Web Appendix to
Multivariate High-Frequency-Based Volatility (HEAVY) Models

Diaa Noureldin
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
diaa.noureldin@economics.ox.ac.uk

Neil Shephard
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
neil.shephard@economics.ox.ac.uk

Kevin Sheppard
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
kevin.sheppard@economics.ox.ac.uk

April 22, 2011

A Results from Matrix Algebra and Calculus

The following results can be found in Magnus (1988). The *vec* operator stacks the columns of a $(k \times k)$ matrix into a $(k^2 \times 1)$ vector. The *vech* operator stacks the lower triangular part including the main diagonal of a $(k \times k)$ symmetric matrix into a $(k^* \times 1)$ vector, $k^* = k(k+1)/2$.

For any matrices $A(k \times l)$, $B(l \times m)$ and $C(m \times n)$, $vec(ABC) = (C' \otimes A)vec(B)$. For any $(k \times k)$ symmetric matrix A, the $(k^* \times k^2)$ elimination matrix L_k is defined such that $vech(A) = L_k vec(A)$, and the $(k^2 \times k^*)$ duplication matrix D_k is defined such that $vec(A) = D_k vech(A)$. Let Λ_{ij} be a $(k \times k)$ matrix with 1 in its $(i,j)^{th}$ position and zeros elsewhere. Then L_k and D_k can be obtained using $L_k = \sum_{i \geq j} vech(\Lambda_{ij})(vec(\Lambda_{ij}))'$ and $D_k = \sum_{i>j} vec(\Lambda_{ij} + \Lambda_{ji})(vech(\Lambda_{ij}))' + \sum_{i=j} vec(\Lambda_{ij})(vech(\Lambda_{ij}))'$.

The above results can be combined to give the following result for any $(k \times k)$ matrices A and B, with B being symmetric

$$vech(ABA') = L_k vec(ABA') = L_k(A \otimes A)vec(B) = L_k(A \otimes A)D_k vech(B). \tag{A.1}$$

For a $(k \times l)$ matrix function $F(X)$ and a $(m \times n)$ matrix of variables X, the derivative of $F(X)$ with respect to X, denoted by the $(kl \times mn)$ matrix $DF(X)$, is given by

$$DF(X) = \frac{\partial vec(F(X))}{\partial (vec(X))'}.$$

Also for any $(k \times k)$ matrix A,

$$\frac{\partial A^{-1}}{\partial A} = -((A^{-1})' \otimes A^{-1}), \frac{\partial \log |A|}{\partial A} = (vec((A^{-1})'))', \text{ and } \frac{\partial tr(AX)}{\partial X} = (vec(A'))'.$$

B The Wishart Distribution: An Overview

The Wishart distribution is the matrix-variate generalization of the χ^2 distribution. It is the sampling distribution of the sample covariance matrix for random draws from the multivariate normal distribution; see Gupta and Nagar (2000) for a detailed treatment. We begin our overview by relating the Wishart distribution to the multivariate normal and the matrix-variate normal. Let $x_1, ..., x_n$ be independent random vectors drawn from the (centered) multivariate normal distribution where each $(k \times 1)$ vector $x_i \sim N_k(0, \Sigma)$, $i = 1, ..., n$. Then the $(k \times n)$ matrix $X = (x_1, ..., x_n)$ has a matrix variate normal distribution denoted as $X \sim N_{k,n}(0, \Sigma \otimes I_{\mathbf{n}})$. If $n \geq k$, then $S = XX'$ is a positive definite $(k \times k)$ matrix and follows a (centered) Wishart distribution, denoted as $S \sim W_k(n, \Sigma)$, where n is integer degrees of freedom and Σ is the scale matrix. The unconditional moments of S are given by $\mathrm{E}(S) = n\Sigma$, and $\mathrm{Var}(vec(S)) = 2nD_k D_k^+(\Sigma \otimes \Sigma)$.

The density of S is given by

$$W_k(n, \Sigma) = \frac{|S|^{\frac{n-k-1}{2}}}{2^{\frac{nk}{2}} \Gamma_k(\frac{n}{2}) |\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} tr(\Sigma^{-\mathbf{1}} S)\right), \qquad n \geq k,$$

where $\Gamma_k\left(\frac{n}{2}\right) = \pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma\left(\frac{n}{2} + (1-j)/2\right)$ is the multivariate gamma function. It is also useful to define a standardized Wishart distribution such that if $S \sim SW_k(n, \Sigma)$, where SW_k denote a standardized Wishart distribution of dimension k, we have $\mathrm{E}(S) = \Sigma$ instead of $\mathrm{E}(S) = n\Sigma$ under the non-standardized Wishart. Note that $SW_k(n, \Sigma)$ is equivalent to $W_k(n, n^{-1}\Sigma)$. In this case, the density of S is given by

$$SW_k(n, \Sigma) = \frac{n^{\frac{nk}{2}} |S|^{\frac{n-k-1}{2}}}{2^{\frac{nk}{2}} \Gamma_k(\frac{n}{2}) |\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{n}{2} tr(\Sigma^{-\mathbf{1}} S)\right), \qquad n \geq k. \tag{B.1}$$

If the $(k \times n)$ matrix X does not have full column rank (i.e. $n < k$), $S = XX'$ follows instead a singular Wishart distribution. Srivastava (2003) derives the density of the singular Wishart as

$$SINGW_k(n, \Sigma) = \frac{\pi^{(-kn+n^2)/2} \left|\widetilde{S}\right|^{\frac{n-k-1}{2}}}{2^{\frac{nk}{2}} \Gamma_n(\frac{n}{2}) \left|\Sigma\right|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} tr(\Sigma^{-\mathbf{1}} S)\right), \quad n < k, \tag{B.2}$$

where $\widetilde{S}$ is a diagonal matrix containing the non-zero eigenvalues of S along the main diagonal.

One of the key results on Wishart distributions is that if $S \sim W_k(n, \Sigma)$, then $ASA' \sim W_k(n, A\Sigma A')$ for any $(k \times k)$ nonsingular matrix A. Srivastava (2003) extends this result to the singular Wishart case. Based on the ranks of P_t and V_t, we use this result in specifying the distribution of the innovation matrices ε_t and η_t as discussed in Section 3.1.

C Further Empirical Results

C.1 Bivariate Scalar HEAVY Model: Other Asset Pairs

We estimate the scalar HEAVY model for other pairs of assets selected from the ten DJIA stocks. The pairing of the assets is chosen by selecting companies in the same sector (e.g. BAC-JPM and IBM-MSFT), where we expect more persistent correlation dynamics, and also pairs of companies in different sectors. The objective is to track the HEAVY model's performance in each case. Table 1 includes the parameter estimates for the HEAVY, GARCH and GARCH-X models. One notable feature is that the estimates do not display large variation across the different pairs. As in the SPY-BAC case, inclusion of the realized measure crowds out the outer product of returns as the coefficient D_{GX} is statistically insignificant in all cases.

The decomposition of the log-likelihood gains shows that the HEAVY model gains are obtained for each pair with respect to both margins and the copula with only one exception. The HEAVY model log-likelihood gain for the joint distribution is uniform across all pairs. The predictive ability test results indicate that the HEAVY model performs better than GARCH for all asset pairs, with the gains being particularly significant at short forecast horizons. We do not report the margins-copula decomposition for these tests in the interest of brevity, but they show that the HEAVY model gains are maintained for some of the margins and also for the copula of some of the pairs. In no case was the GARCH model significantly favoured at any horizon except for the BAC and JPM margins towards the end of the forecast horizon.

C.2 Bivariate Diagonal HEAVY Model: SPY-BAC and Other Asset Pairs

We discuss the estimation and forecast evaluation results only for the diagonal HEAVY and GARCH models. We exclude the GARCH-X model results to improve presentation noting that its results are in line with those of the scalar model. The top panel of Table 2 presents estimates of the diagonal elements of the parameter matrices in (5)-(6), in order, along with those of the corresponding

	HEAVY-P		GARCH		GARCH-X			HEAVY-V	
	A_H	B_H	A_G	B_G	A_{GX}	B_{GX}	D_{GX}	A_M	B_M
BAC - JPM	0.260	0.639	0.062	0.938	0.180	0.702	0.046	0.450	0.550
IBM - MSFT	0.179	0.762	0.051	0.941	0.135	0.794	0.026	0.309	0.676
XOM - AA	0.188	0.737	0.057	0.935	0.114	0.804	0.034	0.315	0.667
AXP - DD	0.201	0.743	0.045	0.951	0.183	0.754	0.010	0.357	0.638
GE - KO	0.220	0.727	0.039	0.957	0.205	0.740	0.007	0.344	0.651

	HEAVY-P log-likelihood gains/losses (+/-) relative to GARCH			
	Margin 1	Margin 2	Copula	Joint LL
BAC - JPM	44	58	8	110
IBM - MSFT	30	37	17	84
XOM - AA	44	31	-7	69
AXP - DD	56	49	16	121
GE - KO	47	34	9	90

	Joint distribution predictive ability tests at different forecast horizons (days)					
	(1)	(2)	(3)	(5)	(10)	(22)
BAC - JPM	-4.09	-3.41	-2.78	-1.98	0.45	1.32
IBM - MSFT	-2.92	-2.60	-2.38	-1.94	-1.56	-0.46
XOM - AA	-2.76	-2.09	-2.05	-1.76	-1.32	-0.80
AXP - DD	-3.46	-3.09	-2.84	-2.39	-1.28	0.18
GE - KO	-2.80	-2.66	-2.41	-2.21	-0.91	1.00

Table 1: Scalar HEAVY estimation and forecast evaluation results for other pairs of assets. Top panel: parameter estimates of HEAVY, GARCH and GARCH-X. BAC-JPM HEAVY-V estimates add up to 1 due to rounding. All coefficients are statistically significant at the 5 percent significance level. Middle panel: HEAVY-P log-likelihood gains/losses relative to GARCH. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

	HEAVY-P				GARCH				HEAVY-V			
	$\overline{A}_H$		$\overline{B}_H$		$\overline{A}_G$		$\overline{B}_G$		$\overline{A}_M$		$\overline{B}_M$	
SPY-BAC (st. error)	0.447 (0.048)	0.477 (0.057)	0.858 (0.033)	0.844 (0.041)	0.238 (0.023)	0.262 (0.035)	0.968 (0.006)	0.964 (0.010)	0.632 (0.025)	0.663 (0.033)	0.768 (0.022)	0.748 (0.028)
	A_H		B_H		A_G		B_G		A_M		B_M	
Var. eqn. (SPY)	0.200		0.736		0.057		0.938		0.400		0.590	
Var. eqn. (BAC)	0.228		0.713		0.069		0.929		0.439		0.560	
Cov. eqn.	0.213		0.725		0.063		0.933		0.419		0.575	

	Log-likelihood decomposition (HEAVY-P versus GARCH)		
	HEAVY-P	GARCH	HEAVY-P gains
Margin 1 (SPY)	-659	-713	54
Margin 2 (BAC)	-1,592	-1,647	55
Copula	816	809	7
Joint distribution	-1,435	-1,552	117

	Predictive ability tests at different forecast horizons (days)					
	(1)	(2)	(3)	(5)	(10)	(22)
Margin 1 (SPY)	-3.86	-3.24	-2.77	-1.78	0.27	1.50
Margin 2 (BAC)	-3.17	-2.45	-1.79	-0.80	0.89	1.90
Copula	-3.29	-3.02	-3.04	-3.04	-2.99	-3.88
Joint distribution	-4.38	-3.79	-3.27	-2.54	-0.46	0.70

Table 2: Diagonal HEAVY estimation and forecast evaluation results for SPY-BAC. Top panel: parameter estimates of HEAVY and GARCH with standard errors reported in parentheses. Middle panel: decomposition of the log-likelihood (excluding constant terms) at the estimated parameter values. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

GARCH model. These are easier to interpret when expressed in terms of the parameters of the *vech* representation in (7)-(8), which are reported underneath. Note that if $\overline{A}_H$ is, for instance, a (2×2) diagonal matrix, then A_H will be a (3×3) diagonal matrix. The first and third diagonal elements of A_H will be the squares of the diagonal elements of $\overline{A}_H$, and the second diagonal element of A_H will be the product of the two diagonal elements of $\overline{A}_H$.

The estimates of the diagonal elements are rather similar within each parameter matrix, except for the HEAVY-V equation. Since the diagonal HEAVY model nests the scalar HEAVY model, we can test for the restriction using a Wald test. The scalar restriction is not rejected for both the HEAVY and GARCH models at the 5% significance level. The log-likelihood decomposition results are similar to the scalar model. The bottom panel of Table 2 shows that the diagonal HEAVY model provides superior forecasts with the gains being particularly significant at short forecast horizons.

We also report estimation results for the diagonal model using other pairs of assets in Table 3. For brevity, we only report parameter estimates for the *vech* representation. The parameter estimates show some variation within and across pairs. The Wald test results indicate that the scalar model restrictions are rejected at the 5% significance level only for the XOM-AA pair in

	HEAVY-P		GARCH		HEAVY-V	
	A_H	B_H	A_G	B_G	A_M	B_M
Variance (BAC)	0.267	0.638	0.051	0.947	0.433	0.566
Variance (JPM)	0.256	0.634	0.074	0.925	0.473	0.526
Covariance (BAC-JPM)	0.262	0.636	0.061	0.936	0.452	0.546
Variance (IBM)	0.187	0.761	0.052	0.939	0.331	0.652
Variance (MSFT)	0.172	0.764	0.049	0.945	0.291	0.695
Covariance (IBM-MSFT)	0.180	0.763	0.050	0.942	0.311	0.673
Variance (XOM)	0.175	0.713	0.073	0.907	0.338	0.644
Variance (AA)	0.180	0.784	0.043	0.952	0.285	0.698
Covariance (XOM-AA)	0.178	0.748	0.056	0.929	0.310	0.670
Variance (AXP)	0.218	0.738	0.065	0.931	0.371	0.628
Variance (DD)	0.186	0.740	0.031	0.963	0.338	0.645
Covariance (AXP-DD)	0.201	0.739	0.045	0.947	0.354	0.636
Variance (GE)	0.211	0.750	0.042	0.956	0.354	0.645
Variance (KO)	0.283	0.610	0.037	0.957	0.331	0.653
Covariance (GE-KO)	0.245	0.676	0.039	0.956	0.342	0.649

Log-likelihood decomposition (HEAVY-P versus GARCH)

	HEAVY-P	GARCH	HEAVY-P gains
BAC - JPM	-2,828	-2,936	108
IBM - MSFT	-2,295	-2,380	85
XOM - AA	-3,415	-3,485	71
AXP - DD	-3,155	-3,271	116
GE - KO	-2,211	-2,304	93

Predictive ability tests at different forecast horizons (days)

	(1)	(2)	(3)	(5)	(10)	(22)
BAC - JPM	-3.78	-3.13	-2.47	-1.68	0.69	1.84
IBM - MSFT	-2.91	-2.59	-2.39	-1.94	-1.57	-0.49
XOM - AA	-2.84	-2.15	-2.14	-1.87	-1.42	-0.93
AXP - DD	-3.24	-2.94	-2.75	-2.33	-1.43	-0.24
GE - KO	-2.87	-2.71	-2.50	-2.25	-1.04	0.83

Table 3: Diagonal HEAVY parameter estimates for other pairs of assets. Top panel: parameter estimates of HEAVY and GARCH. All coefficients are statistically significant at the 5 percent significance level. Middle panel: HEAVY-P and GARCH log-likelihood (excluding constant terms) at the estimated parameter values. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

the HEAVY-P equation, and only for the AXP-DD pair in the GARCH model. The scalar model restrictions for the HEAVY-V equation are not rejected in any of the pairs. The figures in the middle panel shows that the HEAVY model gains over GARCH in terms of the joint distribution log-likelihood are uniform across all pairs. The gains in the margins and the copula - not reported for brevity - mirror the results of the corresponding scalar models; see middle panel of Table 1. The t-statistics of the predictive ability tests in the bottom panel indicate that the HEAVY model consistently outperforms the GARCH model.

References

Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*. New York: Chapman & Hall.

Magnus, J. R. (1988). *Linear Structures*. London: Griffin.

Srivastava, M. S. (2003). Singular Wishart and multivariate Beta distributions. *Annals of Statistics 31*, 1537–1560.