2010

Understanding Analysts Forecasts

R. J. Louth
University of Cambridge

P. Joos

S. E. Satchell
University of Cambridge

Guy Weyns
Singapore Management University, guyweyns@smu.edu.sg

DOI: https://doi.org/10.1080/13518470902853582

Understanding Analysts Forecasts

R. J. Louth, Faculty of Economics, University of Cambridge, Sidgwick Ave, Cambridge CB3 9DD, UK;

P. Joos, Morgan Stanley Asia Ltd, International Commerce Centre, 1 Austin Road West, Kowloon, Hong Kong, China;

S. E. Satchell, Faculty of Economics, University of Cambridge, Sidgwick Ave, Cambridge CB3 9DD, UK;

G. Weyns, Morgan Stanley & Co. International plc., 20 Bank Street, Canary Wharf, London, E14 4AD, UK

Final version of article published in European Journal of Finance, 2010, 16 (2), 97-118.
DOI:10.1080/13518470902853582

Abstract

The purpose of this paper is to model analysts' forecasts. The paper differs from the previous research in that we do not focus on how accurate these predictions may be. Accuracy may indeed be an important quality but we argue instead that another equally important aspect of the analysts' job is to predict and describe the impact of jump events. In effect, the analysts' role is one of scenario prediction. Using a Bayesian-*inspired* generalised method of moments estimation procedure, we use this notion of scenario prediction combined with the structure of the Morgan Stanley analysts' forecasting database to model normal (base), optimistic (bull) and pessimistic (bear) forecast scenarios for a set of reports from Asia (excluding Japan) for 2007–2008. Since the estimation procedure is unique to this paper, a rigorous derivation of the asymptotic properties of the resulting estimator is also provided.

Keywords: analysts' reports, price forecasts, scenario prediction, jump diffusions, risk management

1. Introduction

Understanding the activities of financial analysts has long occupied the minds of both academics and practitioners alike. Nevertheless, the vast majority of this research effort has focussed on the predictive accuracy and efficacy of analysts' reports. Instead, we argue in favour of an alternative interpretation of the role played by analysts; namely, to predict and describe the impact of jump events. The primary focus of this paper is, therefore, to produce a model of analysts' forecasts for normal (base), optimistic (bull), and pessimistic (bear) forecast scenarios using a database of forecasts collected by Morgan Stanley (MS) analysts in Asia, excluding Japan, throughout 2007. Since each set of forecasts is accompanied by a detailed report, the database constitutes a welcome distillation of the often considerable amounts of information pertaining to the key drivers underlying future stock movements elicited by MS's analysts.

As we have already mentioned, the majority of past studies relating to analyst reports have focussed on the recommendation aspect of the report – buy, sell, or hold – and how such recommendations translate into stock price changes. Although a full literary survey is beyond the scope of this paper, the general conclusion from this literature, rather unsurprisingly, is that a positive stock price reaction generally follows an upgrade and a negative price reaction follows a downgrade. Less surprising, however, is the evidence relating to asymmetric responses, that is, there is now mounting evidence to suggest that downgrades are associated with much larger market reactions than upgrades. An excellent survey of this literature with accompanying references is provided by Kerl and Walter (2008) who find that aside from recommendations, earning forecasts and target price forecasts, the reputation of the analysts, and the detailed textual content of the report (when reduced to an appropriately coded variable) also influence market reactions. The idea to use the textual content of the report is an interesting one and something that we build on in this paper. Specifically, we take as our latent jump event the textual descriptions of factors that might increase or decrease the assets price; such an approach is facilitated by the particular structure of the MS analyst's reports which we describe next.

The MS database used in the empirical part of the paper contains data on the report publication date, the duration the report was 'active', and the forecasted bull, bear, and base prices. These prices represent forecasts by the analyst, which are conditioned on a number of events; it is helpful in what follows to think of these events as jump events. In some cases there are also written descriptions of the possible events that may lead to the occurrence of the bull or bear price. Presumably therefore, if none of these occurs, the base price is the relevant forecast. Thus, although we do not know whether a jump event – either good or bad – occurred over the course of the year for a particular stock and a particular analyst, it is often the case that a report that was currently 'active' is replaced. It is thus reasonable to suppose that this was the result of a jump occurring that had been observed by the analyst. In other cases, the active report at the start of the time period was also active at the end, indicating the absence of a jump event. To the extent that only some of the analysts report the forecasting horizon (which, when reported, is always 1 year), we take as the default forecasting horizon time a period of 1 year. Consequently, although the database has both temporal and cross-sectional dimensions, the fact that the majority of the reports are overlapping and the maximum time difference between report issues is only 12 months motivates our decision to treat all forecasted prices as realisations from an underlying time-homogeneous process when conducting the empirical analysis; ergo, we will proceed with the estimation under the assumption that the data only have a cross-sectional dimension.

In order to maintain internal consistency between the reporting and prediction methods used by MS analysts, Weyns, Perez, and Jenkins (2007) emphasise a number of key principles on which the forecasting database is built. Chief among them is the idea that analysts should think probabilistically about the ranges of uncertainty related to underlying jump events; the authors believe that single-point estimates only obscure valuable insights and give a false impression of precision. Moreover, rather than trying to communicate views on *all* of the key drivers and possible scenarios that may occur over the forthcoming year, analysts , 5are encouraged to seek transparency and avoid unnecessary complexity by only forming probabilistic views over a manageable number of scenarios; 'In most cases, we have found that three scenarios are plenty' (Weyns, Perez, and Jenkins 2007, 5).[1] Thus, the fact that analysts already think in terms of probabilistic statements over small number of events provides suitable justification for both the parametric jump-diffusion approach (Section 2), and our method of Bayesian-*inspired* generalised method of moments (GMM) explained (Section 4) which was developed specifically for the purpose of restricting analysts' prior beliefs to small number of discrete jump events.

However, before we explain the details of our empirical methodology, it is useful to illustrate the level of textual detail captured by the MS analysts' forecasting database. The following is a short extract from a randomly selected recent MS report on *China Infrastructure Machinery Holdings* (2007),

> 'Risk to Our Price Target: Downside risks to our price target include higher-than-expected steel prices, a significant FAI slowdown and tougher industry competition. Upside risks include strategic alliances with global names, M&A opportunities, faster-than-expected volume growth and lower-than-expected steel prices. The company announced on 16 January 2007 that it was in preliminary discussions with an independent third part on a proposal, which may lead to an offer for part of the listed securities of the company. We believe CIMH is likely to pursue M&A opportunities to further consolidate the market.' (*China Infrastructure Machinery Holdings*, 12 February 2007, 10).

Even from this short extract, the analyst's role in identifying circumstances that might lead to a dramatic improvement or fall in the stock price is immediately apparent; and this is, indeed, the case for the remaining reports in the database. It seems reasonable, therefore, to make the assumption in what follows that the analyst is an informed agent about the magnitude and number of jumps and reporting these constitutes an important part of her job.

2. The jump-diffusion model

In order to model the 'scenario' nature of the price forecasts, we now present a technical development of Lévy processes which are especially well suited to capture jump events in asset returns (Cont and Tankov 2004; Tankov 2007). At their most basic level, all Lévy processes can be decomposed into four components: a drift term, a diffusive part which is based on Brownian motion, a 'large jumps' component in the form of a compound Poisson process, and a 'small jumps' component in the form of a pure jump martingale. As we have already outlined above, however, it is infeasible for analysts to consider small jumps and they are constrained to predict only a finite number of jumps in any time period. Such constraints naturally shrink the class of Lévy pure jump processes that could underlie the analysts' predictions. In fact, it is well known that there is only one Lévy pure jump process that exhibits a finite number of jumps in a finite time interval: the compound Poisson process – all other pure jump processes exhibit an infinite number of small jumps in any finite time interval (Applebaum 2004).

Accordingly, we choose to model the latent prediction process by combining a drifting Brownian motion with two independent compound Poisson processes in order to capture the three essential elements of scenario prediction: namely, the base, bull, and bear price. It is, therefore, implicitly

[1] See Bunn and Salo (1993) for an early elucidation of the benefits to be gained from scenario prediction over forecasting, and the early popularity of scenario-based approaches to prediction within various business organisations.

assumed in what follows that the base process follows a geometric drifting Brownian motion that constitutes the prediction in the absence of jumps, while the bull and bear processes correspond to the addition of the relevant jump component to this base process. Formally, the prediction of the asset price at time t is given by $S_t = S_0 e^{X_t}$, where the Lévy process describing the log-return prediction, X_t, has the double-jump-diffusive specification in the spirit of Merton (1976),

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t + Y_t^{\text{Bull}} \mathrm{d}J_t^{\text{Bull}} + Y_t^{\text{Bear}} \mathrm{d}J_t^{\text{Bear}} \tag{1}$$

where W_t is a standard Brownian motion, $J^i = J^i(\lambda^i)$ denotes a Càdlàg Poisson process with stationary independent increments and intensity λ^i, and the log-jump size Y_t^i is a Gaussian random variable with mean α_i and variance δ_i^2 for $i = \{\text{Bull}, \text{Bear}\}$.[2] The solution to this stochastic differential equation is given by

$$z_\tau = X_{t+\tau} - X_t = \mu\tau + \sigma W_\tau + \int_0^\tau Y_s^{\text{Bull}} \mathrm{d}J_s^{\text{Bull}} + \int_0^\tau Y_s^{\text{Bear}} \mathrm{d}J_s^{\text{Bear}} \tag{2}$$

which implies that log-returns are independent and identically distributed – a consequence of the assumed independence of the Brownian and jump components.

Conditioning on the event of $J_\tau^{\text{Bull}} = n^{\text{Bull}}$ bull jumps and $J_\tau^{\text{Bear}} = n^{\text{Bear}}$ bear jumps, it follows that there must have been exactly n^{Bull} and n^{Bear} times, say τ_l, $l = 1, \ldots, n^{\text{Bear}}$, and τ_m, $m = 1, \ldots, n^{\text{Bull}}$, between t and $t+\tau$ such that $dJ_{\tau_l}^{\text{Bull}} = 1$ and $dJ_{\tau_m}^{\text{Bear}} = 1$, respectively. Thus, the jump component in Equation (2) can be written as

$$\int_0^\tau Y_s^{\text{Bull}} \mathrm{d}J_s^{\text{Bull}} + \int_0^\tau Y_s^{\text{Bear}} \mathrm{d}J_s^{\text{Bear}} = \sum_{l=1}^{n^{\text{Bull}}} Y_{\tau_l}^{\text{Bull}} + \sum_{m=1}^{n^{\text{Bear}}} Y_{\tau_m}^{\text{Bear}} \tag{3}$$

which is the sum of n^{Bull} and n^{Bear} independent jump terms with distributions $N(\alpha_{\text{Bull}}, \delta_{\text{Bull}}^2)$ and $N(\alpha_{\text{Bear}}, \delta_{\text{Bear}}^2)$, respectively. Hence, the transition density of $X_{t+\tau}$ given X_t is given by

$$\begin{aligned} p(x|x_t, \tau; \theta) = & \sum_{n^{\text{Bull}}=0}^{\infty} \sum_{n^{\text{Bear}}=0}^{\infty} p(x|x_t, \tau, J_\tau^{\text{Bull}} = n^{\text{Bull}}, J_\tau^{\text{Bear}} = n^{\text{Bear}}; \theta) \\ & \times p(J_\tau^{\text{Bull}} = n^{\text{Bull}}; \theta) p(J_\tau^{\text{Bear}} = n^{\text{Bear}}; \theta) \end{aligned} \tag{4}$$

where

$$\begin{aligned} & p(x|x_t, \tau, J_\tau^{\text{Bull}} = n^{\text{Bull}}, J_\tau^{\text{Bear}} = n^{\text{Bear}}; \theta) \\ & \quad = N(x_t + \mu\tau + n^{\text{Bull}}\alpha_{\text{Bull}} + n^{\text{Bear}}\alpha_{\text{Bear}}, \sigma^2\tau + n^{\text{Bull}}\delta_{\text{Bull}}^2 + n^{\text{Bear}}\delta_{\text{Bear}}^2) \end{aligned}$$

$$p(J_\tau^i = n^i; \theta) = \frac{\exp(-\lambda^i \tau)(\lambda^i \tau)^{n^i}}{n^i!}$$

[2] For computational simplicity, it is further assumed that W_t, J_t, and Y_t are independent.

for $i = \{\text{Bull}, \text{Bear}\}$. The functional form of the transition density will play a crucial role in delineating the risk management implications of our findings. However, before this can be accomplished, it remains to describe our method of Bayesian-*inspired* GMM estimation (Section 4) – a crucial prerequisite of which is the moments of the log-return predictions described by Equation (2). We conclude this section by illustrating the method by which they can be derived.

Although the law of the log-return prediction between time *t* and *t*+τ does not admit a closed form expression, the Lévy-Khintchine theorem can be applied to obtain the following analytical expression for the characteristic function of the jump-diffusion process followed by z_τ,

$$\phi_z(x) = \exp\left[ix\mu\tau - \frac{1}{2}x^2\sigma^2\tau + \lambda_{\text{Bull}}\left(\exp\left(ix\alpha_{\text{Bull}} - \frac{1}{2}x^2\delta_{\text{Bull}}^2\right) - 1\right)\tau + \lambda_{\text{Bear}}\left(\exp\left(ix\alpha_{\text{Bear}} - \frac{1}{2}x^2\delta_{\text{Bear}}^2\right) - 1\right)\tau\right] \quad (5)$$

It, therefore, follows that the raw moments of z_τ can be calculated according to

$$E(z_{t+\tau}^j) = \left.\frac{\phi_z^{[j]}(x)}{i^j}\right|_{x=0} \quad (6)$$

for $j = 1, \ldots, j_{\text{Max}}$, where $\phi_z^{[j]}$ denotes the *j*th-order derivative of ϕ_s with respect to *x*. The first three moments are thus given by

$$M[1, \tau] = E[z_\tau] = (\mu + \alpha_{\text{Bull}}\lambda_{\text{Bull}} + \alpha_{\text{Bear}}\lambda_{\text{Bear}})\tau \quad (7)$$

$$M[2, \tau] = E[(z_\tau - E[z_\tau])^2] = (\sigma^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^2 + \delta_{\text{Bull}}^2) + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^2 + \delta_{\text{Bear}}^2))\tau \quad (8)$$

$$M[3, \tau] = E[(z_\tau - E[z_\tau])^3] = \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^3 + 3\alpha_{\text{Bull}}\delta_{\text{Bull}}^2)\tau + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^3 + 3\alpha_{\text{Bear}}\delta_{\text{Bear}}^2)\tau \quad (9)$$

Expressions for the remaining central moments used in the empirical section of this paper are described in the Appendix.[3]

3. Identification issues

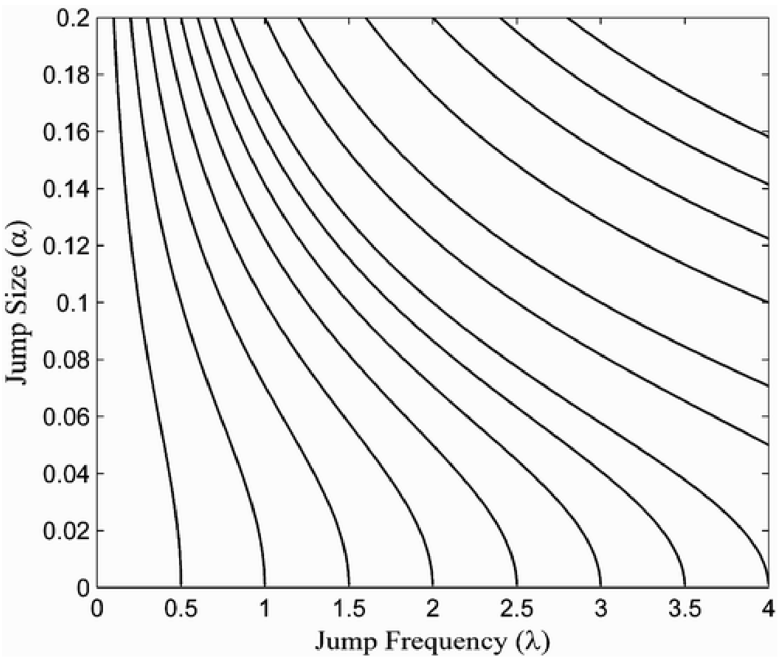
Before outlining our method for estimating the parameters of the double-jump-diffusion process, we first consider the issue of weak identification which motivates our Bayesian-*inspired* GMM approach to estimation. Although each parameter is separately identified within our model, it can nevertheless be the case that several combinations of the parameters correspond to approximately the same jump-diffusion process. The result is that the parameters are often imprecisely estimated. This issue was first discussed by Aït-Sahalia (2004) in the context of disentangling the jump component from the Brownian component when estimating these types of Lévy processes. In particular, he showed that as the sampling frequency tends towards infinity, even infinite activity jump and Brownian components can be perfectly disentangled. In the present context, however, the low-frequency nature of analysts'

[3] See Press (1967) for an early derivation of these moments using the transition density function in the case of one compound Poisson process.

reports means that we cannot appeal to these asymptotic results and so the issue of weak identification remains.

To illustrate why this problem arises, we consider a simple moment-matching example analogous to that presented in Aït-Sahalia (2004) and which is designed to show that different combinations of the parameters can satisfy the same moment conditions. For instance, Figure 1 plots the combinations of jump intensity and mean jump size that result in the same observable conditional variance of log-returns, $E[(z_\tau - E[z_\tau])^2] = \text{constant}$, with all other parameters held fixed. Intuitively, any two combinations of parameters on the same curve cannot be distinguished by the method of moments using that moment condition. Thus, we see that there is a clear trade-off between concluding that there are a high frequency of small jumps and a low frequency of large jumps.

Figure 1. Trade-off between jump frequency (λ) and mean jump size (α).



The overwhelming conclusion from this simple example is, therefore, that one must exploit a range of moments in order to identify the parameters.[4] However, the fact remains that with a finite sample there can still be several combinations of the model parameters that correspond to approximately the same jump-diffusion process. In other words, large changes in the parameter values can result in small changes in the moment function, thereby rending the parameter estimates highly imprecise. In this situation, one would like to improve the precision of the estimates by fixing one of the parameters; but which parameter should we fix? And at what value should it be assigned?

Our proposed answer to these questions derives from empirical Bayesianism. Given the richness of our data set and the constraints on analysts' reporting behaviour, we believe that we have enough prior information at our disposal to facilitate the separate estimation of the prior distribution of the jump frequency parameter. Thus, by estimating the hyper-parameters of this prior distribution we are effectively assigning fixed values to the jump frequency parameter, which in turn provides the

[4] In fact, Carrasco et al. (2007) propose a GMM estimator that uses a continuum of moment conditions implied by the characteristic function.

additional numerical stability that we desire. To illustrate this idea in more detail, we make a temporary detour from our GMM setting and consider the use of maximum likelihood in the presence of complete non-identification – a much stronger problem than we are faced with.

Let $y = \{y_i\}_{i=1}^{N}$ denote a set of N draws from a Gaussian distribution with unknown mean, $\mu_1+\mu_2$, and known variance, σ^2. It is obvious that μ_1 and μ_2 are not separately identified and so conventional maximum likelihood techniques are not directly applicable. However, placing a suitable prior distribution on μ_2 allows us to identify μ_1. To see how this can be achieved, we first of all note that the joint density of μ_1 and μ_2 can be written as

$$p(\mu_1, \mu_2|y, \sigma^2) \propto p(\mu_1, \mu_2) l(\mu_1, \mu_2|y, \sigma^2) \tag{10}$$

where we assume that $p(\mu_1, \mu_2) \propto p(\mu_2)^d = N(\tilde{\mu}_2, N^{-\alpha}\tilde{\sigma}_2^2)$ with the hyper-parameters, $\tilde{\mu}_2$ and $\tilde{\sigma}_2^2$, known, $\tilde{\mu}_2 - \mu_2 = o(N^{\beta})$, and $\alpha, \beta > 0$. The existence of such a prior distribution implies that, given an infinite amount of data, the distribution of our prior beliefs converge to a degenerate distribution centred on the true value of the parameter.[5]

Conditional on the existence of this prior distribution, our ultimate objective is to identify μ_1, which is equivalent to finding

$$p(\mu_1|y, \sigma^2) = \int_{-\infty}^{\infty} p(\mu_1, \mu_2|y, \sigma^2) \mathrm{d}\mu_2 \tag{11}$$

Substituting the functional form of the likelihood and prior density, we obtain

$$\begin{aligned} p(\mu_1|y, \sigma^2) &\propto \int_{-\infty}^{\infty} p(\mu_2) l(\mu_1, \mu_2|y, \sigma^2) \mathrm{d}\mu_2 \\ &\propto \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{N/2}} \frac{N^{-\alpha/2}}{(\tilde{\sigma}_2^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mu_1 - \mu_2)^2 \right. \\ &\qquad \left. - \frac{1}{2N^{-\alpha}\tilde{\sigma}_2^2}(\mu_2 - \tilde{\mu}_2)^2\right] \mathrm{d}\mu_2 \\ &\propto \int_{-\infty}^{\infty} \exp\left[-\frac{N}{2\sigma^2}(\mu_2 - (\bar{y} - \mu_1))^2 - \frac{N^{\alpha}}{2\tilde{\sigma}_2^2}(\mu_2 - \tilde{\mu}_2)^2\right] \mathrm{d}\mu_2 \end{aligned}$$

where $\bar{y} = N^{-1}\sum_{i=1}^{N} y_i$. Furthermore, by completing the square we arrive at the following expression

$$p(\mu_1|y, \sigma^2) \propto \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\delta^2}(\mu_2 - \gamma)^2 - \frac{1}{2(N^{-1}\sigma^2 + N^{-\alpha}\tilde{\sigma}_2^2)}(\mu_1 - (\bar{y} - \tilde{\mu}_2))^2\right] \mathrm{d}\mu_2$$

where

[5] From our empirical Bayesian perspective where the hyper-parameters are estimated from observed data, it is likely that we will become more confident in the prior belief the more data we have at our disposal and so the assumption that a prior distribution exists with these properties is not overly optimistic.

$$\delta^2 = \left[\frac{N}{\sigma^2} + \frac{N^\alpha}{\tilde{\sigma}_2^2}\right]^{-1}$$

$$\gamma = -2\delta^4 \left[\frac{N(\bar{y} - \mu_1)}{\sigma^2} + \frac{\tilde{\mu}_2}{N^{-\alpha}\tilde{\sigma}_2^2}\right]$$

Hence, defining, $z = (\mu_2 - \gamma)/\delta$, it follows that $dz \propto \delta^{-1} d\mu_2$ and thus,

$$p(\mu_1|y, \sigma^2) \propto \exp\left[-\frac{1}{2(N^{-1}\sigma^2 + N^{-\alpha}\tilde{\sigma}_2^2)}(\mu_1 - (\bar{y} - \tilde{\mu}_2))^2\right] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}z^2\right] dz$$

$$\propto \exp\left[-\frac{1}{2(N^{-1}\sigma^2 + N^{-\alpha}\tilde{\sigma}_2^2)}(\mu_1 - (\bar{y} - \tilde{\mu}_2))^2\right]$$

which is clearly a Gaussian kernel, and so

$$p(\mu_1|y, \sigma^2) \sim N\left(\bar{y} - \tilde{\mu}_2, \frac{\tilde{\sigma}_2^2}{N^\alpha} + \frac{\sigma^2}{N}\right) \quad (12)$$

Hence, through our prior knowledge of the distribution of μ_2 we can identify μ_1.
Although this example is highly stylised and its conclusions are too strong for present purposes, it nevertheless conveys the important idea that through the appropriate use of a prior distribution one can circumvent the problems associated with unidentified or weakly identified parameters. However, it also serves to illustrate the importance of selecting an appropriate prior distribution: if, for instance, the hyper-parameters were 'inaccurately' chosen, e.g. $\tilde{\mu}_2 - \mu_2 - c = o(N^\beta)$, then the distribution of μ_1 would not be centred correctly and so the estimator would be inconsistent. Since the choice of prior distribution is critical to our analysis, Section 5 will discuss this issue in greater detail.

4. GMM estimation

To motivate the choice of GMM estimation, we note that the class of exponential Lévy processes implies that asset prices derive from a mixture of *N* Gaussian distributions, where *N* goes to infinity. In these cases, results from the mixture-of-distributions literature show that the likelihood function can become unbounded and thus maximum likelihood estimation is infeasible (Kiefer 1978; Honoré 1998). To remedy this, Honoré (1998) suggests a concentrated maximum likelihood estimator which restricts the volatility parameters in the jump diffusion to be in a compact set which includes the true values. Specifically, he proposes a re-parametrised model whereby the baseline volatility parameter and jump volatility parameter are linked via $\sigma^2 = m\delta^2$, where *m* is a positive constant that fixes the relative sizes of the volatilities. The estimation methodology proceeds by fixing *m* and calculating the maximised likelihood; a procedure which is repeated for many values of $m \in M$. The parameter estimates are those corresponding to the m^* that maximises the profile likelihoods.

In spite of the elegance of this approach, the profile likelihood is not a true likelihood as it is not based directly on a probability distribution and this can lead to some unsatisfactory properties, including a loss of efficiency.[6] In contrast, we have seen that the characteristic function is available in analytical form and so GMM methods can be utilised. Moreover, the asymptotic properties of this estimator are rigorously proved in the Appendix and so, although we lose some efficiency when compared with the

[6] Attempts have been made to improve this, resulting in the modified profile likelihood (Barndorff-Nielsen 1988).

full information maximum likelihood, we prefer this methodology to the profile likelihood approach of Honoré (1998). In spite of this preference, however, there nevertheless remain obvious parallels between both methods arising from the fact that both 'fix' a parameter value in order to improve the stability of parameter estimation.

To illustrate the methodology underlying the Bayesian-*inspired* GMM procedure, we first of all note that our assumptions about the data set imply that our data are essentially base, bull, and bear realisations from the underlying time-homogeneous stochastic process given by Equation (1). Accordingly, we let $z = (z'_{\text{Base}}, z'_{\text{Bull}}, z'_{\text{Bear}})$ denote the concatenation of a single realisation of this log-return prediction data, and let $h(\theta, Z)$ denote the $3 \times j_{\text{Max}}$ vector-valued function with the typical element $(z_i - E[z_i]1[j > 1])^j - M_i[j, \tau]$ $i = \{\text{Base, Bull, Bear}\}$ and $j = 1, \ldots, j_{\text{Max}}$, where the expectation is with respect to the distribution of z, and $1[j>1]$ is an indicator function taking value of unity if $j > 1$ and zero otherwise.[7]

Note that the relevant moment conditions for the base, bull, and bear predictions, $M_i[j, \tau]$, will correspond to the coefficient restrictions $\lambda_{\text{Bull}} = \lambda_{\text{Bear}} = 0$, $\lambda_{\text{Bear}} = 0$, and $\lambda_{\text{Bull}} = 0$ in Equations (A2)–(A7). In all of the other cases we let the supports of the two jump intensity parameters, λ_{Bear} and λ_{Bull}, be the set of integers between zero and four (inclusive). Our decision to restrict the supports of these parameters is a consequence of the nature of the reporting process and the textual content of the analyst reports themselves – a supposition which is confirmed by the empirical evidence contained in Tables 1–3. Thus, in order to incorporate these constraints on prior beliefs into the methodology, we introduce a series of prior probabilities $\{\pi_w\}_{w=1}^{W} \geq 0$ with $\sum_{w=1}^{W} \pi_w = 1$ which attach weights to each of the combinations of $\{\lambda_{\text{Bear}}, \lambda_{\text{Bull}}\} \in \{0, \ldots, 4\}^2$ over which the analyst might have beliefs over. Hence, letting θ_0 denote the true value of
$\theta = (\mu, \sigma^2, \alpha_{\text{Bull}}, \delta^2_{\text{Bull}}, \alpha_{\text{Bear}}, \delta^2_{\text{Bear}})$
we can utilise the orthogonality conditions implied by

$$E\left[\sum_{w=1}^{W} \pi_w h_w(\theta_0)\right] = 0 \qquad (13)$$

to estimate θ_0, where, once again, the expectation is with respect to the distribution of z.

Table 1. Total re-issues.

[7] For notational convenience, we suppress the time subscript of the log-return, i.e. $z \equiv z_\tau$.

	1	2	3	4	5	6	7	8	9	10	11	Average	No. of re-issues	Total reports
CDI	19	10	4	5	1	1	0	0	0	0	0	2.05	82	164
CS	5	5	4	1	0	0	0	0	0	0	0	2.07	31	57
EU	19	8	7	1	0	0	0	1	0	0	0	1.89	68	128
F	16	2	11	0	3	0	1	0	0	0	0	2.27	75	164
Mat	16	9	6	0	2	0	0	0	0	0	0	1.88	58	128
Med	7	2	3	1	0	0	0	0	0	0	0	1.85	24	45
P	7	5	0	0	0	0	0	0	0	0	0	1.42	17	40
R	9	3	0	0	0	0	0	0	0	0	0	1.25	15	41
Tech	29	17	20	3	5	1	1	0	0	0	1	2.39	179	301
TS	8	3	4	1	4	1	1	0	1	0	0	3.13	72	103
Tr	10	13	6	1	3	3	3	0	0	1	0	3.05	112	167
As	7	5	3	0	1	0	0	0	0	0	0	1.93	28	63
Au	17	4	5	0	4	0	1	0	1	0	0	2.38	75	156
Ch	31	21	27	3	0	2	0	0	0	1	0	2.21	187	343
HK	10	5	3	0	0	0	1	0	0	0	0	1.89	36	73
In	23	5	6	1	2	1	0	0	0	0	0	1.87	71	159
SK	28	11	11	7	6	3	2	1	1	0	1	2.83	199	300
Sin	5	8	1	0	1	1	1	0	0	0	0	2.47	35	63
Tai	14	18	10	1	4	1	1	0	0	0	0	2.39	113	198
Whole	135	78	66	12	18	8	6	1	2	1	1	2.33	744	1355

Note: This table describes the total frequencies of report re-issues for consumer discretionary industrials (CDI), consumer staples (CS), energy utilities (EU), financials (F), materials (Mat), media (Med), Property (P), retail (R), technology (Tech), telecom services (TS), transportation (Tr), Asean (As), Australia (Au), China (Ch), Hong Kong (HK), India (In), South Korea (SK), Singapore (Sin), Taiwan (Tai).

Table 2. Bull re-issues.

	1	2	3	4	5	6	7	8	Average	% Bull
CDI	17	10	2	3	0	0	0	0	1.72	69.62
CS	6	3	1	0	0	0	0	0	1.50	62.50
EU	20	5	4	0	0	0	0	1	1.67	78.13
F	15	5	5	0	0	1	0	0	1.77	66.67
Mat	20	10	1	1	1	0	0	0	1.58	82.54
Med	7	1	1	0	0	0	0	0	1.33	50.00
P	4	0	0	0	0	0	0	0	1	26.67
R	5	1	0	0	0	0	0	0	1.17	63.64
Tech	34	13	7	2	0	1	0	1	1.78	58.52
TS	7	4	3	3	2	0	0	0	2.42	63.89
Tr	14	14	9	1	2	1	0	0	2.17	71.77
As	8	1	0	0	0	0	0	0	1.11	43.48
Au	13	4	1	1	0	1	0	0	1.70	47.89
Ch	39	21	14	0	2	1	0	0	1.80	75.96
HK	5	3	2	0	0	0	0	0	1.70	48.57
In	21	5	1	2	2	0	0	0	1.68	76.47
SK	26	16	10	5	2	1	0	3	2.35	75.13
Sin	4	4	2	1	0	0	0	0	2.00	56.41
Tai	26	12	3	1	0	0	0	0	1.50	58.33
Whole	143	68	35	10	6	3	0	3	1.85	67.30

Note: This table describes the frequencies of bull report re-issues. The acronyms are described in Table 1.

Table 3. Bear re-issues.

	1	2	3	4	5	Average	% Bear
CDI	20	2	0	0	0	1.09	30.38
CS	5	2	0	0	0	1.29	37.50
EU	14	0	0	0	0	1.00	21.88
F	8	4	1	1	0	1.64	33.33
Mat	9	1	0	0	0	1.10	17.46
Med	5	2	1	0	0	1.50	50.00
P	3	4	0	0	0	1.57	73.33
R	2	1	0	0	0	1.33	36.36
Tech	31	15	4	0	0	1.46	41.48
TS	9	2	3	1	0	1.73	36.11
Tr	16	6	1	1	0	1.49	28.23
As	5	2	0	1	0	1.63	56.52
Au	12	4	3	2	0	1.76	52.11
Ch	25	4	2	0	1	1.38	24.04
HK	9	3	1	0	0	1.38	51.43
In	8	4	0	0	0	1.33	23.53
SK	23	7	4	0	0	1.44	24.87
Sin	7	5	0	0	0	1.42	43.59
Tai	23	11	0	0	0	1.32	41.67
Whole	112	41	10	3	1	1.44	32.70

Note: This table describes the frequencies of bear report re-issues. The acronyms are described in Table 1.

To the extent that this Bayesian-*inspired* GMM approach is a novel addition to the existing GMM literature, we now provide a detailed explanation of the two-step estimation algorithm and the necessary mathematical justifications for the consistency and asymptotic normality of the estimator. Letting $\Im_N = \{z_{\text{Base},n}, z_{\text{Bull},n}, z_{\text{Bear},n}\}_{n=1}^{N}$ denote N realisations from the log-return prediction process and letting $g(\theta, z)$ denote the vector of orthogonality conditions weighted by the prior probabilities, i.e. $g(\theta, z) = \sum_{w=1}^{W} \pi_w h_w(\theta, z)$, we can define the sample orthogonality conditions as

$$\hat{g}(\theta, \Im_N) = \frac{1}{N}\sum_{n=1}^{N} g(\theta, z_n) \tag{14}$$

where the expectation is now with respect to the empirical distribution function. Hence, an estimator of θ_0 can thus be obtained by minimising the modified GMM criterion function

$$\hat{\theta}_N = \arg\min_{\theta} Q_N(\theta) = \hat{g}(\theta, \Im_N)' W_N \hat{g}(\theta, \Im_N) \tag{15}$$

where W_N is an appropriately defined positive semi-definite weighting matrix with $W_N \xrightarrow{p} W$ and W positive definite, and $\sum_{w=1}^{W} \pi_w = 1$. More specifically, we use the standard two-step GMM algorithm based on the following procedure.

1. Maximise $Q_N(\theta)$ using $W_N = I_{3\times j\text{Max}}$, and obtain $\tilde{\theta}$,
2. Calculate $\hat{\Omega}_N(\tilde{\theta}) = \left[\frac{1}{N}\sum_{n=1}^{N} g(\tilde{\theta}, z_n) g(\tilde{\theta}, z_n)'\right]$,
3. Maximise $Q_N(\theta)$ using $W_N = \hat{\Omega}_N(\tilde{\theta})^{-1}$ and obtain ˆ θ.

Under the following regularity conditions, we can show that this Bayesian-*inspired* two-step GMM estimator is consistent and asymptotically normal.

Assumption 1

(*a*) $\theta_0 \in \Theta$ *is the unique solution to* $E[g(\theta, z)] = 0$, (*b*) Θ *is compact*, (*c*) $g(\theta, z)$ *is continuous at each* $\theta \in \Theta$ *with probability* 1, (*d*) $E[\sup_{\theta\in\Theta} \|g(\theta, z)\|^2] < \infty$, (*e*) $\Omega = E[g(\theta_0, z) g(\theta_0, z)']$ *is positive definite*.

Assumption 2

(*a*) $\theta_0 \in \text{int}(\Theta)$; (*b*) $g(\theta, z)$ *is continuously differentiable on a neighbourhood* $\mathbb{N}$ *of* θ_0, $E[\sup_{\theta\in\mathbb{N}} \|\partial g(\theta, z)/\partial\theta'\|^2] < \infty$; (*c*) $E[\sup_{\theta\in\mathbb{N}} \|\partial g(\theta_0, z)/\partial\theta'\|^2]$ *is full column rank.*

Theorem (consistency) *Under* Assumption 1, $\hat{\theta} \xrightarrow{p} \theta_0$.

Proof See Appendix Section A2.

Theorem (asymptotic normality) Under Assumptions 1 and 2, it follows that

$$N^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1}) \tag{16}$$

where

$$G = E\left[\sum_{w=1}^{W} \pi_w \frac{\partial h(\theta_0, z)}{\partial \theta'}\right]$$

$$\Omega = E\left[\sum_{w=1}^{W} \pi_w h(\theta_0, Z) \sum_{w=1}^{W} \pi_w h(\theta_0, z)'\right]$$

Proof See Appendix Section A2.

Although the algorithm could be iterated, the estimates based on a single iteration have the same asymptotic distribution as those based on an arbitrarily large number of iterations (Newey and McFadden 1994).

5. Informative priors

To operationalise our Bayesian-*inspired* GMM approach, it remains to define the method of setting the prior probabilities $\{\pi_w\}_{w=1}^{W} \geq 0$which attach weights to each of the combinations of bull and bear jumps, $\{\lambda_{\text{Bear}}, \lambda_{\text{Bull}}\} \in \{0, \ldots, 4\}^2$, for each of the groups under consideration. To do so, we adopt an empirical Bayesian stance whereby the probabilities are partly based on the known number of re-issues and partly based on the report statements such as the one provided in Section 1 that lists up to four specific events.[8]

Formally, we estimate the intensity hyper-parameter λ from a Poisson distribution for the total number of re-issue frequencies, *x*, for each group in the data set using the data contained in Table 1. From these estimates we then calculate the probabilities associated with the total number of re-issues, *x*, using the Poisson density

$$p(x) = \frac{\sum_{x=0}^{\infty} e^{-\lambda} \lambda^x}{x!} \tag{17}$$

which implies that

$$\hat{p}_0 = \hat{p}(x = 0) = e^{-\hat{\lambda}} \tag{18}$$

$$\hat{p}_1 = \hat{p}(x = 1) = e^{-\hat{\lambda}}(1 + \hat{\lambda}) \tag{19}$$

and so forth. Thus, under the assumption that the numbers of re-issues due to bull and bear events are independent, the prior probabilities for *j* bull and *k* bear jumps can be calculated according to

$$\hat{\pi}_{j,k} = \frac{\hat{p}_{j+k}}{j + k + 1} \tag{20}$$

for$j, k = 0, \ldots, 4$.

An alternative approach would be to estimate different intensity hyper-parameters for the bull and bear distributions using the information contained in Tables 2 and 3 rather than assuming a common intensity hyper-parameter, ˆ λ, for both. We choose not to pursue this approach for the simple reason

[8] See Weyns, Perez, and Jenkins (2007) for further evidence on the reporting process and suggested constraints on the number of jump events reported.

that there are a significant number of zeros in the data set corresponding to no report re-issue, and thus it is not possible to fit a Poisson distribution to both bull and bear intensity parameters separately without making some arbitrary assumptions about how to deal with the zeros beforehand. One method of circumventing this problem would be to fit a zero-truncated Poisson distribution to the bull and bear re-issue data. Unfortunately, however, this leads to the added complication that for some groups the maximum number of report re-issues per firm is 1 and so fitting a zero-truncated Poisson distribution to the data is not possible. Furthermore, the number of re-issue observations is small for some groups, meaning that estimation of the intensity parameter is likely to be imprecise and highly sensitive to outliers.[9]

6. Results

Due to data limitations arising from the fact that there are 102 analysts and we only have approximately 13 reports per analyst in the data set, we prefer to assume a collective analyst whose forecasts represent the forecast of all the analysts within the particular group (i.e. country or sector). Thus, we are implicitly assuming that the data constitute multiple realisations from the relevant underlying base, bull, or base stochastic process for the particular group under consideration. It is possible to weaken this assumption by taking into account different base models for individual stocks and identifying the common jumps with respect to the particular stock returns. However, this requires that we make some arbitrary decisions about what part of individual stock return represents base and what represents jumps, and so we prefer our group-based interpretation. All results are obtained using Asian Open Platform 1 data from the year 2007, which yields a total of 1355 cross-sectional observations of which 760 are uncensored; the remaining 595 observations contain no 'report expiry' information, i.e. they are still 'active' at the end of the sample period.

In view of the fact that the well-known *J*-test of over-identifying restrictions over-rejects in small samples (see Hall and Horowitz (1996) and Hall (2000) for a discussion), we propose an alternative method of selecting the appropriate number of moment conditions based on K-fold cross-validation.[10] This approach is based on a comparison of the out-of-sample prediction abilities of the jump-diffusion specification estimated using different moment conditions: the final choice of moment conditions being the set that has the best predictive ability. Intuitively, the fewer the moment conditions that are required, the more Gaussian are the data and thus the less information and the more noise are added by employing more moment conditions.

To implement this procedure, we let $\kappa : \{1, \ldots, N\} \mapsto \{1, \ldots, K\}$ denote an indexing function that indicates to which of the $K \leq N$ partitions observation n is allocated by randomisation, and let $\hat{Z}^{i(-k)}_{t+\tau}$ denote the simulated log-return between t and $t+\tau$ for $i = \{\text{Base, Bull, Bear}\}$ with the kth part of the data removed. The cross-validation process is then repeated $K-1$ times, with each of the k subsamples used exactly once as the validation data. The quadratic loss cross-validation estimate of the prediction error is then calculated according to

$$\text{CV} = \frac{1}{3N} \sum_{n=1}^{N} \sum_{i} (Z^{i}_{t+\tau} - \hat{Z}^{i(-k)}_{t+\tau})^2 \qquad (21)$$

where $i = \{\text{Base, Bull, Bear}\}$. Since we have 18 moment conditions in total (i.e. three moment conditions for each of the six moments), we use the 10-fold cross-validation prediction error to select

[9] Results for the groups whose maximum number of report re-issues is greater than 1 are available from the authors upon request

[10] See Hastie, Tibshirani, and Friedman (2001) for a general overview of cross-validation.

the appropriate number of moments to use. In particular, we define a tuning parameter, $m \in \{1, \ldots, 13\}$, which indexes the number of moment conditions that are used: corresponding to the use of the first 5+*m* elements of the vector $g(\theta, z)$ defined in Section 4. That is, we calculate for each $m \in \{1, \ldots, 13\}$,

$$\mathrm{CV}(m) = \frac{1}{3N} \sum_{n=1}^{N} \sum_{i} (Z^{i}_{t+\tau} - \hat{Z}^{i(-k)}_{t+\tau}(m))^2 \quad (22)$$

where $\hat{Z}^{i(-k)}_{t+\tau}(m)$ is the simulated log-return between *t* and *t*+τ for $i = \{\mathrm{Base}, \mathrm{Bull}, \mathrm{Bear}\}$ with the *k*th part of the data removed and using the first 5+*m* elements of the vector $h(\theta, \tau, Z)$ as moment conditions. The function CV(*m*) provides an estimate of the test error curve for each point *m*, and so we select the tuning parameter, i.e. the optimum number of moments, 5+*m*, at which the curve attains its minimum. The final chosen number of moment conditions is then fit to all of the data.

Under our simplifying assumption of a single sector/country analyst, the estimates from the 11 sectors and eight countries in the data set are reported in Tables 4 and 5, respectively, using the two-step GMM algorithm outlined in Section 4 and the method of forming a prior density for the intensity parameters outlined in Section 5.

Table 4. Sector Results

	CDI	CS	EU	F	Mat	Med	P	R	Tech	TS	Tr
μ	6.05	5.62	3.10	5.70	2.53	6.52	1.45	4.46	6.59	4.80	7.28
	(0.73)	(0.70)	(0.78)	(0.37)	(0.72)	(0.97)	(0.81)	(0.49)	(0.46)	(0.78)	(0.43)
σ	4.30	3.05	6.13	2.05	6.26	10.52	8.32	3.25	4.11	8.18	5.55
	(4.43)	(3.76)	(4.66)	(2.00)	(3.85)	(4.65)	(3.97)	(3.43)	(3.47)	(3.86)	(3.08)
α_{Bull}	41.11	41.53	36.46	29.00	37.88	41.95	25.59	43.59	37.65	26.31	25.74
	(1.99)	(2.41)	(1.80)	(0.47)	(2.10)	(0.97)	(1.45)	(4.88)	(0.64)	(0.46)	(0.61)
δ_{Bull}	0.00	0.00	0.00	5.25	0.00	3.21	8.33	0.00	0.00	4.38	6.24
	(8.71)	(10.11)	(9.40)	(2.76)	(7.40)	(5.96)	(5.77)	(12.51)	(6.78)	(2.78)	(3.92)
α_{Bear}	−36.76	−37.14	−31.92	−27.96	−37.28	−39.60	−24.38	−37.37	−32.03	−21.32	−24.90
	(1.72)	(2.35)	(1.66)	(0.52)	(2.11)	(0.80)	(2.08)	(8.29)	(0.41)	(0.89)	(0.68)
δ_{Bear}	0.00	7.18	3.41	5.88	0.00	0.00	10.64	0.00	0.00	6.74	0.00
	(7.64)	(6.98)	(7.10)	(2.78)	(8.45)	(6.98)	(7.90)	(22.71)	(4.60)	(4.17)	(4.87)
N	164	57	128	164	128	45	40	41	301	103	167
λ	0.67	0.76	0.71	0.61	0.60	0.71	0.49	0.39	0.92	1.33	1.28
	(0.07)	(0.14)	(0.09)	(0.07)	(0.08)	(0.14)	(0.12)	(0.10)	(0.07)	(0.16)	(0.12)
m	4	4	4	11	4	9	7	3	7	8	9
Mean	7.51	7.27	4.71	6.02	2.71	7.35	1.74	9.17	9.17	8.08	7.81
STD	32.23	34.63	29.53	22.84	29.80	35.87	20.42	33.78	33.78	29.37	29.47
Skew	0.20	0.13	0.20	0.05	0.03	0.10	−0.02	0.24	0.24	0.18	0.11
Kurt	4.46	4.39	4.34	4.83	4.52	4.21	5.11	4.08	4.08	3.79	3.83

Note: This table reports parameter estimates for each of the 11 sectors: consumer discretionary industrials (CDI), consumer staples (CS), energy utilities (EU), financials (F), materials (Mat), media (Med), Property (P), retail (R), technology (Tech), telecom services (TS), and transportation (Tr) using the GMM procedure outlined in Section 4 with *N* observations. For ease of interpretation, the parameters have been converted into 'percentage' form (i.e. ×100). Standard errors are reported in parentheses, 'λ' denotes the estimate of the intensity hyper-parameter, '*m*' denotes the optimal tuning parameter correspond to the use of '5+*m*' moments. Finally, the final four rows denote the first four 'posterior' moments of the resulting double-jump diffusion taking into account the prior density for each of the jump intensity parameters.

Table 5. Country results

	Asean	Australia	China	Hong Kong	India	South Korea	Singapore	Taiwan
μ	12.66	1.86	4.49	6.91	2.69	8.36	4.75	7.62
	(0.50)	(0.48)	(0.38)	(0.63)	(0.25)	(0.28)	(0.41)	(0.37)
σ	14.31	4.00	5.36	4.44	2.39	4.14	1.66	3.03
	(1.95)	(2.84)	(2.57)	(3.25)	(1.83)	(2.41)	(2.01)	(2.36)
α_{Bull}	36.63	27.75	29.94	32.13	36.90	31.78	29.40	37.11
	(0.37)	(0.80)	(0.59)	(0.92)	(0.66)	(0.41)	(0.52)	(0.23)
δ_{Bull}	0.00	0.00	6.15	0.00	6.94	0.00	4.58	0.00
	(3.33)	(4.65)	(4.37)	(6.93)	(4.73)	(4.71)	(3.24)	(2.79)
α_{Bear}	−32.47	−27.38	−29.94	−32.36	−36.99	−29.67	−26.01	−29.50
	(0.60)	(1.17)	(0.49)	(2.20)	(0.39)	(0.28)	(0.65)	(0.37)
δ_{Bear}	4.97	10.03	7.76	0.00	5.22	0.00	5.37	5.24
	(3.26)	(6.65)	(3.16)	(7.58)	(2.72)	(3.54)	(3.47)	(2.60)
N	63	156	343	73	159	300	63	198
λ	0.61	0.67	0.78	0.64	0.56	1.17	0.93	0.87
	(0.11)	(0.08)	(0.06)	(0.11)	(0.07)	(0.08)	(0.14)	(0.08)
m	12	7	12	6	11	8	11	12
Mean	13.92	1.98	4.49	6.84	2.66	9.59	6.32	10.93
STD	30.65	23.67	27.65	26.21	28.20	33.33	27.23	31.60
Skew	0.13	−0.19	−0.04	−0.01	0.02	0.09	0.17	0.32
Kurt	4.06	4.75	4.43	4.47	4.94	3.84	4.21	4.22

Note: This table reports parameter estimates for each of the eight countries. Refer to Table 4 for the appropriate notes that accompany the table.

In light of these results, it is re-assuring to see that the parameter estimates conform to the predictions of the well-established financial theory. For instance, there is a clear risk of return trade-off implied by the parameters – the correlation between the mean and standard deviation across the sector and country results is 0.67 and 0.75, respectively. Moreover, the tendency for $\hat{\delta}_{\text{Bear}} > \hat{\delta}_{\text{Bull}}$directly supports the ‘leverage hypothesis’ of Black (1976), which implies that stock volatility tends to rise more following a negative shock than a positive shock. As explained by Black (1976), leverage can induce future stock volatility to vary inversely with the stock price: a fall in a firm's stock value relative to the market value of its debt causes a rise in its debt–equity ratio and increases its stock volatility. The fact that these predictions adhere to existing financial theory is re-assuring and supports our belief that the analysts are informed agents who are producing meaningful predictions rather than making random guesses at the likelihood of future events.

Another feature of the results is the remarkable symmetry between bull and bear components at both the country and sector level. In fact, when one takes into consideration the sizeable number of cases where we cannot reject the hypothesis that $\delta_{\text{Bear}} = 0$or $\delta_{\text{Bull}} = 0$, there is strong evidence to suggest that the analysts primarily choose to convey the differing nature of ‘large’ jumps via the use of different probability values rather than through different volatilities of the jump components. Although we cannot assess the epoch dependency of this conclusion, it does conform to the guidelines outlined in the technical report by Weyns, Perez, and Jenkins (2007) who suggest that analysts should adopt this exact procedure. Our results, therefore, imply that the analysts have heeded this advice, at least over the period under consideration.

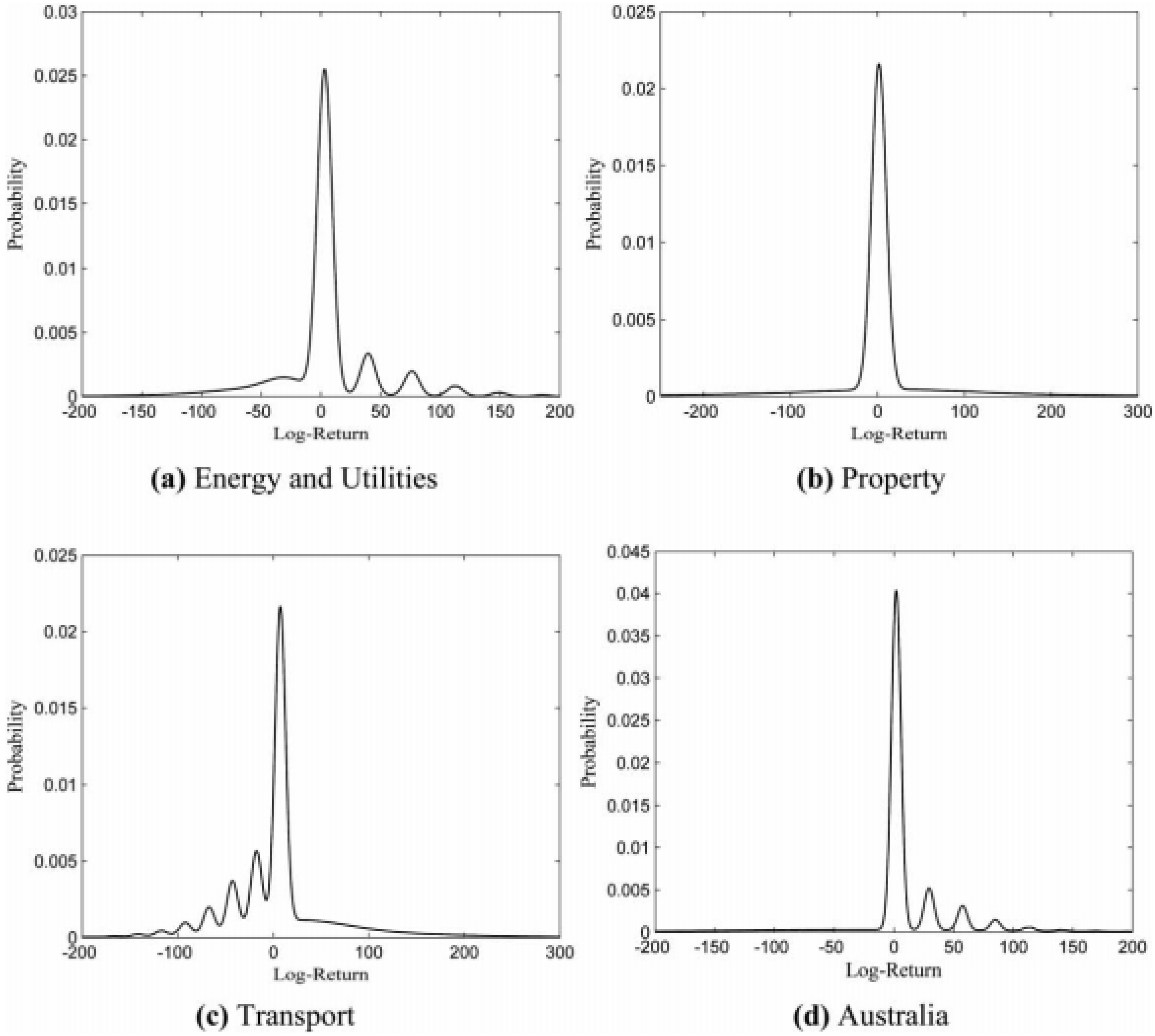
Finally, we conclude with some insights into the properties of particular sectors/countries from the perspectives of the analysts afforded by our modelling framework. Beginning with the sector results, we see that with the exception of Property, the analysts universally believe that all of the return distributions are positively skewed over the forecasting period. Correspondingly, we also estimate that

$\alpha_{Bull} > \alpha_{Bear}$ in each case and so it appears that the analysts believe that positive jumps will be on average greater in magnitude than negative jumps over the forecasting horizon at the sector level. In contrast, the bull and bear jump estimates from the country level exhibit much more symmetry with the exception of Taiwan and Asean – arguably two of the most heterogeneous regions in Asia. Another interesting point of comparison of these results pertains to their implied portfolio recommendations. Under the assumption of a common risk-free rate across Asia ex-Japan, we find that the Sharpe ratios vary dramatically between the sectors and countries (Sharpe 1966) Asean and Taiwan, for instance, have the highest Sharpe ratios, whereas Australia and Property have the lowest. Thus, since these models were calibrated using forward-looking data, these statistics are arguably more useful for portfolio managers than forecasted statistics based on backward-looking historical data.

Further insights into the various sectors/countries as perceived at the analyst level can be gained by calculating the various transition densities implied by the parameter estimates using a truncated version of Equation (4); four examples of these transition densities are provided in Figure 2.[11] The remarkable conclusion from these diagrams is that the obvious presence of a bull and bear jump component in the parameter estimates does not necessarily translate into a multimodal transition density. The starkest illustration of this is the unimodal Property transition density, Figure 2b, which implies that the majority of the jumps are completely off-setting. What is more, even when the parameter estimates convey that both bull and bear jumps are prevalent, the left and right tails of the density can display markedly different characteristics. Consider, for instance, the case of Transport, Figure 2c, which exhibits significant discontinuous mass at various spikes in the left tail, whereas the right tail displays continuous, albeit slow, decay. These finding have obvious risk management implications since the analysts are essentially predicting different amounts of mass in the left tail of the predictive density. This will be further explored in the following section.

[11] Although each of the remaining 15 densities is subtly different from those described in the figure, they nevertheless broadly adhere to one of the four shapes described in the figure.

Figure 2. Transition density examples.



7. Risk management implications

In the preceding section we saw how the database of analysts' forecasts could be used to estimate the parameters of the latent double-jump diffusion and the associated transition densities for the various sectors and countries. We also suggested how calibration of these processes could yield additional insights into portfolio formation beyond those afforded by backward-looking methods. Arguably, a more natural application of these results, however, is in the area of risk management. After all, the primary role of the analysts is one of jump prediction and so we believe that these findings provide a more reliable indication of the probability of negative extremal events when compared with alternative risk measurement techniques such as the popular RiskMetrics (1993) methodology, which is based entirely on historical data.

To illustrate the risk management connotations of the results, Tables 6 and 7 provide estimates of the 5% value-at-risk (VaR) and expected shortfall (ES) for each sector and country and the corresponding estimates based on a Gaussian distribution with the same mean and variance.[12] Notably, there are only four cases in which the VaR calculated using the transition density is greater in magnitude than the VaR calculated using the relevant Gaussian distribution. Consequently, one can conclude that the analysts are predicting a distinct lack of large negative jumps over the forecasting horizon. Nevertheless, conditional on the return being in the lowest 5% quantile, the estimates of ES imply that the analysts predict much larger average losses in these circumstances than those occurring under Gaussianity – a consequence of the slow tail decay of the transition densities.[13]

Table 6. Sector VaR and ES estimates.

	CDI	CS	EU	F	Mat	Med	P	R	Tech	TS	Tr
VaR	−63.66	−21.83	−38.89	−24.72	−36.25	−40.38	−17.96	−30.14	−52.00	−40.07	−58.68
G-VaR	−45.51	−49.69	−43.86	−31.54	−46.31	−51.65	−31.84	−36.60	−46.39	−40.24	−40.66
ES	−77.98	−64.71	−66.66	−58.67	−64.70	−71.62	−56.28	−52.80	−71.80	−78.97	−80.36
G-ES	−58.50	−64.05	−56.22	−41.09	−58.70	−66.77	−40.60	−47.48	−60.36	−52.45	−52.96

Note: This table reports value-at-risk (VaR) and expected shortfall (ES) estimates for each of the 11 sectors based on the double-jump transition densities. Estimates based on the Gaussian density are also reported for comparative purposes; these estimates are denoted by G-VaR and G-ES, respectively.

Table 7. Country VaR and ES estimates

	Asean	Australia	China	Hong Kong	India	South Korea	Singapore	Taiwan
VaR	−25.25	−7.25	−27.51	−31.15	−31.76	−49.97	−37.75	−31.86
G-VaR	−36.49	−36.95	−40.98	−36.28	−43.73	−45.24	−38.47	−41.05
ES	−59.07	−50.04	−69.97	−66.44	−66.41	−69.57	−69.96	−65.55
G-ES	−48.86	−46.85	−52.46	−46.87	−55.63	−59.59	−49.68	−54.13

Note: This table reports value-at-risk (VaR) and expected shortfall (ES) estimates for each of the eight countries. Refer to Table 6 for the appropriate notes that accompany the table.

Notwithstanding the calculation of risk measures such as VaR and ES, the estimated jump-diffusion process can also be used to stress-test existing investment strategies. Under our assumption that the analysts are informed agents about the magnitude and number of jumps, stress-tests based on extreme realisations from the latent underlying process may constitute a more reliable test of the performance of an investment strategy under realistic extremal conditions.

8. Conclusion

The purpose of this paper has been to model analysts' forecasts for normal (base), optimistic (bull), and pessimistic (bear) scenarios using a set of MS reports for Asia (excluding Japan) for 2007–2008. To this end we developed a modified method of GMM estimation inspired by Bayesian econometrics and tailored to suit the discrete nature of analysts' prior beliefs regarding the numbers of possible jump events. A rigorous derivation of the asymptotic properties of this estimator was also provided.

Based on the results of estimation, we uncovered evidence to suggest that analysts' beliefs conform to the predictions of a number of well-established theories, including the positive correlation between risk and return and Black's leverage hypothesis. We also documented a number of epoch dependent predictions made by the analysts that indicate how this approach can be used for portfolio construction purposes. In particular, the analysts' predictions imply that investments in Asean and

[12] See, Duffie and Pan (1996) for an introduction to VaR, and Artzner et al. (1999) for a nice exposition of ES.

[13] The fact that the conclusions based on VaR and ES differ should not be surprising since the former does not give information on the potential size of the loss beyond a particular quantile (Artzner et al. 1999).

Taiwan constitute the 'best' investment, both in terms of Sharpe ratio forecasts and the dominance of bull jumps over bear jumps. As a second application, we demonstrated how the implied transition densities can be used to provide an augmented measure of the VaR and ES for each of the sectors and countries and how, within a stress-testing framework, the calibrated double-jump process could provide a more reliable test of the performance of various investment strategies under extremal stock movements.

To summarise, this paper has sought to contribute to the ongoing discussion between practitioners and academics in order to better understand the activities of financial analysts and to advance the methodological basis for the use of analysts' predictions as an additional input into existing risk management and portfolio construction systems. In contrast to much of the existing literature, which focuses mainly on the predictive accuracy and efficacy of analysts' reports, we hope that the clear benefits gained from our alternative interpretation of the role played by analysts – namely, to describe and predict the impact of jump events – have been brought to the fore.

Acknowledgements

The authors would like to thank, without implicating in our conclusions, Morgan Stanley International Plc for generously providing their analysts' database for our research. We are also grateful to John Knight for helpful discussions and comments.

References

1. Aït-Sahalia, Y. 2004. Disentangling diffusion from jumps. *Journal of Financial Economics*, 74(3): 487–528.

2. Applebaum, D. 2004. "Lévy processes and stochastic calculus". In *Cambridge studies in advanced mathematics*, Vol. 93, Cambridge: Cambridge Univ. Press.

3. Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. 1999. Coherent measures of risk. *Mathematical Finance*, 9(3): 203–28.

4. Barndorff-Nielsen, O. 1988. "Parametric statistical models and likelihood". In *Lecture notes in statistics*, Vol. 50, Heidelberg: Springer-Verlag.

5. Black, F. Studies of stock market volatility changes. Proceedings of the American Statistical Association, business and economic statistics section. pp.177–81.

6. Bunn, D. W. and Salo, A. A. 1993. Forecasting with scenarios. *European Journal of Operational Research*, 68(3): 291–303.

7. Carrasco, M., Chernov, M., Florens, J. P. and Ghysels, E. 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2): 529–73.

8. Cont, R. and Tankov, P. 2004. *Financial modelling with jump processes*, Boca Raton: Chapman & Hall.

9. Duffie, D. and Pan, J. 1996. An overview of value at risk. *Journal of Derivatives*, 4(3): 13–32.

10. Hall, A. R. 2000. Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica*, 68(6): 1517–27.

11. Hall, P. and Horowitz, J. 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4): 891–916.

12. Hastie, T., Tibshirani, R. and Friedman, J. 2001. *The elements of statistical learning: Data mining, inference, and prediction*, New York: Springer.

13. Honoré, P. 1998. "Pitfalls in estimating jump diffusion models". University of Aarhus Working Paper Series no. 18 Centre for Analytical Finance.

14. Kerl, A. and Walter, A. 2008. Never judge a book by its cover: What security analysts have to say beyond recommendations. *Financial Markets and Portfolio Management*, 22(4): 289–321.

15. Kiefer, N. M. 1978. Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, 46(2): 427–34.

16. Merton, R. C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1–2): 125–144.

17. Newey, W. K. and McFadden, D. 1994. "Large sample estimation and hypothesis testing". In *Handbook of econometrics*, Edited by: Engle, R. F. and McFadden, D. L. Vol. IV, 2111–245. Amsterdam: North-Holland.

18. Press, J. 1967. A compound events model for security prices. *Journal of Business*, 40(3): 317–35.

19. Sharpe, W. F. 1966. Mutual fund performance. *Journal of Business*, 39(1): 119–38.

20. Tankov, P. 2007. Lévy processes in finance and risk management. *Wilmott Magazine*, 31 September/October

21. Weyns, G., Perez, J. L. and Jenkins, V. December 2007. "Risk-reward views: Unlocking the full potential of fundamental analysis". December, Morgan Stanley Global Research Paper

Appendix 1

A1. Raw moments

The raw moments of z_τ are calculated using

$$E(z_{t+\tau}^{j}) = \left. \frac{\phi_z^{[j]}(x)}{i^j} \right|_{x=0} \quad \text{(A1)}$$

for $j = 1, \ldots, j_{\text{Max}}$, where $\phi_z^{[j]}$ denotes the jth-order derivative of ϕ_s with respect to x. Accordingly, the first six central moments of z_τ which form the basis for the GMM estimation procedure can be calculated as follows

$$M[1, \tau] = E[z_\tau] = (\mu + \alpha_{\text{Bull}}\lambda_{\text{Bull}} + \alpha_{\text{Bear}}\lambda_{\text{Bear}})\tau \quad \text{(A2)}$$

$$M[2, \tau] = E[(z_\tau - E[z_\tau])^2] = (\sigma^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^2 + \delta_{\text{Bull}}^2) + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^2 + \delta_{\text{Bear}}^2))\tau \quad \text{(A3)}$$

$$\begin{aligned} M[3, \tau] &= E[(z_\tau - E[z_\tau])^3] \\ &= \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^3 + 3\alpha_{\text{Bull}}\delta_{\text{Bull}}^2)\tau + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^3 + 3\alpha_{\text{Bear}}\delta_{\text{Bear}}^2)\tau \end{aligned} \quad \text{(A4)}$$

$$\begin{aligned} M[4, \tau] &= E[(z_\tau - E[z_\tau])^4] \\ &= 3(\sigma^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^2 + \delta_{\text{Bull}}^2) + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^2 + \delta_{\text{Bear}}^2))\tau^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^4 + 6\alpha_{\text{Bull}}^2\delta_{\text{Bull}}^2 + 3\delta_{\text{Bull}}^4)\tau \\ &\quad + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^4 + 6\alpha_{\text{Bear}}^2\delta_{\text{Bear}}^2 + 3\delta_{\text{Bear}}^4)\tau \end{aligned} \quad \text{(A5)}$$

$$\begin{aligned} M[5, \tau] &= E[(z_\tau - E[z_\tau])^5] \\ &= \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^5 + 10\alpha_{\text{Bull}}^3\delta_{\text{Bull}}^2 + 15\alpha_{\text{Bull}}\delta_{\text{Bull}}^4)\tau + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^5 + 10\alpha_{\text{Bear}}^3\delta_{\text{Bear}}^2 + 15\alpha_{\text{Bear}}\delta_{\text{Bear}}^4)\tau \\ &\quad + 10[(\sigma^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^2 + \delta_{\text{Bull}}^2) + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^2 + \delta_{\text{Bear}}^2))(\lambda_{\text{Bull}}(\alpha_{\text{Bull}}^3 + 3\alpha_{\text{Bull}}\delta_{\text{Bull}}^2) \\ &\quad + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^3 + 3\alpha_{\text{Bear}}\delta_{\text{Bear}}^2))]\tau^2 \end{aligned} \quad \text{(A6)}$$

$$\begin{aligned} M[6, \tau] &= E[(z_\tau - E[z_\tau])^6] \\ &= (\sigma^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^2 + \delta_{\text{Bull}}^2) + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^2 + \delta_{\text{Bear}}^2))^3\tau^3 + 10(\lambda_{\text{Bull}}(\alpha_{\text{Bull}}^3 + 3\alpha_{\text{Bull}}\delta_{\text{Bull}}^2)\tau \\ &\quad + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^3 + 3\alpha_{\text{Bear}}\delta_{\text{Bear}}^2))^2\tau^2 + 15(\sigma^2 + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^2 + \delta_{\text{Bull}}^2) + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^2 + \delta_{\text{Bear}}^2)) \\ &\quad \times (\lambda_{\text{Bull}}(\alpha_{\text{Bull}}^4 + 6\alpha_{\text{Bull}}^2\delta_{\text{Bull}}^2 + 3\delta_{\text{Bull}}^4)\tau + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^4 + 6\alpha_{\text{Bear}}^2\delta_{\text{Bear}}^2 + 3\delta_{\text{Bear}}^4))\tau^2 \\ &\quad + \lambda_{\text{Bull}}(\alpha_{\text{Bull}}^6 + 15\alpha_{\text{Bull}}^4\delta_{\text{Bull}}^2 + 45\alpha_{\text{Bull}}^2\delta_{\text{Bull}}^4 + 15\delta_{\text{Bull}}^6)\tau \\ &\quad + \lambda_{\text{Bear}}(\alpha_{\text{Bear}}^6 + 15\alpha_{\text{Bear}}^4\delta_{\text{Bear}}^2 + 45\alpha_{\text{Bear}}^2\delta_{\text{Bear}}^4 + 15\delta_{\text{Bear}}^6)\tau \end{aligned} \quad \text{(A7)}$$

A2. Asymptotic properties of the two-step estimator

Following Newey and McFadden (1994), we now prove the consistency and asymptotical normality of our Bayesian-*inspired* two-step GMM estimator. However, before beginning, we remind the reader of a theorem and lemma which play central roles in both proofs.

Theorem A1

If there is a function $Q_0(\theta)$ *such that*: (i) $Q_0(\theta)$ *is uniquely minimised at* $\theta_0 \in \Theta$; (ii) Θ *is compact*; (iii) $Q_0(\theta)$ *is continuous*; (iv) $Q_N(\theta)$ *converges in probability to* $Q_0(\theta)$, *then* $\hat{\theta} \xrightarrow{p} \theta_0$.

Proof See Newey and McFadden (1994, 2121, Theorem 2.1). □

Lemma A1

(*Uniform weak law of large numbers*) *If the data are iid*, Θ *is compact*, $a(z_n, \theta)$ *is continuous at each* $\theta \in \Theta$ *with probability* 1, *and hence there is a d*(*z*) *with* $\|a(z, \theta)\| \leq d(z)$ *for all* $\theta \in \Theta$ *and* $E[d(z)] < \infty$, *then* $E[a(z, \theta)]$ *is continuous and* $\sup_{\theta \in \Theta} \|N^{-1} \sum_{n=1}^{N} a(z_n, \theta) - E[a(z, \theta)]\| \xrightarrow{p} 0$.

Proof See Newey and McFadden (1994, 2129, Lemma 2.4).

Assumption A1 (a) $\theta_0 \in \Theta$ *is the unique solution to* $E[g(\theta, z)] = 0$; (b) Θ *is compact*; (*c*) *g*(θ, *z*) *is continuous at each* $\theta \in \Theta$ *with probability* 1; (d) $E[\sup_{\theta \in \Theta} \|g(\theta, z)\|^2] < \infty$; (e) $\Omega = E[g(\theta_0, z) g(\theta_0, z)']$ *is positive definite*.

Theorem A2 *Under* Assumption A1, $\hat{\theta} \xrightarrow{p} \theta_0$.

Proof Evoking Theorem A1, it follows that we need to verify that for

$$Q(\theta) = g(\theta, z)' \Omega^{-1} g(\theta, z) \qquad \textbf{(A8)}$$

the following conditions hold: (i) *Q*(θ) is uniquely minimised at $\theta = \theta_0$; (ii) Θ is compact; (iii) *Q*(θ) is continuous; (iv) $Q_N(\theta) \xrightarrow{p} Q(\theta)$ uniformly on $\theta \in \Theta$. However, before we begin, we need to show that $\hat{\Omega}_N(\tilde{\theta})^{-1} \xrightarrow{p} \Omega^{-1}$ is positive definite where $\tilde{\theta}$ is a preliminary consistent estimate of θ_0. By the uniform weak law of large numbers (UWLLN; Lemma 1) Ω is continuous on $\theta \in \Theta$ and

$$\sup_{\theta \in \Theta} \left\| \hat{\Omega}_N(\theta) \xrightarrow{p} \Omega(\theta) \right\| \xrightarrow{p} 0$$

and thus by Assumption A1(c), (d), and (e), $\hat{\Omega}_N(\tilde{\theta}) \xrightarrow{p} \Omega$ since,

$$\begin{aligned} \|\hat{\Omega}_N(\tilde{\theta}) - \Omega\| &\leq \|\hat{\Omega}_N(\tilde{\theta}) - \Omega(\tilde{\theta})\| + \|\Omega(\tilde{\theta}) - \Omega\| \\ &\leq \sup_{\theta \in \Theta} \|\hat{\Omega}_N(\tilde{\theta}) - \Omega(\tilde{\theta})\| + \|\Omega(\tilde{\theta}) - \Omega\| \xrightarrow{p} 0 \end{aligned}$$

using the triangle inequality, followed by the UWLLN (i.e. $\sup_{\theta \in \Theta} \|\hat{\Omega}_N(\tilde{\theta}) - \Omega(\tilde{\theta})\| \xrightarrow{p} 0$), and continuity of Ω (.) (i.e. $\|\Omega(\tilde{\theta}) - \Omega\| \xrightarrow{p} 0$). Hence, $\hat{\Omega}_N(\tilde{\theta})$ is positive definite with probability approaching 1 since Ω is positive definite by Assumption A1(e). Thus, the inverse $\hat{\Omega}_N(\tilde{\theta})^{-1}$ exists and is also positive definite with probability approaching 1.

Thus, we are now ready to verify the conditions of Theorem A1 for *Q*(θ). Condition (i) follows from Assumption A1(a) and (e). Condition (ii) follows from Assumption A1(b). Condition (iii) follows from the UWLLN applied to *g*(θ, *z*) using Assumption A1(b), (c), and (d); that is, we have that $\sup_{\theta \in \Theta} \|\hat{g}(\theta) - g(\theta)\| \xrightarrow{p} 0$ and *g*(θ) is continuous. Thus, (iii) holds because $Q(\theta) = g(\theta, z)' \Omega^{-1} g(\theta, z)$ is continuous. Finally, to show that (iv) holds we use the triangle inequality to show that

$$|Q_N(\theta) - Q(\theta)| \leq |(\hat{g}(\theta) - g(\theta))'\hat{\Omega}_N(\tilde{\theta})^{-1}(\hat{g}(\theta) - g(\theta))| + 2|g(\theta)\hat{\Omega}_N(\tilde{\theta})^{-1}(\hat{g}(\theta) - g(\theta))|$$
$$+ |g(\theta)'(\hat{\Omega}_N(\tilde{\theta})^{-1} - \Omega^{-1})g(\theta)|$$

and the Cauchy–Schwartz inequality to show that

$$|Q_N(\theta) - Q(\theta)| < \|\hat{g}(\theta) - g(\theta)\|^2 \|\hat{\Omega}_N(\tilde{\theta})^{-1}\| + 2\|g(\theta)\| \|\hat{\Omega}_N(\tilde{\theta})^{-1}\| \|\hat{g}(\theta) - g(\theta)\| + \|g(\theta)\|^2 \|\hat{\Omega}_N$$

It, therefore, follows that $Q_N(\theta) \xrightarrow{p} Q(\theta)$ uniformly on $\theta \in \Theta$ from the UWLLN and Assumption A1(d). Hence, we have shown that conditions (i)–(iv) of Theorem A1 are satisfied and thus $\hat{\theta} \xrightarrow{p} \theta_0$.
□

Assumption A2 (a) $\theta_0 \in \text{int}(\Theta)$, (b) $g(\theta, z)$ *is continuously differentiable on a neighbourhood* $\mathbb{N}$ *of* θ_0, $E[\sup_{\theta \in \mathbb{N}} \|\partial g(\theta, z)/\partial \theta'\|^2] < \infty$, (c) $E[\sup_{\theta \in \mathbb{N}} \|\partial g(\theta_0, z)/\partial \theta'\|^2]$ *is full column rank.*

Theorem A3 *Under* Assumptions A1 *and* A2, *it follows that*

$$N^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1}) \quad \text{(A9)}$$

where $G = E[\partial g(\theta_0, z)/\partial \theta']$.

Proof Assumption A1 implies that $\hat{\theta} \xrightarrow{p} \theta_0$. Thus, using Assumption A2(a) and (b), the first-order conditions for a minimum

$$\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}\hat{g}(\hat{\theta}) = 0 \quad \text{(A10)}$$

are satisfied with probability approaching 1, where $\hat{G}(\theta) = 1/N \sum_{n=1}^{N} (\partial g(\theta, z_n))/\partial \theta'$. A Taylor expansion of $\hat{g}(\hat{\theta})$ about θ_0 yields

$$\hat{g}(\hat{\theta}) = \hat{g}(\theta_0) + \hat{G}(\theta^*)(\hat{\theta} - \theta_0) \quad \text{(A11)}$$

where θ^* lies on the line segment joining ˆ θ and θ_0. Therefore, with probability approaching 1,

$$\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}\hat{G}(\theta^*)(\hat{\theta} - \theta_0) = -\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}\hat{g}(\theta_0) \quad \text{(A12)}$$

Using Assumption A2(b) and thus the UWLLN, we know that

$$\sup_{\theta \in \mathbb{N}} \|\hat{G}(\theta) - G(\theta)\| \xrightarrow{p} 0$$

and $G(\theta)$ is continuous on $\theta \in \mathbb{N}$. Hence, it follows that $\hat{G}(\hat{\theta}), \hat{G}(\theta^*) \xrightarrow{p} G$, and so, it follows from Slutsky's theorem that

$$\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}\hat{G}(\theta^*) \xrightarrow{p} G'\Omega^{-1}G$$

and

$$\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1} \xrightarrow{p} G'\Omega^{-1}$$

Moreover, since $G'\Omega^{-1}G$ is positive definite by Assumptions A1(e) and A2(c), $\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}\hat{G}(\theta^*)$ is positive definite with probability approaching 1. Therefore, with probability approaching 1, Equation A11 becomes

$$N^{1/2}(\hat{\theta}-\theta_0) = -[\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}\hat{G}(\theta^*)]^{-1}\hat{G}(\hat{\theta})'\hat{\Omega}_N(\tilde{\theta})^{-1}N^{1/2}\hat{g}(\theta_0) \qquad \text{(A13)}$$

Evoking the Lindeberg–Lévy central limit theorem we have that $N^{1/2}\hat{g}(\theta_0) \xrightarrow{d} N(0,\,\Omega)$, and so, applying Cramer's theorem yields the result

$$N^{1/2}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\,(G'\Omega^{-1}G)^{-1})$$

where

$$G = E\left[\sum_{w=1}^{W}\pi_w \frac{\partial h(\theta_0, z)}{\partial \theta'}\right]$$

$$\Omega = E\left[\sum_{w=1}^{W}\pi_n h(\theta_0, z)\sum_{w=1}^{W}\pi_n h(\theta_0, z)'\right]$$