (1992). Second, while I propose a kernel-based nonparametric method to estimate the time-varying risk premium parameter, Anyfantaki and Demos's (2016) estimation strategy is based on Bayesian methods (Markov chain Monte Carlo (MCMC) likelihood based estimation procedure).

I combine Linton and Perron's (2003) iterative semiparametric estimator with Giraitis, Kapetanios, and Yates's (2013) kernel-based least squares framework to estimate the free parameters $\theta = \left(\underline{\lambda}, \phi\right)'$, where $\underline{\lambda} = \left(\lambda_1, .., \lambda_T\right)'$. This method consists of recursively updating estimates of σ_t and u_t on each iteration, and then computing estimates of $\underline{\lambda}$ and ϕ. To this end, consider moment conditions based on (2) and (5),

$$\mathbb{E}\left[\sigma_t\left(r_t - \lambda_t\sigma_t\right)\right] = 0, \quad \text{for each } \ t = 1, 2, ..., T, \tag{6}$$

$$\mathbb{E}\left[z_t u_t\right] = 0, \quad \text{with } \ z_t := \frac{\partial\left(\psi_0 + \sum_{i=1}^{\bar{q}} \psi_i u_{t-i}\right)}{\partial\phi}, \tag{7}$$

where (7) is truncated at some lag-order $\bar{q}$ with $\bar{q} > p+q+1$. Notably, (7) holds because u_t is a martingale difference sequence and z_t is a function of lagged values of u_t. It follows that estimating θ by the standard generalized method of moments (GMM) using the moments defined in (6) and (7) is not operational, as z_t and σ_t are latent variables. Using Linton and Perron's (2003) approach, rewrite (6) and (7) using estimates of σ_t and u_t obtained at some j iteration,

$$\mathbb{E}\left[\sigma_{j,t}\left(r_t - \lambda_{j+1,t}\sigma_{j,t}\right)\right] = 0, \quad \text{for each } \ t = 1, 2, ..., T, \tag{8}$$

$$\mathbb{E}\left[z_{j,t} u_{j+1,t}\right] = 0, \tag{9}$$

where $\sigma_{j,t}$ and $z_{j,t}$ denote the filtered estimates of σ_t and z_t based on $\widehat{\theta}_j$, and $u_{j+1,t} = \epsilon_{j,t}^2 - \psi_{j+1,0} - \sum_{i=1}^{\bar{q}} \psi_{j+1,i} u_{j,t-i}$ with $\epsilon_{j,t}^2 = \left(r_t - \lambda_{j+1,t}\sigma_{j,t}\right)^2$. While the finite sample counterpart of (9) is given by the usual sample mean, computing the sample counterpart of (8) is less obvious. The work of Giraitis, Kapetanios, and Yates (2013) suggests the use of local kernels to construct operational sample counterparts of (8). In turn, a feasible moment condition based on (8) reads

$$K_t^{-1} \sum_{\tau=1}^{T} k_{t,\tau} \sigma_{j,\tau} \left(r_\tau - \widehat{\lambda}_{j+1,t} \sigma_{j,\tau} \right) = 0, \quad \text{for each } t = 1, 2, ..., T, \tag{10}$$

where $k_{t,\tau} = K\left((t-\tau)/H\right)$ denotes a kernel function such that $K(x) \geq 0$ for any $x \in \mathbb{R}$ is a continuous bounded function with a bounded first derivative and $\int K(x)dx = 1$; H is the bandwidth parameter satisfying $H = o\left(T/\ln(T)\right)$ as $H \to \infty$; and $K_t = \sum_{\tau=1}^{T} k_{t,\tau}$. Notably, writing the moment conditions as in (10) is consistent with previous studies in the time-varying parameter literature which maximizes kernel weighted log-likelihood functions (see Robinson (1989), Giraitis, Kapetanios, Wetherilt, and Žikeš (2016), among others).

I use the fact that (10) is exactly identified for each t, and hence estimates of λ_t can be obtained independently of ϕ. In turn, estimates of θ are computed iteratively by a two-step procedure. The first step consists of solving (10) for each t, while the second step mimics the work of Linton and Perron (2003) and consists of estimating ϕ using the sample counterpart of (9). In practise, the kernel-based iterative estimator is as follows:

Step 1: Choose starting values $\widehat{\underline{\lambda}}_0$ and $\widehat{\phi}_0$, such that $\widehat{\phi}_0$ satisfies the second-order stationarity conditions of the GARCH(1,1) model. Using $\widehat{\theta}_{0,t} = \left(\widehat{\underline{\lambda}}_0, \widehat{\phi}_0\right)'$, compute recursively $\left\{\sigma_{0,t}^2\right\}_{t=1}^T$, and $\left\{u_{0,t}\right\}_{t=1}^T$ from (2)-(5).

Step 2: Given $\left\{\sigma_{0,t}^2\right\}_{t=1}^T$, calculate

$$\widehat{\lambda}_{1,t} = \left(\sum_{\tau=1}^{T} k_{t,\tau} \sigma_{0,\tau}^2 \right)^{-1} \sum_{\tau=1}^{T} k_{t,\tau} \sigma_{0,\tau} r_\tau, \quad \text{for each } t = 1, 2, .., T. \tag{11}$$

Step 3: Solving the sample counterpart of (9) is equivalent to estimate $\widehat{\phi}_1$ by nonlinear least squares. Calculate

$$\widehat{\phi}_1 = \arg\min_{\widehat{\phi}_1} \sum_{t=1}^{T} \left\{ \left(r_t - \widehat{\lambda}_{1,t} \sigma_{0,t} \right)^2 - \widehat{\psi}_{1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{1,i} u_{0,t-1-i} \right\}^2. \tag{12}$$

Step 4: Update recursively $\left\{\sigma_{1,t}^2\right\}_{t=1}^T$ and $\left\{u_{1,t}\right\}_{t=1}^T$ based on $\widehat{\theta}_1$.

Repeat steps 2-4 j times until $\widehat{\theta}_j$ converges. Convergence occurs when $\left\| \widehat{\underline{\lambda}}_j - \widehat{\underline{\lambda}}_{j-1} \right\|_2 \leq \varepsilon$ and $\left\| \widehat{\phi}_j - \widehat{\phi}_{j-1} \right\|_2 \leq \varepsilon$, with ε set to 10^{-5}. Parameters on the j^{th} iteration are given by:

$$\widehat{\lambda}_{j,t} = \left[\sum_{\tau=1}^{T} k_{t,\tau}\sigma^2_{j-1,\tau}\right]^{-1} \sum_{\tau=1}^{T} k_{t,\tau}\sigma_{j-1,\tau} r_\tau, \quad \text{for each } t = 1, 2, .., T, \tag{13}$$

$$\widehat{\phi}_j = \arg\min_{\widehat{\phi}_j} \sum_{t=1}^{T} \left[\left[r_t - \widehat{\lambda}_{j,t}\sigma_{j-1,t}\right]^2 - \widehat{\psi}_{j,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{j,i} u_{j-1,t-1-i}\right]^2. \tag{14}$$

Finally, three inputs are still necessary to implement the above algorithm: the kernel function, the bandwidth parameter H, and the truncation lag $\bar{q}$. As in Giraitis, Kapetanios, and Yates (2013), three kernel functions are used: the Epanechnikov, Gaussian, and flat kernels. The choice of H follows from the Monte Carlo study conducted in Section 3.1, while $\bar{q}$ is chosen to be proportional to $\ln(T)$, (Dufour and Jouini (2005)).

Asymptotic theory for the Quasi-Maximum Likelihood (QMLE) estimator in the GARCH-in-mean models is yet to be fully established. Conrad and Mammen (2016) give an important step forward and prove the asymptotic distribution of the QMLE estimator for the simple GARCH(1,1)-in-mean. As discussed in Linton and Perron (2003), the semiparametric GARCH-in-mean models offer additional complications compared to the standard GARCH-in-mean models, and hence rigorous inference is still not available. Similar difficulties arise in the TVGARCH-in-mean specification. In turn, this note follows Linton and Perron's (2003) approach as it briefly discusses the general conditions required for consistency and asymptotic normality; uses the wild bootstrap to conduct inference; and adopts a Monte Carlo study to assess the finite sample properties of the proposed iterative estimator.

The concept of asymptotic contraction mapping (ACM) developed in Dominitz and Sherman (2005) is useful to guide the discussion on the asymptotic properties of the kernel iterative estimator. If a collection is an ACM, then it will have a unique fixed point that depends on the sample characteristics and hence the iterative procedure converges.[2] While the two-step procedure given in (13) and (14) is seen as the sample mapping, (8) and (9) are their population counterpart. Consistency and asymptotic normality require the population mapping to be an ACM, which implies, under some uniform convergence conditions, that the sample mapping is also an ACM and hence has an unique fixed point (regardless of the initial values). Combining Theorem 4 in Dominitz and Sherman (2005) with Giraitis, Kapetanios, and Yates's (2013) results, estimates of λ_t are expected to be $\sqrt{H}$ consistent and asymptotically normally distributed, and estimates of ϕ are expected to be consistent and asymptotically normally distributed at the usual $\sqrt{T}$ rate.

[2]See Dominitz and Sherman, 2005, p. 841 for a formal definition of the ACM.

3. Numerical illustrations

3.1. Monte Carlo

I simulate data from (2)-(4) where $p = q = 1$, η_t is normally distributed with zero mean and variance equal to one, and λ_t follows a bounded random walk process (see detailed discussion in the online Supplement). The sample size and the number of replications are set to 2,000 and 1,000, respectively.

Table 1 displays the results for the kernel-based iterative estimator computed with alternative bandwidth choices and the Epanechnikov, Gaussian, and flat kernels. Results are reported in terms of the root mean squared error (RMSE) and pointwise correlation between λ_t and the kernel-based estimates. The best choices of bandwidth parameters, in terms of minimizing the RMSE, are $H = T^{1/2}$ and $H = T^{6/10}$. These are also the bandwidths that deliver the highest pointwise correlation (about 0.85) between the kernel-based estimates and the true latent time-varying risk premium parameter. All combinations of kernel methods and bandwidth parameters deliver unbiased estimates of $\phi = (\omega, \alpha, \beta)'$ (apart from $H = T^{2/10}$). Finally, convergence rates are greater than 98% for all specifications, suggesting that (8) and (9) are ACMs.

3.2. Empirical results

I use excess returns of the CRSP value-weighted index aggregated on a monthly basis. Figure 1 plots monthly estimates of λ_t and their 90% wild bootstrap confidence bands from a TVGARCH(1,1)-in-mean model with bandwidth $H = T^{6/10}$ (see the Supplement material for a Monte Carlo study showing that the wild bootstrap produces valid inference). Not surprisingly, likewise the semiparametric GARCH-in-mean models, the empirical confidence bands are relatively wide, which reflects the difficulties associated with estimating the risk-return trade-off (Linton and Perron (2003) and Christensen, Dahl, and Iglesias (2012)). I find that the risk premium parameter is indeed time-varying, with $\widehat{\lambda}_t$ assuming both positive (generally significant) and negative (insignificant) values. This finding sheds light on the mixed evidence on the risk-return literature regarding the sign and significance of the risk premium parameter. Additionally, in periods where $\widehat{\lambda}_t$ is statistically significant, market volatility is low. On contrary, when $\widehat{\lambda}_t$ is not statistically significant, market volatility is high. This indicates that identification of the risk premium parameter is problematic in periods of high volatility (Rossi and Timmermann (2010)).

4. Conclusion

I introduce a kernel-based iterative estimator that combines the estimators in Giraitis, Kapetanios, and Yates (2013) and Linton and Perron (2003) to estimate the stochastic time-varying risk premium parameter in the TVGARCH(1,1)-in-mean model. The Monte Carlo study shows that the kernel-based estimator presents a good finite sample performance. I investigate the time-varying risk premium for the CRSP index and find strong evidence that λ_t is indeed time-varying.

Anyfantaki, S., and A. Demos (2016): “Estimation and Properties of a Time-Varying EGARCH(1,1) in Mean Model,” *Econometric Reviews*, 35(2), 293–310.

Brandt, M. W., and L. Wang (2010): “Measuring the Time-Varying Risk-Return Relation from the Cross-Section of Equity Returns,” Discussion paper, Duke University.

Chou, R., R. F. Engle, and A. Kane (1992): “Measuring risk aversion from excess returns on a stock index,” *Journal of Econometrics*, 52(1), 201 – 224.

Christensen, B. J., C. M. Dahl, and E. M. Iglesias (2012): “Semiparametric inference in a GARCH-in-mean model,” *Journal of Econometrics*, 167, 458–472.

Conrad, C., and E. Mammen (2016): “Asymptotics for parametric GARCH-in-Mean models,” *Journal of Econometrics*, 194(2), 319 – 329, Financial Statistics and Risk Management.

Dominitz, J., and R. Sherman (2005): “Some Convergence Theory for Iterative Estimation Procedures with an Application to Semiparametric Estimation,” *Econometric Theory*, 21, 838–863.

Dufour, J.-M., and T. Jouini (2005): “Asymptotic Distribution of a Simple Linear Estimator for Varma Models in Echelon Form,” in *Statistical Modeling and Analysis for Complex Data Problems*, ed. by P. Duchesne, and B. Rémillard, pp. 209–240. Springer US.

Giraitis, L., G. Kapetanios, A. Wetherilt, and F. Žikeš (2016): “Estimating the Dynamics and Persistence of Financial Networks, with an Application to the Sterling Money Market,” *Journal of Applied Econometrics*, 31(1), 58–84, jae.2457.

Giraitis, L., G. Kapetanios, and T. Yates (2013): “Inference on stochastic time-varying coefficient models,” *Journal of Econometrics*.

Linton, O., and B. Perron (2003): “The Shape of the Risk Premium: Evidence From a Semiparametric Generalised Autoregressive Condidional Heteroscedasticity Model,” *Journal of Business & Economic Statistics*, 21, 354–367.

Merton, R. C. (1973): “An Intertemporal Capital Asset Pricing Model,” *Econometrica*, 41(5), 867–887.

Robinson, P. (1989): “Nonparametric Estimation of Time-Varying Parameters,” in *Statistics Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl, chap. 15, pp. 253–264. Springer, Berlin.

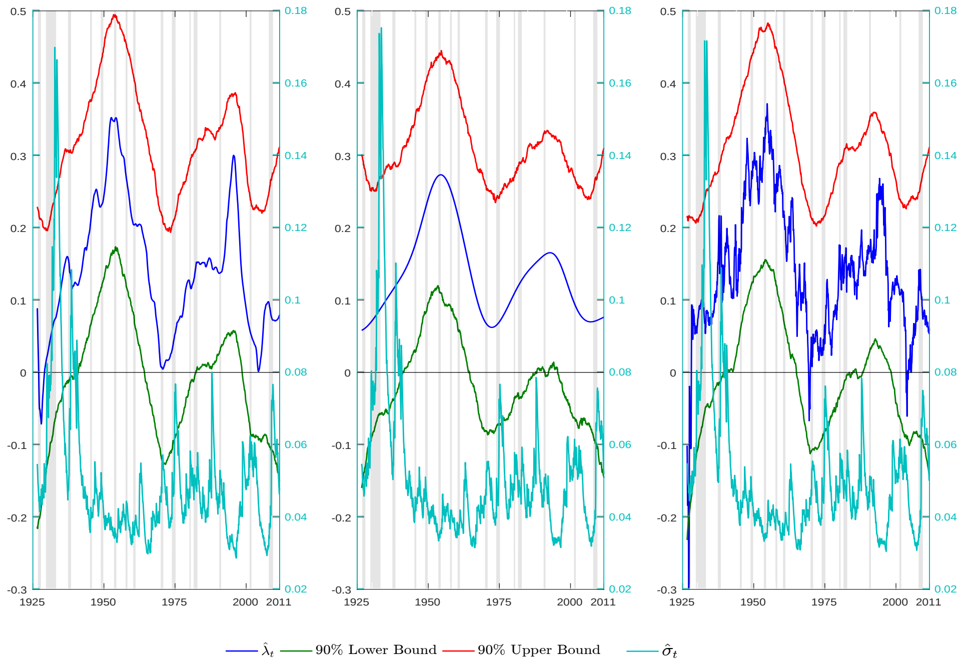
Rossi, A., and A. Timmermann (2010): “What is the Shape of the Risk-Return Relation?,” Discussion paper, The Rady School of Management, University of California, San Diego.

Table 1: Finite sample performance of the kernel-based iterative estimator

		λ_t		RMSE			Mean		
Bandwidth - H	Kernel	RMSE	Corr	ω	α	β	ω	α	β
$T^{2/10}$	Epanechnikov	0.39	0.60	0.01	0.03	0.10	0.02	0.07	0.84
$T^{2/10}$	Gaussian	0.25	0.71	0.01	0.02	0.06	0.01	0.05	0.88
$T^{2/10}$	Flat	0.36	0.61	0.01	0.02	0.07	0.01	0.06	0.86
$T^{3/10}$	Epanechnikov	0.26	0.72	0.01	0.02	0.05	0.01	0.05	0.88
$T^{3/10}$	Gaussian	0.18	0.81	0.01	0.02	0.05	0.01	0.05	0.89
$T^{3/10}$	Flat	0.24	0.73	0.01	0.02	0.05	0.01	0.05	0.89
$T^{4/10}$	Epanechnikov	0.18	0.80	0.01	0.02	0.05	0.01	0.05	0.90
$T^{4/10}$	Gaussian	0.13	0.86	0.01	0.02	0.06	0.01	0.05	0.90
$T^{4/10}$	Flat	0.17	0.81	0.01	0.02	0.05	0.01	0.05	0.90
$T^{5/10}$	Epanechnikov	0.13	0.86	0.01	0.02	0.05	0.01	0.05	0.90
$T^{5/10}$	Gaussian	0.11	0.88	0.01	0.02	0.05	0.01	0.05	0.90
$T^{5/10}$	Flat	0.13	0.86	0.01	0.02	0.05	0.01	0.05	0.90
$T^{6/10}$	Epanechnikov	0.12	0.88	0.01	0.02	0.06	0.01	0.05	0.89
$T^{6/10}$	Gaussian	0.13	0.86	0.01	0.02	0.06	0.01	0.05	0.89
$T^{6/10}$	Flat	0.12	0.86	0.01	0.02	0.06	0.01	0.05	0.89
$T^{7/10}$	Epanechnikov	0.13	0.85	0.01	0.02	0.05	0.01	0.05	0.90
$T^{7/10}$	Gaussian	0.16	0.79	0.01	0.02	0.05	0.01	0.05	0.90
$T^{7/10}$	Flat	0.15	0.81	0.01	0.02	0.05	0.01	0.05	0.90
$T^{8/10}$	Epanechnikov	0.17	0.77	0.01	0.02	0.05	0.01	0.05	0.89
$T^{8/10}$	Gaussian	0.21	0.68	0.01	0.02	0.05	0.01	0.05	0.89
$T^{8/10}$	Flat	0.19	0.67	0.01	0.02	0.05	0.01	0.05	0.89

RMSE accounts for root mean squared error; Corr is the pointwise correlation between λ_t and $\widehat{\lambda}_t$; λ_t is defined as a bounded random walk process with upper and lower bounds given by 0.90 and -0.90, respectively; and $\phi = (0.01, 0.05, 0.9)'$.

Figure 1: Time-varying risk premium estimation and conditional standard deviation



The left-hand-side, center and right-hand-side graphs display estimates of λ_t with the Epanechnikov, Gaussian, and the flat kernel functions, respectively. Estimates of λ_t and the 90% confidence intervals are on the left axis. The bandwidth parameter is equal to $H = T^{6/10}$. The conditional standard deviation is in light blue on the right axis. I perform 1000 replications in the bootstrap algorithm. The shaded areas account for the recession periods from the National Bureau of Economic Research (NBER).