Essays

# Financial econometrics – A new discipline with new methods

## Robert Engle

*Department of Economics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA and Stern School of Business, New York University, 44 West 4th Street, Suite 9-160, New York, NY 10012, USA*

Financial econometrics is simply the application of econometric tools to financial data. For many years, least-squares techniques provided satisfactory tools. Stock market forecasts, efficient market tests, and even tests of portfolio models such as the CAPM and APT were essentially implemented with least squares on cleverly manipulated data sets. More recently, however, the field has developed its individual character as new statistical tools have been invented to analyze new questions.

In this short overview, I would like to suggest a framework that includes much of the recent literature and important tools of financial econometrics. Let $P_t$ be a vector of asset prices observed at time $t$, and let $F_t$ be the information set known to the econometrician at time $t$ which automatically must include these prices. Corresponding to the price change and dividend payment from $t$ to $t + k$ is a return vector $R_{t+k,t}$. A central concern of financial econometrics is to discover the joint conditional density $f(P_{t+k}|F_t)$. Estimates of the conditional mean, $\mu_{t+k/t} = E_t(P_{t+k})$ of this density were used to test the efficient markets hypothesis, which, in its simplest form, supposed that expected excess returns should be zero. New econometric methods were then introduced to estimate the conditional variances and covariances, $\Omega_{t+k,t} = E_t(P_{t+k} - \mu_{t+k/t})(P_{t+k} - \mu_{t+k/t})'$ of these prices and returns. The first models were the ARCH and GARCH models of Engle (1982) and Bollerslev (1986) and then the stochastic volatility models of Taylor (1986) and Harvey et al. (1994). Multivariate GARCH methods were implemented by Bollerslev et al. (1988), Bollerslev (1990), and Engle and Kroner (1995). Engle and Gonzales-Rivera (1991) allowed general non-normal errors but a paper by Hansen (1994) is one of the few successful efforts to estimate time-varying higher moments of this density. In response to the needs of regulators and risk managers for calculations of value at

risk, new methods now are being designed to examine the tails of this distribution. It is not clear whether the tails have the same dynamic behavior as the rest of the distribution as would be assumed by GARCH style models. The new models include the Hybrid model of Boudoukh et al. (1998), the CAViaR model of Engle and Manganelli (1999), and extreme value theory estimation of tail shapes as in Embrechts et al. (1997), and McNiel and Frey (2000).

This collection of methods has generally been successful in parameterizing and estimating conditional densities. The most significant unsolved problem is the multivariate extension of many of these methods. Although various multivariate GARCH models have been proposed, there is no consensus on simple models that are satisfactory for big problems. There has been little work on multivariate tail properties or other conditional moments. There is intriguing evidence of interesting non-linearities in correlation. A second important extension that is receiving a great amount of attention, is the development of methods to use intra-daily data and ultimately transactions data, called by Engle (2000) ultra-high-frequency data (UHP), to improve estimation of conditional densities.

All these methods are focussed on moments of the empirical conditional density, $f$. However, another object of interest is the risk neutral conditional density, $f^*$, which is the set of probabilities of returns or prices under which an agent would value assets by their discounted expected value. This density is known to exist and to be given by

$$f^*(P_{t+k}|F_t) = b_t^{-1} m(P_{t+k}) f(P_{t+k}|F_t) \tag{1}$$

if and only if there are no arbitrage opportunities. The positive function $m$ is the pricing kernel, interpreted as price per unit probability, which is unique if markets are complete. The scalar $b_t = E_t(m)$ ensures that $f^*$ is a density. The risk neutral density has the property that assets with payoff $g(P_{t+k})$ should have a price

$$S_t = \int m(u) g(u) f(u|F_t) \, du = b_t \int g(u) f^*(u|F_t) \, du = b_t E_t * (g(P_{t+k})). \tag{2}$$

Since this formula applies to all assets including ones with a sure payment, it is clear that $b_t$ is the price of a pure discount bond. When $g$ is the identity, it gives an expression for the risk premium for the underlying asset. When the payoff function is

$$g(P_{t+k}) = \max\{P_{t+k} - K, 0\}, \tag{3}$$

this expression prices European Call Options.

Two different strands of financial econometrics have approached the option-pricing problem. The first observes options data and infers $f^*$; the second specifies $m$, estimates $f$, and computes the option price. The estimation of $f^*$ is sometimes called arbitrage free pricing, since the existence of a density which

prices all assets, insures no arbitrage opportunities. A variety of methods have been proposed which involve some interpolation of options prices to all strikes and then inversion of (2) to get $f^*$. See for example, Shimko (1993) and Rubinstein (1994). These methods have no implication for risk management as the density $f^*$ is simply estimated at each point in time and may have any sort of time series behavior. For a dramatic demonstration, see Dumas Fleming and Whaley (1998).

The second approach is closer to conventional econometric research. It is typical to appeal to the Black and Scholes continuous hedging argument and Girsanov's Theorem to infer the pricing kernel; however, much recent research which is aimed directly at parameterizing and estimating $m$, has shown that this may be too simple a specification to replicate the properties of options prices. See for example, Jackwerth and Rubinstein (1996) and Rosenberg and Engle (1999).

An important class of models assumes that the true data generating process is a continuous-time diffusion process. In this case, both the empirical and risk neutral density can be computed. Black and Scholes derived their option pricing model based upon geometric Brownian motion but many other models have been proposed and have their own options prices. The econometric problem is to estimate the parameters of the diffusion process based upon the discretely observed data. It has generally been found that simple diffusion processes do not fit observed data so interest has focussed on mean reverting processes such as Ornstein–Uhlenbeck, jump diffusions, and affine family models such as the square root diffusion. The econometrics of these models is difficult, often requiring simulated method of moments or characteristic function estimation. There is a challenge here in finding a satisfactory diffusion model that also supports option pricing. However, models that are most convenient for option pricing incorporate the assumption that options are redundant assets and perhaps this is not the case.

The use of higher and higher frequency data potentially could provide information on the appropriate class of diffusion models to use for pricing both underlying and derivative assets. To analyze UHF data, it is necessary to model not only the characteristics of each trade, but also the timing. Engle and Russell (1998) introduce the Autoregressive Conditional Duration model that estimates the distribution of arrival times for the next event conditional on all past information. Dufour and Engle (2000) following Hasbrouck's (Hasbrouck, 1991) vector autoregression, show that the more frequent the transactions, the greater the volatility and price response to trades and transaction arrivals are predictable based upon economic variables such as the bid ask spread. Econometric models of transactions, quotes, prices and volumes support many of the implications of market microstructure theory. Potentially, these models should yield valuable information for market designers and risk managers. These models may also serve as underlying data generating processes for calculating empirical

and risk neutral density forecasts and consequently for options pricing. This is an important challenge for models of empirical microstructure.

## References

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.
Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Review of Economics and Statistics 72, 498–505.
Bollerslev, T., Engle, R., Wooldridge, J.M., 1988. A capital asset pricing model with time varying covariances. Journal of Political Economy 96, 116–131.
Boudoukh, J., Richardson, M., Whitelaw, R.F., 1998. The best of both worlds. Risk 11, 64–67.
Dufour, A., Engle, R., 2000. Time and the price impact of a trade. Journal of Finance, forthcoming.
Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: Empirical tests. Journal of Finance 53 (6), 2059–2106 DEC.
Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer, Berlin.
Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1008.
Engle, R., 2000. The econometrics of ultra high frequency data. Econometrica 68, 1–22.
Engle, R., Gonzalez-Rivera, G., 1991. Semi-parametric ARCH models. Journal of Business and Economic Statistics 9, 345–359.
Engle, R., Manganelli, S., 1999. CAViaR: conditional value at risk by regression quantiles. Manuscript, UCSD.
Engle, R., Kroner, K., 1995. Multivariate simultaneous GARCH. Econometric Theory 11, 122–150.
Engle, R., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.
Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.
Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic models. Review of Financial Studies 61, 247–264.
Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179–207.
Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51 (5), 1611–1631.
McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance, forthcoming.
Rosenberg, J., Engle, R., 1999. Empirical pricing kernels. Manuscript UCSD, and NYU.
Rubinstein, M., 1994. Implied binomial trees. Journal of Finance 49 (3), 771–818.
Shimko, D.C., 1993. Bounds of probability. Risk 6, 33–37.
Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, New York.