A Practical Guide to GMM

(with Applications to Option Pricing)

Tom Arnold and Timothy Falcon Crack*

October 1999

*Assistant Professor, Ourso College of Business Administration, Finance Department, Louisiana State University, 2159 CEBA, Baton Rouge, LA 70803, E-mail: tmarnol@unix1.sncc.lsu.edu, Tel: (225) 388 6369, Fax: (225) 388 6266, and Assistant Professor, Finance Department, Kelley School of Business, Indiana University, 1309 East Tenth Street, Bloomington, IN 47405-1701. E-mail: tcrack@indiana.edu, Tel: (812) 855 2695; Fax: (812) 855 5875.

A Practical Guide to GMM (with Applications to Option Pricing)

Abstract

Generalized Method of Moments (GMM) is underutilized in financial economics because it is not adequately explained in the literature. We use a simple example to explain how and why GMM works. We then illustrate practical GMM implementation by estimating and testing the Black-Scholes option pricing model using S&P500 index options data. We identify problem areas in implementation and we give tactical GMM estimation advice, troubleshooting tips, and pseudo code. We pay particular attention to proper choice of moment conditions, exactly-identified versus over-identified estimation, estimation of Newey-West standard errors, and numerical optimization in the presence of multiple local extrema.

JEL Classification: A23, C13, C23, G13.

Keywords: Generalized Method of Moments, GMM, Newey-West, Option Pricing, Black-Scholes.

Generalized Method of Moments (GMM) suffers from a lack of adequate explanation in the financial economics and financial econometrics literature. This shortfall causes significant barriers to entry for researchers who might otherwise benefit from the use of GMM techniques. This paper removes those barriers to entry.

We use a simple example to explain how and why GMM works. We then demonstrate GMM by estimating and testing the famous Black-Scholes option pricing model (Black and Scholes (1973)). We choose an option pricing application because the literature on the application of GMM to option pricing is particularly limited. We use simple explanations and we give tactical estimation advice, troubleshooting tips, and pseudo-code. Our intended audience includes empirical financial economics researchers, students of econometrics, and derivatives practitioners.

The paper proceeds as follows: in Section I we explain how and why GMM works via an extended yet simple example; in Section II we explain the attractive features of GMM; in Section III we give a brief review of Black-Scholes option pricing theory; in Section IV we demonstrate GMM in an option pricing framework; Section V concludes; Appendix A discusses our data; and Appendix B presents pseudo code for GMM estimation.

I Understanding GMM – Simple Example

Generalized method of moments is a generalization of the classical Method of Moments (MOM) estimation technique. The classical MOM technique equates sample and population moments to enable estimation of population parameters. We think that MOM, GMM, and the relationship between them are best understood via a simple example (ours is a much-expanded version of one given by Hamilton (1994, pp409-412)). A much more sophisticated application of GMM to actual option pricing data appears in Section IV.

Suppose we have data $Y_1, \ldots, Y_T$ distributed Student-t with ν degrees of freedom, and we want to estimate ν. It is well known that the mean and variance of a Student-t are $E(Y_t) = 0$, and $E(Y_t^2) = \frac{\nu}{\nu-2}$, respectively, for $\nu > 2$ (Evans et al (1993)). By equating the sample second moment $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} Y_t^2$ and the population second moment $\frac{\nu}{\nu-2}$, we may deduce a

MOM estimator for ν as in Equation (1).

$$\hat{\nu}^{(1)} = \frac{2\hat{\sigma}^2}{\hat{\sigma}^2 - 1}. \tag{1}$$

For data distributed Student-t, it is also known that the population fourth moment satisfies $\mu_4 = E(Y_t^4) = \frac{3\nu^2}{(\nu-2)(\nu-4)}$ when $\nu > 4$ (Evans et al (1993)). Suppose we equate the sample fourth moment $\hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} Y_t^4$ and the population fourth moment $\frac{3\nu^2}{(\nu-2)(\nu-4)}$, and solve for ν. After manipulation, the quadratic nature of the problem yields two additional MOM solutions for ν. Without loss of generality call both solutions "$\hat{\nu}^{(2)}$" as in Equation (2).

$$\hat{\nu}^{(2)} = \frac{3\hat{\mu}_4 \pm \sqrt{\hat{\mu}_4(\hat{\mu}_4 + 24)}}{(3\hat{\mu}_4 - 3)}. \tag{2}$$

In repeated simulations we find that only one of the two quadratic solutions in Equation (2) is close to the $\hat{\nu}^{(1)}$ of Equation (1) – the other is spurious and should be discarded.[1] Thus we obtain two estimators of the degrees of freedom for the Student-t distribution: $\hat{\nu}^{(1)}$, and $\hat{\nu}^{(2)}$.

The estimators $\hat{\nu}^{(1)}$ and $\hat{\nu}^{(2)}$ are based on the sample statistics $\hat{\sigma}^2$, and $\hat{\mu}_4$ respectively. If these latter statistics are by chance equal to the true parameters (i.e. $\hat{\sigma}^2 = \sigma^2$, and $\hat{\mu}_4 = \mu_4$), then our two estimators $\hat{\nu}^{(1)}$ and $\hat{\nu}^{(2)}$ are identical and are equal to the true ν. However, both $\hat{\sigma}^2$, and $\hat{\mu}_4$ are necessarily estimated with error. Their sampling distributions are continuous probability densities, so it follows that although our two estimators of ν are similar, the chance that they are the same in practice is zero. Thus the parameter ν is over identified and no single $\hat{\nu}$ will solve both Equations (1) and (2).

To illustrate we simulate 1,000 independent drawings $Y_1, \ldots, Y_{1000}$ from a Student-t distribution with $\nu = 10$ degrees of freedom. For our simulated data set we get $\hat{\nu}^{(1)} = 9.252$ and $\hat{\nu}^{(2)} = 10.408$ (the other quadratic root is $\hat{\nu}^{(2)} = 1.529$ and we discard it as spurious). We see that although both $\hat{\nu}^{(1)}$ and $\hat{\nu}^{(2)}$ are close to the true $\nu = 10$, the sampling error in the second and fourth moments respectively yields distinct estimators of the degrees of freedom ν.

Equations (1) and (2) are transformations of the original moment-matching conditions repeated here in Equations (3).

$$\hat{\sigma}^2 = \frac{\nu}{\nu - 2}, \quad \text{and} \quad \hat{\mu}_4 = \frac{3\nu^2}{(\nu-2)(\nu-4)}. \tag{3}$$

From our sampling error arguments it follows that no single MOM estimator $\hat{\nu}$ can solve both moment-matching conditions in Equations (3). With two estimators $\hat{\nu}^{(1)} = 9.252$ and $\hat{\nu}^{(2)} = 10.408$, how then are we to choose a single sensible MOM estimator for $\hat{\nu}$?

The solution is to generalize our MOM approach for estimating ν by introducing a 2×2 weighting matrix W that reflects our confidence in each moment-matching condition in Equations (3). We execute this by stacking our previous moment-matching conditions into a vector g as in Equation (4).

$$g(\nu) \equiv \begin{bmatrix} \hat{\sigma}^2 - \frac{\nu}{\nu-2} \\ \hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix}. \tag{4}$$

We then minimize the scalar quadratic objective function $Q(\nu) = g(\nu)'W g(\nu)$ with respect to choice of ν. The contents of W describe the relative importance of each moment-matching condition in determining $\hat{\nu}$. If the matrix W is simply the 2×2 identity,[2] then the objective function reduces to

$$\begin{aligned} Q(\nu) &= g(\nu)'g(\nu) \\ &= \left[\hat{\sigma}^2 - \frac{\nu}{\nu-2}\right]^2 + \left[\hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)}\right]^2. \end{aligned}$$

If we think of the moment conditions as residuals (i.e. deviations from their ideal value of zero), then an identity weighting matrix reduces the quadratic objective function to a sum of squared residuals, and it reduces the optimization to a traditional least squares problem.

More generally, the weighting matrix W should place more weight on the moment-matching conditions in which we have more confidence. The obvious choice for a W that is directly related to our confidence in the moment conditions in Equation (4) is the inverse of the variance-covariance matrix (VCV) of the moment conditions. In practice, rather than inverting the VCV of the vector g, we shall invert the asymptotic VCV defined as $\Omega \equiv \mathrm{var}(\sqrt{T}g) = T\mathrm{var}(g)$. This VCV is typically a function of the estimator itself, but we suppress the dependence of Ω on ν for ease of notation. Thus we shall minimize $Q(\nu) = g(\nu)'\Omega^{-1}g(\nu)$.

How are we to think of the minimization of the objective function $Q(\nu)$? A simple analogy is that when evaluated at the optimum $\hat{\nu}$, the objective $Q(\hat{\nu}) = g(\hat{\nu})'\hat{\Omega}^{-1}g(\hat{\nu})$ is similar to a squaring of a scaled version of the traditional t-statistic for testing whether a population mean is zero as reported in Equation (5).

$$\Bigg\{\underbrace{\left(\frac{1}{\sqrt{T}}\right)}_{\text{scale factor}} \cdot \underbrace{\left[\frac{\hat{\mu}-0}{\hat{\sigma}/\sqrt{T}}\right]}_{t\text{-statistic}}\Bigg\}^2 = (\hat{\mu}-0)\left[T\left(\frac{\hat{\sigma}}{\sqrt{T}}\right)^2\right]^{-1}(\hat{\mu}-0). \tag{5}$$

The analogy follows because: $g(\hat{\nu})$ and $(\hat{\mu}-0)$ should be zero under their respective null hypotheses; $g(\hat{\nu})$ and $(\hat{\mu}-0)$ appear fore and aft in their respective expressions; $\hat{\Omega}$ and $[T(\hat{\sigma}/\sqrt{T})^2]$ are the estimated asymptotic variances of $g(\hat{\nu})$ and $\hat{\mu}$ respectively (i.e. you divide them by T to get actual variances of $g(\hat{\nu})$ and $\hat{\mu}$ respectively in large sample); and finally, the asymptotic variances are inverted in the kernels of both expressions.

When minimizing the objective function $Q(\nu)$, we are building through choice of $\hat{\nu}$ a test statistic least likely to reject the hypothesis that the moments are zero. If we multiply the optimized objective function by T, we get an asymptotically chi-squared test statistic for whether the moment conditions are zero.[3] Thus by construction the GMM estimator of ν is that value of ν statistically least likely to reject the null hypothesis that the moments $g(\nu)$ are zero. This optimal ν is selected via weighted least squares minimization of a quadratic function of moment conditions.

In the case of our simulation of 1,000 observations from a Student-t, the objective function $Q(\hat{\nu}) = g(\hat{\nu})'\hat{\Omega}^{-1}g(\hat{\nu})$ is shown in Figure 1. A simple numerical minimization of the objective function in Figure 1 locates the GMM estimator $\hat{\nu} = 11.337$. Our two MOM estimators $\hat{\nu}^{(1)} = 9.252$ and $\hat{\nu}^{(2)} = 10.408$ bracket the true ν, but the GMM estimator is outside of this range.[4] The numerical ordering of $\hat{\nu}^{(1)}$, $\hat{\nu}^{(2)}$, and the GMM estimator vary with the random seed used in the simulation.

The next question is how to get the standard error of our GMM estimator. We need to give more details of the formal GMM setup and the VCV estimations to answer this question. Hansen's (1982) GMM is a formalization of our technique. Assume the underlying data X_t are stationary and ergodic.[5] We use economic theory and intuition (see examples in our Section IV) to obtain q unconditional moment restrictions on f (a vector of functions) for true parameter vector β_0:

$$f(X_t, \beta_0) = \begin{bmatrix} f_1(X_t, \beta_0) \\ f_2(X_t, \beta_0) \\ \vdots \\ f_q(X_t, \beta_0) \end{bmatrix}, \quad \text{where} \quad E[f(X_t, \beta_0)] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

For admissible β, we let $g_T(\beta) \equiv \frac{1}{T}\sum_{t=1}^{T} f(X_t,\beta)$. In our Student-$t$ example we are implicitly using

$$f(Y_t,\nu) = \begin{bmatrix} f_1(Y_t,\nu) \\ \\ f_2(Y_t,\nu) \end{bmatrix} = \begin{bmatrix} Y_t^2 - \frac{\nu}{\nu-2} \\ \\ Y_t^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix} \text{ for } t = 1,\ldots,T$$

$$\text{to get} \quad g(\nu) = \frac{1}{T}\sum_{t=1}^{T} f(Y_t,\nu) = \begin{bmatrix} \hat{\sigma}^2 - \frac{\nu}{\nu-2} \\ \\ \hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix}.$$

Let W_T be positive definite such that $\lim_{T\to\infty} W_T = W$, where W is positive definite, then the GMM estimator $\hat{\beta}_{GMM}$ is the choice of β that minimizes the scalar quadratic objective function $Q_T(\beta) = g_T(\beta)' W_T g_T(\beta)$ (we argue shortly that $W_T = \Omega^{-1}$). If we assume that a weak law of large numbers applies to the average g, so that $g_T(\beta) \xrightarrow{T} E[g_T(\beta)]$, and in particular $g_T(\beta_0) \xrightarrow{p} 0$, then $\hat{\beta}_{GMM}$ is consistent for the true β_0. Assume a central limit theorem applies to $f(X_t,\beta_0)$, so that the (appropriately scaled) sample mean g of the f_t's satisfies $\sqrt{T} g_T(\beta_0) \overset{\text{“a”}}{\sim} \mathcal{N}(0,\Omega)$ (where Ω is the asymptotic VCV of g).

If $q = p$ (the number of restrictions in f equals the number of parameters in β_0), then β_0 is exactly identified, and $\hat{\beta}_{GMM}$ is independent of choice of the weighting matrix. However, if $q > p$, β_0 is over identified. In this case different weighting matrices lead to different $\hat{\beta}_{GMM}$. In either case, a numerical optimization routine (e.g. Newton-Raphson, or Berndt et al (1974)) is typically but not always needed to find $\hat{\beta}_{GMM}$. Hansen (1982) shows that in the over-identified case, $W_0 = \Omega^{-1}$ gives the asymptotically efficient GMM estimator (consistent with our earlier intuition that the weights should be inversely related to our uncertainty regarding the moments). The GMM estimator $\hat{\beta}_{GMM}$ is asymptotically Normal, with

$$\begin{aligned}
\sqrt{T}(\hat{\beta}_{GMM} - \beta_0) &\overset{\text{“a”}}{\sim} \mathcal{N}[0, V_{GMM}], \quad \text{where} \\
V_{GMM} &= (\Gamma'\Omega^{-1}\Gamma)^{-1}, \\
\Gamma &= E\left(\frac{\partial g_T(\beta_0)}{\partial \beta'}\right) = E\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(X_t, \beta_0)}{\partial \beta'}\right), \quad \text{and} \\
\Omega &= E[\sqrt{T} g_T(\beta_0) \cdot \sqrt{T} g_T(\beta_0)'] \\
&= E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\underbrace{f(X_t, \beta_0)}_{q\times 1}\underbrace{f(X_s, \beta_0)'}_{1\times q}\right].
\end{aligned}$$

The matrix Ω gives the asymptotic VCV of the moment conditions g.[6] To get from Ω to the asymptotic VCV of the parameter vector $\hat{\beta}_{GMM}$ we need the matrix Γ to capture the relationship between the moments and the parameters. This is why Γ pre- and post-multiplies Ω^{-1} in the calculation of the VCV matrix of the parameters. Γ can sometimes be calculated explicitly. Otherwise it is estimated using $\hat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(X_t,\hat{\beta}_{GMM})}{\partial \beta'}$, which is typically a function of $\hat{\beta}_{GMM}$. In our Student-t example, the data are IID, so Γ reduces to

$$\Gamma = E\left[\frac{\partial f(Y_t,\nu)}{\partial \nu}\right] = \left[\begin{array}{c} \frac{2}{(\nu-2)^2} \\ \\ \frac{6\nu(3\nu-8)}{(\nu-2)^2(\nu-4)^2} \end{array}\right], \tag{6}$$

and we estimate Γ during the optimization by substituting in $\hat{\nu}$.

Assume the components of the moment vector $f(X_t,\beta_0)$ are not auto- or cross-correlated at any non-zero lag (i.e. $E[f_i(X_t,\beta_0)f_j(X_s,\beta_0)] = 0$ for all $t \neq s$ and for any i and j). If f is potentially heteroskedastic (i.e. $\mathrm{var}(f_i) \neq \mathrm{var}(f_j)$ for some i and j), then the matrix Ω may be estimated using the White (1980) estimator in Equation (7).

$$\hat{\Omega}_{WHITE} = \frac{1}{T}\sum_{t=1}^{T} f(X_t,\hat{\beta}_{GMM})f(X_t,\hat{\beta}_{GMM})'. \tag{7}$$

It may be seen that the ij^{th} element of $\hat{\Omega}_{WHITE}$ estimates the ij^{th} element of the asymptotic VCV of the vector g as follows.

$$\begin{aligned}
\operatorname{var}\left(\sqrt{T}g\right)_{ij} &= \operatorname{cov}\left(\sqrt{T}g_i, \sqrt{T}g_j\right) \\
&= E\left(\sqrt{T}g_i\sqrt{T}g_j\right) \\
&= TE(g_i g_j) \\
&= TE\left[\frac{1}{T}\sum_{t=1}^{T} f_i(X_t,\beta_0)\cdot\frac{1}{T}\sum_{s=1}^{T} f_j(X_s,\beta_0)\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E[f_i(X_t,\beta_0)f_j(X_s,\beta_0)] \\
&= \frac{1}{T}\sum_{t=1}^{T} E[f_i(X_t,\beta_0)f_j(X_t,\beta_0)] + \text{cross terms} \\
&= E[f_i(X_t,\beta_0)f_j(X_t,\beta_0)] \\
&\approx \frac{1}{T}\sum_{t=1}^{T} f_i(X_t,\beta_0)f_j(X_t,\beta_0),
\end{aligned}$$

which is just the ij^{th} element of the White estimator in Equation (7). In the above, we use the fact that $E(g_i) = E(g_j) = 0$, we denote the i^{th} and j^{th} elements of the vector f as f_i, and f_j respectively, the cross-terms are zero because the moments are assumed uncorrelated (i.e. they have no own- or cross-correlation at any non-zero lag), and at the last step we use the Weak Law of Large Numbers.

If the components of the vector of moments $f(\beta_0)$ do exhibit auto- or cross-correlation (i.e. $E[f_i(X_t, \beta_0) f_j(X_s, \beta_0)] \neq 0$ for some i, j, and some $t \neq s$), and are also potentially heteroscedastic, then the Newey-West (1987) estimator of Ω may be used (Equation (8)). Practical choice of lag length m is discussed in Section D.

$$\begin{aligned} \hat{\Omega}_{NW} &= \hat{\Phi}_0 + \sum_{j=1}^{m} w(j,m)(\hat{\Phi}_j + \hat{\Phi}'_j), \quad m \ll T, \quad \text{where} \qquad (8) \\ \hat{\Phi}_j &\equiv \frac{1}{T} \sum_{t=j+1}^{T} f(X_t, \hat{\beta}_{GMM}) f(X_{t-j}, \hat{\beta}_{GMM})', \quad \text{and} \\ w(j,m) &= 1 - \frac{j}{(m+1)}. \end{aligned}$$

Note that $\hat{\Phi}_0$ in Equation (8) is just White's estimator, which Newey-West extends. Thus the Newey-West estimator is robust to both heteroskedasticity and autocorrelation of the components of the moment vector f. The White and Newey-West estimators are not the only estimators available for Ω (see Ogaki(1993), Hamilton (1994)).

To avoid any misunderstanding, let us emphasize that the White and Newey-West VCV matrix estimators as used here provide standard errors that are robust to heteroskedasticity and autocorrelation in the *moment conditions* f_t, but not necessarily in the *underlying data* X_t. However, if you have autocorrelation or heteroskedasticity in the underlying data, and this generates autocorrelation or heteroskedasticity in the moments, then your standard errors are robust to these latter deviations. Using White and Newey-West are thus analogous to assuming your data are stationary when using OLS, but allowing for non-spherical residuals – the GMM moment conditions are effectively model residuals. Note also that although different moments may have different variances, the variance structure must be stationary else GMM is not valid.

If $q > p$ (so that the parameters are over-identified), then as suggested earlier, multiplying the objective function by the sample size yields a chi-squared test statistic at the optimum (Equation (9)).

$$T \times Q_T(\hat{\beta}_{GMM}) = Tg(\hat{\nu})'\hat{\Omega}^{-1}g(\hat{\nu}) \overset{\text{“a”}}{\sim} \chi^2_{q-p}, \tag{9}$$

The distribution in Equation (9) assumes $Q_T(\beta)$ is minimized using $W = \Omega^{-1}$ (Hansen (1982)). This is a large sample test of whether the sample moments $g_T(\hat{\beta}_{GMM})$ are as close to zero as would be expected if the expectation of the population moments $E[f(X_t, \beta_0)]$ are truly zero. It is a test of model specification and is particularly strong if it rejects (after all, we choose $\hat{\beta}_{GMM}$

specifically to minimize the likelihood that this test will reject).

The chi-squared test statistic in Equation (9) is a quadratic function of the moment conditions. If the test statistic rejects, then the underlying model that generated the system of moment conditions is declared invalid. It is thus important that we select an informative set of moment conditions if the chi-squared test statistic is to truly test the model being estimated. Passing the chi-squared test is no guarantee of statistical significance of individual parameter estimators. Thus the best set of moment conditions will be those that not only pass the chi-squared test, but also admit the least number of possible values for a given model parameter (i.e. a low standard error). There is some potential for an unscrupulous econometrician to game the chi-squared test via choice of moment conditions. We illustrate both sensible and non-sensible moment condition choices and discuss gaming the chi-squared statistic in Section IV.

Continuing our Student-t example, Table I reports results from several simulations of increasing sample size. In each case the two MOM estimators $\hat{\nu}^{(1)}$, and $\hat{\nu}^{(2)}$ are reported, along with the GMM estimator $\hat{\nu}_{GMM}$, the standard error of the GMM estimator, and the chi-squared goodness-of-fit test (there are $q = 2$ moments, and $p = 1$ parameters, so it is a χ^2_1 statistic).

Only in the case $T = 10,000$ does the statistic reject, and that is probably a Type-I error. The standard errors fall as the sample size rises and the GMM estimator is quite close to the true degrees of freedom ($\nu = 10$) in the larger samples, but in each case (at least in this simulation) $\hat{\nu}_{GMM} > \nu = 10$.

II Attractive Features of GMM

The first attractive feature of GMM is that it is distributionally nonparametric. Unlike MLE, it does not place distributional assumptions on the data. However, the GMM moment conditions are certainly *functionally* parametric – you must impose a specific functional form. GMM's initial assumptions of stationarity and ergodicity are also relatively weak compared to the more traditional assumption that the data are IID. However, the resulting statistics are all asymptotic, so a sizable sample is required

GMM is well-suited to horribly non-linear models. This is partly because no matter how ugly the option pricing model, it still generates natural moment conditions. Examples include "model price less market price equals zero." These moment conditions can easily be chosen to test particularly interesting phenomena (e.g. the volatility smile discussed in Section III).

If you do not use GMM, and you fit option prices to model prices by minimizing sum of squared errors, then it is not clear how you conduct tests of goodness of fit, or tests for significance of individual parameters (e.g. Bakshi, Cao and Chen (1998) are unable to perform statistical tests). However, for example, if you allow for different implied volatilities across strike prices, then once the parameters are estimated, GMM allows for the traditional Wald-type tests of whether the parameters are the same.

GMM subsumes OLS, 2SLS, 3SLS, and other methods, so it is a very general technique (Ogaki (1993)). Like these methods, it is very easy to incorporate VCV estimations that allow for heteroskedasticity and autocorrelation in the moment conditions – using White or Newey-West VCV estimators.

The model specifications are used to create the GMM moment conditions. It follows that you do not have to proxy for these elements of the model (which would introduce error). That is, rather than proxying for a parameter and then regressing a left hand side dependent variable on this and other proxies to see if there is a relationship, you set up moment conditions that allow you to deduce the parameter from the data and the functional form of the model. The relationship is then tested using standard errors on individual parameters, and the chi-squared test for the overall model. For example, a consumption-based asset pricing model would not necessarily need a proxy for a consumption parameter.

You can look at the minimizations of specific GMM moment conditions to determine what aspect of the data is not being properly captured by the model. However, you should be careful of data mining in this instance. In fact if you look at the errors in any moment condition over time you may be able to capture specific sections of time where the model is not consistent with the data and this could indicate a regime shift.

III Black-Scholes-Merton Option Pricing Review

Black and Scholes (1973) and Merton (1973) derive pricing formulae for European-style call and put options. The authors assume that the underlying security price follows a geometric Brownian motion in continuous time with constant volatility. It follows that continuously compounded returns on the underlying security are normally distributed, and future asset prices are lognormally distributed. Interest rates are assumed constant. The present is time t. The options expire at time $t+\tau$. Under these assumptions (and some others concerning market frictions), the time-t value of the European-style put and call on a non-dividend paying security are given by $c(t)$ and $p(t)$ in Equations (10).

$$\begin{aligned}
c(t) &= S(t)N(d_1) - e^{-r\tau}XN(d_2), \text{ and} \\
p(t) &= e^{-r\tau}XN(-d_2) - S(t)N(-d_1), \text{ where} \\
d_1 &= \frac{\ln\left(\frac{S(t)}{X}\right) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \quad \text{and} \\
d_2 &= d_1 - \sigma\sqrt{\tau},
\end{aligned} \tag{10}$$

where $N(\cdot)$ is the cumulative normal function, S is the value of the underlying security, r is the riskless interest rate per annum, X is the strike price of the option, and σ is the annual standard deviation of returns to the stock. If the underlying security pays dividends continuously through time at rate q (a reasonable approximation for an index), then we replace $S(t)$ by $S(t)e^{-q\tau}$ throughout Equations (10) to get the dividend-adjusted Black-Scholes formula (also known as the Merton model (Merton (1973)) as shown in Equations (11).

$$
\begin{aligned}
c(t) &= S(t)e^{-q\tau}N(d_1) - e^{-r\tau}XN(d_2), \text{ and} \\
p(t) &= e^{-r\tau}XN(-d_2) - S(t)e^{-q\tau}N(-d_1), \text{ where} \\
d_1 &= \frac{\ln\left(\frac{S(t)}{X}\right) + (r - q + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \quad \text{and} \\
d_2 &= d_1 - \sigma\sqrt{\tau}.
\end{aligned} \tag{11}
$$

Four of the six inputs to the dividend-adjusted Black-Scholes formula are directly observable in the market place: $S(t)$, X, τ, and r. The dividend yield q can be deduced using the prices of traded options together with put-call parity (see Section IV). The final input σ may be estimated from historical asset returns. Alternatively, given the prices of traded options and the four observable inputs, the value of σ may be inferred directly by comparing the Black-Scholes formula to market prices of traded options. This so called "implied volatility" is the subject of much attention. For example, it is well known that for options on the same underlying security and with the same maturity but with differing strike prices, the implied volatilities can differ markedly (e.g. Black (1975), MacBeth and Merville (1979), Murphy (1994), Bates (1995, p43), Hull (1997, p503)). One of the aims of more complicated option pricing models is to better capture the behaviour of the underlying security price, and in so doing remove this variation in implied volatility with strike price (typically called a "volatility smile," "volatility skew," of "strike price bias").

IV Using GMM to Estimate and Test Black-Scholes

A Prior GMM-Based Work

For a broad review of applications of GMM in financial economics see Ogaki (1993). We are more interested here in option pricing applications, but very few option pricing papers use GMM. Scott (1987) and Wiggins (1987) both estimate stochastic volatility option pricing models using MOM. These are methodologically close to exactly-identified GMM estimations. Similarly, Chesney and Scott (1989) also use MOM to estimate a stochastic volatility option pricing model but they explicitly use Hansen's (1982) GMM results for calculating standard errors. Melino and Turnbull (1990) use an over-identified GMM estimation of a stochastic volatility option pricing model.

These papers find that allowing volatility to be stochastic gives better option pricing than forcing volatility to be constant. Curiously, Chesney and Scott find that the dividend-adjusted Black-Scholes (our Equation (11)) with daily revisions of implied volatility performs better than the more complicated stochastic volatility models for pricing FOREX options (Chesney and Scott (1989, Table 3)).

B Data

We purchased tick-by-tick Market Data Report (MDR) data for all market maker bid-ask quotes and all customer trades on all S&P500 stock index (SPX) options on the Chicago Board Options Exchange (CBOE) for every trading day from 1987 to 1995 (this study uses a subsample of only 96 trading days from 1995). The CBOE SPX index options contract is European style, so the Black-Scholes formula is theoretically correct with the correction for dividends in Equations (11).[7] We convert T-bill yields from the Wall Street Journal (WSJ) into risk free rates. After cutting each trading day into 15 minute windows, we end up with 2529 observations on each of six different moneyness/maturity classes (three moneyness classes crossed with two maturity classes). We relegate all further data details to Appendix A.

C Choosing Moments, Estimation and Tests

We have market prices of traded options. We also have the theoretical Black-Scholes option pricing formula with only one parameter (volatility of stock returns) that is not directly observable. On the face of it, we need only estimate the "implied volatility" number that best matches model prices to market prices and then test to see if the prices match well enough that we are unable to reject the model.[8] Of course, the practical implementation is much more complicated.

Black and Scholes assume the underlying security price process is driven by a random walk (a geometric Brownian motion) with constant volatility. In their model there is no reason for same-maturity options of differing strike price to be priced using different volatilities. That is, fixing maturity, implied volatilities are assumed to be equal across moneyness. It could be argued that Black and Scholes assume that options of different maturity should also be priced using the same volatility number, after all, their random walk has only one σ. However, if a long-dated option expires after some event such as a final judgement in a litigation or an important earnings announcement or a scheduled regulatory announcement, and if a short-dated option expires before this event, then there is good reason to expect that the market consensus volatility forecasts will differ over the lives of the short-dated and long-dated options on affected securities. Thus the two options will have associated with them different implied volatilities. It follows that when testing Black-Scholes we should allow options of different maturity to be priced using different implied volatilities. This is referred to as the "volatility term structure" (Hull, 1997, p503).

We also allow for a term structure of interest rates and we allow options of different maturity to have different dividend yields associated with them – a term structure of anticipated dividend yields. We allow the term structures of interest rates and dividend yields to time-vary. That is, they can be different every day. However, we do not allow the volatility term structure to time-vary.

Even after allowing for a term structure of volatility, and time-varying term structures of interest rates and of anticipated dividend yields, we fully expect that we will strongly (and correctly) reject the Black-Scholes model because of the "volatility smile" mentioned previously. That is, the well known variation in implied volatility across same-maturity options of different strike prices is likely to lead to a strong rejection of the Black-Scholes model. We can model this by allowing for different implied volatilities across strike price for a given maturity and then testing to see whether these volatilities differ enough to reject Black-Scholes.

For each 15 minute window t, $t = 1, \ldots, T$, we form a vector f_t of moments that match market prices to model prices.

$$f_t(\theta) = \begin{bmatrix} f_{t,1}(\theta) \\ f_{t,2}(\theta) \\ \vdots \\ f_{t,6}(\theta) \end{bmatrix} = \begin{bmatrix} c_{t,1,1}^{(\mathrm{mkt})} - c_{t,1,1}^{(\mathrm{BS})}(\sigma_{1,1}) \\ c_{t,1,2}^{(\mathrm{mkt})} - c_{t,1,2}^{(\mathrm{BS})}(\sigma_{1,2}) \\ c_{t,1,3}^{(\mathrm{mkt})} - c_{t,1,3}^{(\mathrm{BS})}(\sigma_{1,3}) \\ c_{t,2,1}^{(\mathrm{mkt})} - c_{t,2,1}^{(\mathrm{BS})}(\sigma_{2,1}) \\ c_{t,2,2}^{(\mathrm{mkt})} - c_{t,2,2}^{(\mathrm{BS})}(\sigma_{2,2}) \\ c_{t,2,3}^{(\mathrm{mkt})} - c_{t,2,3}^{(\mathrm{BS})}(\sigma_{2,3}) \end{bmatrix}$$

where $\theta = [\sigma_{1,1} \ \sigma_{1,2} \ \ldots \ \sigma_{2,3}]'$ is the vector of implied volatilities to be estimated for the six moneyness/maturity classes over the sample period, $c_{t,i,j}^{(\mathrm{mkt})}$ is the market price of a call option falling in the i^{th} maturity class, and the j^{th} moneyness class during window t, and $c_{t,i,j}^{(\mathrm{BS})}(\sigma_{i,j})$ is the dividend-adjusted Black-Scholes call pricing formula for the same option with the implied volatility $\sigma_{i,j}$ as the plug figure.

Once the parameter vector θ is estimated from the data, we can test Black-Scholes by testing whether the volatility smile is flat or not ($H_0: \sigma_{i,1} = \sigma_{i,2} = \sigma_{i,3}$ for each maturity class i), and we also can test to see if there is a term structure of volatility ($H_0 : \sigma_{1,j} = \sigma_{2,j}$ for each j).

The GMM estimation just described is exactly identified (there are six moments and six volatility parameters $\sigma_{i,j}$ to be estimated). There is thus a $\hat{\theta}$ that sets $g(\hat{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\theta})$ identically to zero and this means that whatever the weighting matrix used, the same parameter estimates will be obtained. In our case, we need a numerical optimization to locate the parameter estimates that set g to zero, but in some applications no numerical technique is required in the exactly-identified case.[9] The chi-squared goodness of fit test no longer applies (it is a test of over-identifying restrictions and our moments are exactly identified). We may use a standard Wald test to determine whether the volatility smile is flat and another to test for the term structure of volatility (for details see Table V).

An alternative to the exactly identified approach is to keep the same moments, but impose $\sigma_{i,1} = \sigma_{i,2} = \sigma_{i,3} = \sigma_i$, say, for each maturity class i

during the estimation. This GMM estimation is over identified (there are more moments than parameters to be estimated). In this case we cannot conduct a standard Wald test of whether the smile is flat (we have only one σ for each maturity, purposefully forcing the smile to be flat), but if we have falsely forced the smile to be flat, then the moments will be non-zero, and the chi-squared test of over-identifying restrictions will reject the model.

In summary, we can have one moment for every parameter, the system is exactly identified, and we can test whether the parameters differ from one another using Wald tests, or we can cut down the number of parameters to be estimated by imposing a flat volatility smile, and then use the chi-squared test of over-identifying restrictions to test the fit of the whole model (and still use a Wald test for the volatility term structure). We shall do both and then compare and contrast the results.

D Newey-West Lag Lengths

When implementing GMM using Newey-West standard errors, the choice of lag length m in Equation (8) is important for two reasons. Firstly, insufficient lag length can lead to inappropriate standard errors (either too large or too small depending upon the nature of the dependence in the data). Secondly, insufficient lag length can cause a lack of smoothness in the GMM objective function that can slow numerical optimization (recall that $W = \Omega^{-1}$ is estimated via Newey-West). A practical method for choosing the lag length m in the Newey-West estimator is to estimate the parameters and their standard errors using $m = 5$ and then re-estimate everything using increasing lag lengths until the lag length has negligible effect on the standard errors of the parameters to be estimated. In our Black-Scholes estimation we noticed substantial differences in both standard errors and quality of optimization as our lag lengths increased up to about 35, but beyond lag 50 (which we used) there was no change. One rule of thumb is to use $m = \sqrt{T} + 5$ where T is the sample size. In our case, $T \approx 2500$, so 55 lags is suggested (in agreement with our empirical findings).

E Nonsense Moments and Gaming the Chi-squared Test of Over-Identifying Restrictions

In the over-identified case, we may be tempted to append moments to the vector f_t to reflect Black-Scholes assumptions for geometric Brownian motion. These assumptions include that the continuously compounded returns (i.e. log of the price relative) are normally distributed (no skewness or kurtosis) with no autocorrelation (as in Equation (12)).

$$
f_t(\theta) = \begin{bmatrix} f_{t,1}(\theta) \\ f_{t,2}(\theta) \\ \vdots \\ f_{t,6}(\theta) \\ f_{t,7}(\theta) \\ f_{t,8}(\theta) \\ f_{t,9}(\theta) \\ f_{t,10}(\theta) \\ f_{t,11}(\theta) \end{bmatrix} = \begin{bmatrix} c_{t,1,1}^{(\mathrm{mkt})} - c_{t,1,1}^{(\mathrm{BS})}(\sigma_{1,1}) \\ c_{t,1,2}^{(\mathrm{mkt})} - c_{t,1,2}^{(\mathrm{BS})}(\sigma_{1,2}) \\ \vdots \\ c_{t,2,3}^{(\mathrm{mkt})} - c_{t,2,3}^{(\mathrm{BS})}(\sigma_{2,3}) \\ \log_e\left(\frac{S_t}{S_{t-1}}\right) - \mu \\ \left[\log_e\left(\frac{S_t}{S_{t-1}}\right) - \mu\right]^2 - \sigma_h^2 \\ \left[\log_e\left(\frac{S_t}{S_{t-1}}\right) - \mu\right] \cdot \left[\log_e\left(\frac{S_{t-1}}{S_{t-2}}\right) - \mu\right] - \rho \cdot \sigma_h^2 \\ \frac{\left[\log_e\left(\frac{S_t}{S_{t-1}}\right) - \mu\right]^3}{\sigma_h^3} - \psi \\ \frac{\left[\log_e\left(\frac{S_t}{S_{t-1}}\right) - \mu\right]^4}{\sigma_h^4} - \kappa \end{bmatrix} \tag{12}
$$

where S_t is the SPX index level at time t, and $\theta = [\sigma_{1,1}\ \sigma_{1,2}\ \ldots\ \sigma_{2,3}\ \mu\ \sigma_h\ \rho\ \psi\ \kappa]'$ where the $\sigma_{i,j}$ are as before, μ is the mean return, σ_h is the historical sample standard deviation of returns (as opposed to the implied volatilities which are forward looking), ρ is the first order autocorrelation of returns, ψ is the skewness of returns, and κ is the kurtosis of returns. However, this expanded set of moment conditions does not make sense. In testing the assumptions of the model rather than the actual quality of the pricing we risk rejecting a model with good pricing that is robust to assumptions that are not strictly true. Testing assumptions rather than pricing is one step removed from what is economically important. In other words, we do not really care if stock returns are normally distributed in this case as long as the model prices are close to the market prices.

Adding the above-mentioned or any other “nonsense moments” serves to increase the degrees of freedom of the chi-squared test of over-identifying restrictions and make it less likely that the test will reject the model. This amounts to a gaming of the chi-squared test and thus we believe that any highly over-identified GMM estimation should be viewed with substantial skepticism.

F Results and discussion

We use an elementary Newton routine (Greene (1997, Section 5.5)) to optimize the objective function in the exactly-identified case. Table V reports the implied volatilities and their standard errors and test statistics.[10] Several different starting points lead to the same solution. Splitting the data set into two halves and estimating each half separately gives very similar results and the same inferences. The Wald test for the volatility smile in Table V provides a very strong rejection of the Black-Scholes model. The volatility term structure test is also statistically significant (though perhaps not economically significant). There is a very significant volatility skew with in-the-money calls priced at substantially higher volatilities than out-of-the-money calls. This is consistent with previous literature (e.g. Hull (1997, p503)).

Absolute (i.e. dollar) and relative (i.e. percentage) pricing errors are reported in Table VI. The shorter the maturity and the more in-the-money is the option, the lower is its relative pricing error. The differences in relative pricing error are an immediate consequence of differences in "vega" across the moneyness/maturity classes. The vega of an option is the sensitivity of the option's price to changing implied standard deviation.[11] Other things being equal, vega increases with maturity, and is highest for options that are at-the-money relative to spot prices (our "out-of-the-money" class is very nearly at-the-money relative to the spot price). A higher vega means that when trying to match market prices to model prices we have less latitude in selection of the implied volatility. That is, a narrower range of σ's will do an accurate pricing job when vega is high than when vega is low. Higher vega thus goes hand-in-hand with less accurate option pricing (Table VI), but with lower standard errors on estimated σ's (Table V).

G Exactly versus Over-Identified Approach

Let us now impose upon the estimation routine that the short maturity options are priced using the same implied volatility, and all the long maturity options are priced using the same volatility, but that the two volatilities are allowed to be different. This estimation routine is over identified: there are six moneyness/maturity moment conditions, but only two parameters to be estimated. Table VII reports the implied volatilities and their standard errors. The chi-squared test of over-identifying restrictions is distributed chi-squared with four degrees of freedom (number of moments less number of parameters) and it takes the value 46.92. We reject the model with *p-value* of essentially zero. We can form a Wald test of whether there is a term structure of volatility. In this case the test statistic is asymptotically chi-squared with one degree of freedom and takes the value 93.17 with *p-value* of zero. The pricing errors reported in Table VIII are strictly worse than those in Table VI for the exactly-identified case.

The implied volatilities in Table VII are much lower than those in Table V. We had naively expected that restricting the implied volatility to one value across moneyness for each maturity class would produce some sort of average of the implied volatilities obtained in the exactly-identified case. However, this is not the case. The objective function is plotted in Figure 2 and a simple grid search confirms the optimum reported in Table VII.

The reason for the unexpected solution to the over-identified case can be found in several interrelated factors. The exactly-identified estimation indicates that the Black-Scholes model fails largely because of the substantial volatility smile: different volatilities are needed to price options of different moneyness. However, the over-identified estimation forces all options to be priced using the same volatility. By forcing Black-Scholes pricing with a single volatility, the solution is heavily biased toward the solution found for the out-of-the-money options in the exactly-identified case. The reason is two-fold. First, if we imagine a plot of actual options prices as a function of index level, then this empirical plot is steeper than that obtainable for the Black-Scholes model for options of the same maturity because Black-Scholes does not allow for higher volatilities at higher stock prices. If you fit this Black-Scholes plot to the empirical plot via choice of implied volatility using a least squares criterion, you have to use volatilities close to those for the out-of-the-money options. This is because the in-the-money option prices are not very sensitive to volatility (they have low "vega") and are easily fitted, but to address the afore-mentioned steepness issue, lower volatilities are needed to pull the plot down to match the empirical data. Secondly, the GMM estimation by its nature puts more weight where there is more statistical confidence. The standard errors are lowest for the out-of-the-money options in Table V, and thus the GMM solution in the over-identified gravitates toward the GMM solution for the out-of-the-money options in the exactly-identified case (note the standard errors in Table VII are correspondingly very small).

For comparative purposes we estimate both the exactly- and over- identified cases using relative pricing error moments (i.e. $f = (c^{(\text{mkt})} - c^{(\text{BS})})/c^{(\text{mkt})}$). These results are reported in Tables IX and XI, with pricing errors reported in Table X and XII. The results are quite similar to the absolute pricing error moments case (and again over-identified pricing errors are worse than exactly-identified pricing errors). For further comparison we plot the objective functions for the over-identified estimations in the absolute pricing error moments case (Figure 2) and the relative pricing error moments case (Figure 3). The two plots have similarities and differences. In each case multiple local minima exist where a criss-cross of valleys intersect. The locations of these valleys correspond roughly to the solutions from the exactly-identified case (where individual moments are minimized). This is because the exactly-identified solutions set their respective average moments identically to zero, so other things being equal they are each attractive in their own right. The valleys are more clearly visible with varying $\hat{\sigma}_2$ (long maturity) than with varying $\hat{\sigma}_1$ (short maturity) because the prices of longer maturity options are more sensitive to choice of volatility (they have higher vega). We do not see any reason to favor absolute or relative pricing error moments – the model pricing is very similar (compare Tables VI and X).

With multiple local extrema visible in Figures 2 and 3, great care needs to be taken in the estimation. This is always a concern, but in an over-identified model it seems to us to be especially important – particularly if the model subsequently rejects. As a precaution we start our estimation routine from many different points. In the over-identified case our simple Newton estimation (Greene (1997, p203)) converges to the same solution from every starting point. However, when using BFGS (Press (1996, pp426-428)) the routine stops at several local minima. The Newton routine was slow (two hours), and BFGS was fast (15 minutes), but we had no confidence in the BFGS solutions. The explanation lies in the choices of step length. Our Newton routine was written by us and checked an unorthodox range of step lengths during its line searches but the more complex BFGS routine was an orthodox canned routine in MATLAB that looked for the closest local minimum. The tradeoff for us is a simple and slow-but-sure Newton routine, versus a fast BFGS routine that stops at every lamppost. Computing time is cheap, so we use the Newton routine. We have the luxury of working in only two dimensions, so we also execute grid searches (from 1% to 40% for each σ) that confirm our solutions quite quickly – more quickly in fact than the Newton routine. However, we think a thorough grid search ceases to be viable if there are four or more dimensions. Hints for handling multiple local optima appear in Table XIII.

Let us emphasize that the two approaches we present (exactly-identified model using Wald tests versus over-identified model using test of over-identifying restrictions) are not equally attractive in the option pricing framework. We discovered six advantages to using an exactly-identified model (see summary in Table XIV). *Firstly*, in the exactly-identified case we obtain much more information than the over-identified case. We get implied volatilities for each maturity and moneyness class and we can test to see how any of these differ from any other in any combination we choose. We can test various different hypotheses about the model and draw potentially diverse conclusions from these tests. However, in the over-identified case we obtain one implied volatility estimator only for the short maturity options, and one only for the long maturity options. We can test whether they differ and the test of over-identifying restrictions tells us that the volatility smile is not flat, but it does not tell us what shape the smile is.

The *second* reason we prefer exactly- to over-identified estimation is that the estimators obtained in the over-identified case need not be economically meaningful when the model is rejected. In our case, the implied volatilities in Tables V and IX (exactly-identified cases) represent the volatilities used by market participants in pricing the options, but those in Tables VII and XI do not. We reject the model in both cases, but only in the exactly-identified case do we learn about market participants' pricing.

The *third* reason we prefer exactly- to over-identified estimation is that the over-identified case converges very slowly.[12] We had to start the over-identified estimation with a GMM estimation where the weighting matrix is replaced by the identity matrix.[13] In this case the GMM procedure reduces to a least squares estimation. With the solution to this initial least squares estimation in hand as a starting point, the full GMM estimation was then run with weighting matrices estimated using the parameter estimates. In both the initial identity matrix case and the subsequent estimated weighting matrix case we had to allow for variable step lengths in the iterative solution technique.

The *fourth* reason we prefer exactly- to over-identified estimation is that in the exactly-identified case you have the supreme luxury of running the entire optimization using an identity weighting matrix. This is because the solution is independent of the weighting matrix in the exactly-identified case. Convergence takes only a few seconds whatever the routine (even a simple Newton routine with fixed step length). Once you locate the optimum you have to estimate the correct weighting matrix to calculate the standard errors. In our case we have 50 lags in our Newey-West routine and we appreciate not having to continually re-estimate the matrix.

The *fifth* problem with over-identified estimation is that as previously mentioned we can see clearly in both Figures 2, and 3 that the criss-cross of valleys representing solutions to the exactly-identified case produces multiple local minima for the GMM objective function. These multiple local minima needlessly complicate the optimization routine. The existence of multiple local extrema goes hand-in-hand with the rejection of the over-identified model. If the over-identified model could not be rejected, then its solution would match that of the exactly-identified model, and we would not see the competing solutions. This in turn goes hand-in-hand with the observation that when the over-identified model is rejected, the solution need not be economically meaningful. A summary of hints for handling local extrema appears in Table XIII.

The *sixth* and final reason that we prefer exactly- to over-identified estimation is that you will know whether your solution is a global minimum or not because you know *a priori* that such a solution will set the objective function identically to zero in the exactly-identified case. This is in stark contrast to the over-identified case where local extrema masquerade as global extrema.

It is up to the individual researcher to weigh up the costs and benefits of exactly- versus over-identified estimation. In our case, we are particularly interested in the volatility smile. We are able to test for the existence of the smile using an exactly-identified model together with Wald tests, or using an over-identified model plus the test of over-identifying restrictions. In our case, an exactly-identified model provides more informative and more meaningful results with a less torturous estimation routine than does an over-identified model. This is largely because it is more natural to let our implied volatility parameters roam freely and then test whether they differ than it is to force them to adhere to a model that may well be incorrect. However, we concede that there may be other applications where the benefits of over-identified estimation outweigh the costs but that is certainly not so here.

V Conclusion

Generalized Method of Moments (GMM) is not adequately explained in the finance literature. This lack of explanation creates a barrier to entry for researchers who might otherwise benefit from use of GMM. We address this shortfall by first using a simple example to explain how and why GMM works. We then address implementation issues by using GMM to estimate and test the famous Black-Scholes option pricing model using S&P500 index options data. We identify problem areas in implementation and we give tactical estimation advice and troubleshooting tips. We pay particular attention to proper choice of moment conditions, exactly-identified versus over-identified estimation, estimation of Newey-West standard errors, and numerical optimization in the presence of multiple local extrema.

A Data Appendix

Our initial sample comprises approximately 4.5 million CBOE SPX option market maker bid-ask quotes recorded during the 96 trading days from Jan 03 1995 to May 18 1995. We use filtering and screening to remove bad or needlessly repetitious data.

We first screen the data for any options that fail a basic arbitrage bound.[14] Approximately 1 in 27 options fail this bound. These failing call options are very deep in the money and are worth on average about \$80.00. They typically fail the bound by only about \$0.10. These results are consistent with Bates (1995, p11) and may be due partly to non-synchronous option and index quotes. Deep-in-the-money call option prices are not sensitive to volatility. With little or no vega they are easy to price accurately and excluding them will have little or no impact on our results.

We break our data up using 15 minute windows.[15] The window length is endogenous - it should be the shortest window possible that allows observation of representative prices on all option maturity and moneyness categories through the majority of the data set. If your particular sample involves heavy trading, then a shorter window length may be possible. If trading is slow, then a longer window length may be needed.

For each 15 minute interval we record one and only one bid-ask quote for each option series – the first. That is, if the DEC 350 SPX call (i.e. the S&P500 index call option with strike price 350 maturing on the Saturday following the third Friday in December) is quoted 6 times between 9:15AM and 9:30AM, we record only the first such quote and discard any repetitions. Each option quote goes hand-in-hand with a simultaneous record of the level of the SPX.

We next choose during each 15 minute window from amongst all the available quotes of various moneynesses and maturities. We choose the price (actually a bid-ask quote midpoint) of one SPX call option from each of three moneyness classes (out-, in-, and near-the-money), and each of two maturity classes (short- and long-dated), for a total of six categories of options during each 15 minute window. Options expiring in fewer than 88 days are labelled short term; the remainder are long term. Our longest-dated options expire within one year; we do not use LEAPS (very long-dated options). We cut at moneyness of $\frac{X}{F} = 0.92$ and $\frac{X}{F} = 0.98$, where $F = Se^{(r-q)\tau}$ is the forward price.[16] We label moneyness below 0.92 as "in-the-money," moneyness between 0.92 and 0.98 as "near-the-money," and moneyness above 0.98 is "out-of-the-money." These choices cut the data into relatively similar sized sextants.[17]

If several options with moneyness below 0.92 and maturity in excess of 88 days are quoted between 9:15AM and 9:30AM, then only one of these quotes is taken as representative of the long-dated in-the-money class during this 15 minute window. The selection is made by taking those options in that moneyness/maturity class in that 15 minute window that are of highest price (thereby avoiding any extremely low priced options where microstructure issues can be problematic (Bates (1995, p19))).[18] If more than one option exists in this class at that price, then the earliest quote is taken. This leaves at most one observation of each moneyness/maturity class during each 15 minute window.

We label the timing of these 15 minute windows as $t = 1, \ldots, T$. Regular trading hours (RTH) for the SPX contract are from 8:30AM to 3:15PM, so we get 27 observations per day on each of six categories of options. We use 96 days of data from Jan 03 1995 to May 18 1995, for a total of 2592 observation windows for each moneyness and maturity class. Of these 2592 windows, 710 contain a missing observation on at least one moneyness/maturity option quote. In 647 of these 710 cases we are able to replace the missing observation with a quote from earlier on the same day on an option of the appropriate moneyness and maturity class. Although we look to an earlier (presumably still outstanding) market maker quote for the missing option price observation, we still use a concurrent index level observation drawn from the original 15 minute window. After replacing missing observations, only 63 of the 2592 15 minute windows are excluded because of missing observations. This leaves a total sample size of 2529 observations on each of six moneyness/maturity classes. The missing data comprise less than 2.5% of our data set, and we think the effect must be negligible. Summary statistics for call option premia, term to maturity, and moneyness for the six moneyness/maturity classes are reported in Tables II, III, and IV respectively.

Each day during our sample period we hand collect WSJ T-bill bid and ask discount yields for weekly maturities out to one year. We adjust the recorded days to maturity for the one or two day discrepancy that the WSJ reports. We use the T-bill bid-ask spread midpoints to calculate continuously compounded annualised T-bill yields (Cox and Rubinstein (1985, p255)). For each and every option quote, we calculate the weighted average of the yields on the two T-bills of maturities that bracket the maturity of the option in question. The average yield is weighted according to the closeness of the option expiration dates to the expiration dates of the T-bills used. We use the T-bill quotes from the day closest to the day when the option quote was recorded. Thus, on any given day, we use different yields for options of different maturity, and these yields change every day. If there is a missing data point in the T-bill series, we use the midpoint of the previous and following days.

The market's anticipation of the annualised dividend yield q payable over the life of the option is deduced using put-call parity for SPX index options as in Equations (13).[19]

$$\begin{aligned} Se^{-q\tau} + p &= c + Xe^{-r\tau} \\ \Rightarrow e^{-q\tau} &= \frac{c + Xe^{-r\tau} - p}{S} \\ \Rightarrow q &= \frac{1}{\tau} \ln \left[\frac{S}{c + Xe^{-r\tau} - p} \right]. \end{aligned} \tag{13}$$

On each day and for each and every option maturity we locate all pairs of this maturity same-strike put and call option quotes that are within 20 points of the contemporaneous SPX level.[20] We take only pairs of quotes that are recorded within 15 minutes of each other, but the 15 minute window can be at absolutely any time during the trading day – it need not coincide with the CBOE 15 minute trading day divisions. From all of these pairs of put and call quotes for a given maturity, we choose the single pair that is closest to being at-the-money for that maturity. The SPX level recorded for this pair is the average of the levels when each of the two quotes are recorded respectively. In practice these index levels are virtually identical because of the afore-mentioned 15 minute restriction. We use at-the-money quotes because this is where most of the volume is and where we therefore have most confidence in the "implied dividend yields." This is the only put option data used in our analysis.[21] To each and every option quote we assign a dividend yield and a T-bill rate corresponding to the maturity of that option.

Note the distinction between our use of the word "maturity" and our use of the phrase "maturity class." The longest maturity class contains options of several different maturities and although we allow for dividend yields and T-bill rates to differ for each maturity, we fit only a single volatility number for each moneyness/maturity class.

B Pseudo-Code for GMM

```
% INITIAL DATA MANIPULATION
% load options and t-bill data
  load data=[day time-stamp call put S X tau r]
% infer dividend yields
  q=(1/tau).*log(S./(call-put+X.*exp(-r.*tau))
% screen for no-arbitrage violation
  F=S.*exp((r-q).*tau)
  failure=call<exp(-r.*tau).*(F-X).*(sign(log(F./X))+1)/2
  data(find(failure))=void

% FORM GMM MOMENTS AND OPTIMIZE
  theta=initialguess
  while update./theta>sqrt(my computer's floating point precision)[22]
    d=(log(S./X)+(r-q+(theta.∧2)/2).*tau)./(theta.*sqrt(tau))
    f=(c-(S.*exp(-q.*tau).*N(d)-...
    X.*exp(-r.*tau).*N(d-theta.*sqrt(tau))))'
    g=mean(f')'
    % FIND DERIVATIVE OF g W.R.T. THETA
    dfdth=S.*exp(-q.*tau).*(sqrt(tau)/sqrt(2*pi)).*exp(-((d.∧2)/2))
    dg=diag(mean(dfdth)) % same as Gamma
    % USE WHITE AND NEWEY-WEST TO FIND W=OMEGA
    WHITE=(1/T)*f*f'; NWEST=0
    m=50
    for j=1:m
       phij=(1/T)*f(:,j+1:T)*f(:,1:T-j)'
       NWEST=NWEST+(1-(j/(m+1)))*(phij+phij')
    end
    NWEST=NWEST+WHITE
    W=inverse(NWEST)
    update=inverse(dg'*W*dg)*(dg'*W*g)
    theta=theta+update
  end while loop

% CALCULATE STANDARD ERRORS AND DO TESTS
```

```
  OMEGA=WHITE
  VGMM=inv(Gamma'*inv(OMEGA)*Gamma)
  SEWHITE=sqrt(diag(VGMM)/T)

  OMEGA=NWEST
  VGMM=inv(Gamma'*inv(OMEGA)*Gamma)
  SENWEST=sqrt(diag(VGMM)/T)
  print [theta SEWHITE SENWEST]

% TEST VOLATILITY SMILE
  R1=[...
  1 -1 0 0 0 0
  0 1 -1 0 0 0
  0 0 0 1 -1 0
  0 0 0 0 1 -1]
  testsmile=(R1*theta)'*inv(R1*VGMM*R1'/T)*(R1*theta)
  pvalue=1-cdf('chi2',testskew,#rows(R))]
  print [testsmile pvalue]

% TEST VOLATILITY TERM STRUCTURE
  R2=[...
  1 0 0 -1 0 0
  0 1 0 0 -1 0
  0 0 1 0 0 -1]
  testterm=(R2*theta)'*inv(R2*VGMM*R2'/T)*(R2*theta)
  pvalue=1-cdf('chi2',testterm,#rows(R2))
  print [testterm pvalue]
```

C Footnotes

1. The existence of multiple roots to Equation (2) could cause problems for an optimizing algorithm. However, comparison of $\hat{\nu}^{(2)}$ to $\hat{\nu}^{(1)}$ should enable the correct root to be located. Thus, the use of both moments becomes necessary in this case.

2. We shall see in Section IV that you sometimes need to begin a GMM optimization routine with an identity weighting matrix to get initial values for the full estimation.

3. The reader can confirm that Equation (5) shares the same property: multiply it by T and it is asymptotically chi-squared – with one degree of freedom.

4. Note also a local minimum visible in Figure 1 at approximately $\nu = 3$. We discuss over-identified estimation and the existence of local extrema in Section IV.

5. Stationarity is stronger than "identically distributed," but weaker than IID, since stationarity does not imply independence. Ergodicity is weaker than independence – it is a form of average asymptotic indepen-

dence that restricts dependence or memory in a sequence. Stationary and ergodicity together are strictly weaker than IID. See White (1984, pages 41-46) for details.

6. The ij^{th} element of Ω is given by $E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} f_i(X_t,\beta_0)f_j(X_s,\beta_0)\right]$, where f_i, and f_j are the i^{th} and j^{th} elements respectively of the vector f.

7. Had we used the S&P100 stock index (OEX) options contract, then its American style exercise means that Black-Scholes would not be appropriate.

8. This is a slight abuse of the term "implied volatility." The volatility number we infer does not match market prices of options exactly to model prices. Rather, it is a best match on average across many options.

9. If you are estimating the mean, variance, skewness, kurtosis, and other simple parameters then not only is the GMM estimation exactly identified, but no numerical optimization technique is required because the traditional estimators set the mean of the GMM moments to zero. In

this case, the optimal weighting matrix is not needed in the optimization, but is still used to calculate standard errors.

10. These numbers are reassuringly close to the level of the VIX index which varied from a low of 10.49 to a high of 13.77 during our data period Jan 03 1995 to May 18 1995 (source: Bloomberg terminal). The VIX is a weighted average of implied volatilities on eight OEX (S&P100) index options.

11. Vega for dividend-adjusted Black-Scholes is $Se^{-q\tau}\sqrt{\tau}\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}}$, with d_1 as in Equations (11). The mean values of vega for our moneyness/maturity class are 2.51, 30.14, and 65.20 for the short term, and 14.36, 100.74, and 138.40 for the long term.

12. This can be a symptom of scaling problems. Be sure that your parameters are scaled to be of roughly the same magnitude. Radically different orders of magnitude can confound estimation because of computer arithmetic precision problems.

13. Near the beginning of the optimization the estimator of the VCV of the moments is very sensitive to the parameter values. Either replace

it with the identity matrix and iterate to convergence, or use numerical derivatives for a few iterations.

14. We reject the quote if $c < (F - X)e^{r\tau}\left\{\frac{\mathrm{sgn}\left[\ln\left(\frac{F}{X}\right)\right]+1}{2}\right\}$, where c is the ask price for the call, $F = Se^{(r-q)\tau}$ is the forward price, and $\mathrm{sgn}(x)$ is +1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. The RHS of the inequality is $\lim_{\sigma \to 0} c^{(\mathrm{BS})}(\sigma)$. This is the only place in our analysis where we use raw ask prices instead of bid-ask spread mid-points. We do so because any trading strategy that exploits this opportunity would have to buy the call at the ask (Bhattacharya (1983, p164)). Any call option price that fails this test falls below the lower no-arbitrage bound and would – if priced via Black-Scholes – have no real implied volatility (the solution would be a complex number).

15. The official CBOE trading day is similarly broken up, with a loud buzzer sounding every 15 minutes on the floor.

16. Quoting moneyness via comparison of strike and forward prices is standard practice (Bates (1995, p8)).

17. Using 0.97 and 1.03 as moneyness cutoffs produced very similar results

and the same inferences at the cost of twice as many missing observations. Cutting the data into roughly equal sized portions seems more natural from the standpoint of describing the data and minimizing the number of missing observations. One consequence is that our "near-the-money" options are not the ones with highest vega: because we cut moneyness with respect to forward price at 0.98, this means that the "out-of-the-money options are only marginally out-of-the-money with respect to spot price, and they have the highest vega (Cox and Rubinstein (1985, p225)).

18. We repeated the entire analysis using lowest priced instead of highest priced options here and the inferences were unchanged. In fact, most standard errors reduced slightly and test statistics were correspondingly larger. However, the percentage pricing errors on the out-of-the-money options blew up to 40-50% because the options were so low priced.

19. Alternatively, we could deduce dividends by using $F = Se^{(r-q)\tau}$, where τ is option life, r is T-bill yield, q is dividend yield to be deduced, S is spot level of the index, and F is SPX futures price for a futures contract of life closest to that of the option. SPX futures (on the CME) expire

on the Thursday immediately before the third Friday of the month, but SPX options (on the CBOE) expire on the Saturday following.

20. The SPX index level varied from about 450 to about 530 during our sample. These options are therefore at worst about 5% away from the money, but in practice the minimization of deviations from moneyness means that almost all are right at-the-money.

21. One unexpected result is the very significant term structure of implied dividend yields. For almost every day in the sample the value of q inferred from put-call parity was increasing in maturity of the options. In a few cases the very shortest maturity options actually have slightly negative q values associated with them. We do not report the details here.

22. See discussion of multiplicative tolerance factors in Press *et al.* (1996, p398, p410).

References

Bachelier, L., 1900, Theory of speculation, in Paul Cootner, ed.: *The Random Character of Stock Market Prices* (Massachusetts Institute of Technology Press, Cambridge, MA, 1964).

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003-2049.

Bates, David S., 1995, Testing option pricing models, Working Paper, The Wharton School, University of Pennsylvania.

Berndt, Ernst K., Bronwyn Hall, Robert Hall, and Jerry A. Hausman, 1974, Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement*, Vol 3 No 4, (October), 653-665.

Bhattacharya, Mihir, 1983, Transactions data tests of efficiency of the Chicago Board Options Exchange, *Journal of Financial Economics*, Vol 12 No 2, (August), 161-185.

Black, F., 1975, Fact and fantasy in the use of options, *Financial Analysts Journal*, Vol 31 No 4, (July-August) 36-41, 61-72.

Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, Vol 81 No 3, (May/June), 637-659.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

Cox, John C. and Mark Rubinstein, 1985, *Options Markets*, Prentice-Hall: New Jersey, USA.

Evans, Merran, Nicholas Hastings, and Brian Peacock, 1993, *Statistical Distributions*, (John Wiley and Sons, New York, NY).

Fama, Eugene F., 1965, The behavior of stock market prices, *Journal of Business* 38, 34-105.

Greene, William H., 1997, *Econometric Analysis*, Third Edition, Prentice Hall: Upper Saddle River, NJ.

Hamilton, James D., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Hansen, L, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.

Hull, John C., 1997, *Options, Futures, and Other Derivatives*, Third Edition, Prentice-Hall: New Jersey.

MacBeth, James, and Larry Merville, 1979, An empirical examination of the Black-Scholes call option pricing model, *Journal of Finance*, Vol 34, 1173-1186.

Melino, Angelo and Stuart M. Turnbull, 1990, Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, Vol 45 No 1/2, (July/August), 239-265.

Merton, Robert C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, (Spring), 141-183.

Murphy, Gareth, 1994, When options price theory meets the volatility smile, *Euromoney*, No 299, (March), 66-74.

Newey, W., and K. West, 1987, A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Ogaki, Masao, 1993, Generalized method of moments: Econometric applications, Ch 17 in Maddala, Rao and Vinod, eds.: *Handbook of Statistics*, Vol 11.

Press, William H., Saul Teukolsky, William T. Vetterling, and Brian P. Flannery, 1996, *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition, Cambridge University Press, Cambridge UK.

Scott, Louis O., 1987, Option pricing when the variance changes randomly: Theory, estimation, and an application, *Journal of Financial and Quantitative Analysis*, Vol 22 No 4, (December), 419-438.

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817-838.

Wiggins, James B., 1987, Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics*, Vol 19 No 2, (December), 351-372.

D Tables

Parameter	Simulation Results			
T	100	1,000	10,000	100,000
$\hat{\nu}^{(1)}$	6.030	9.252	9.475	10.491
$\hat{\nu}^{(2)}$	9.164	10.408	10.561	10.392
$\hat{\nu}_{GMM}$	14.899	11.337	10.439	10.481
SE_{WHITE}	9.422	1.745	0.620	0.240
SE_{NW}	9.146	1.742	0.621	0.239
χ_1^2	3.075	1.184	**7.259**	0.142
p-value	0.080	0.277	0.007	0.706

Table I: Simulated Student-t Data.
For sample size T we simulate IID Student-t data with $\nu = 10$ degrees of freedom. The MOM estimators $\hat{\nu}^{(1)}$, and $\hat{\nu}^{(2)}$ are reported along with the GMM estimator $\hat{\nu}_{GMM}$. Both White and Newey-West standard errors are reported along with the chi-squared goodness of fit test (critical 5% value 3.842).

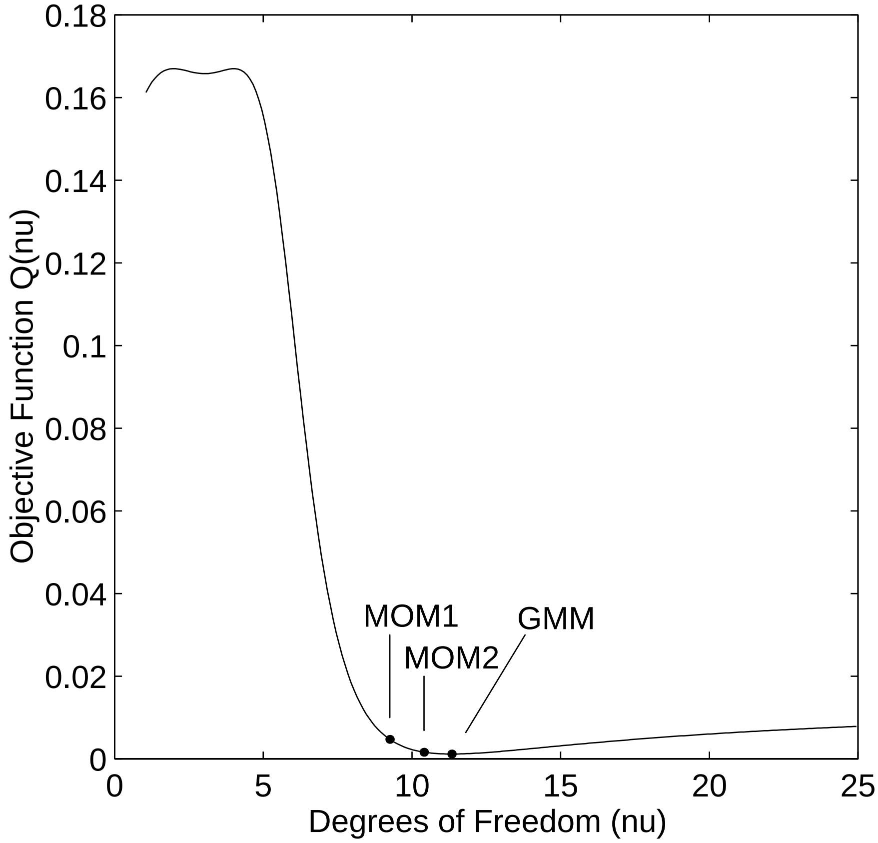

Figure 1: Objective Function for Simulated Student-t data.
For 1,000 independent drawings from a Student-t with $\nu = 10$ degrees of freedom, we calculate the objective function $Q(\hat{\nu}) = g(\hat{\nu})'Wg(\hat{\nu})$ where $W = \hat{\Omega}^{-1}$ is the estimated asymptotic VCV of $g(\nu)$ (estimated using the Newey-West technique described later in this paper). Our original two MOM estimators are $\hat{\nu}^{(1)} = 9.252$ and $\hat{\nu}^{(2)} = 10.408$ (labelled "MOM1," and "MOM2"). The objective function is minimized at the GMM estimator $\hat{\nu} = 11.337$ (labelled "GMM").

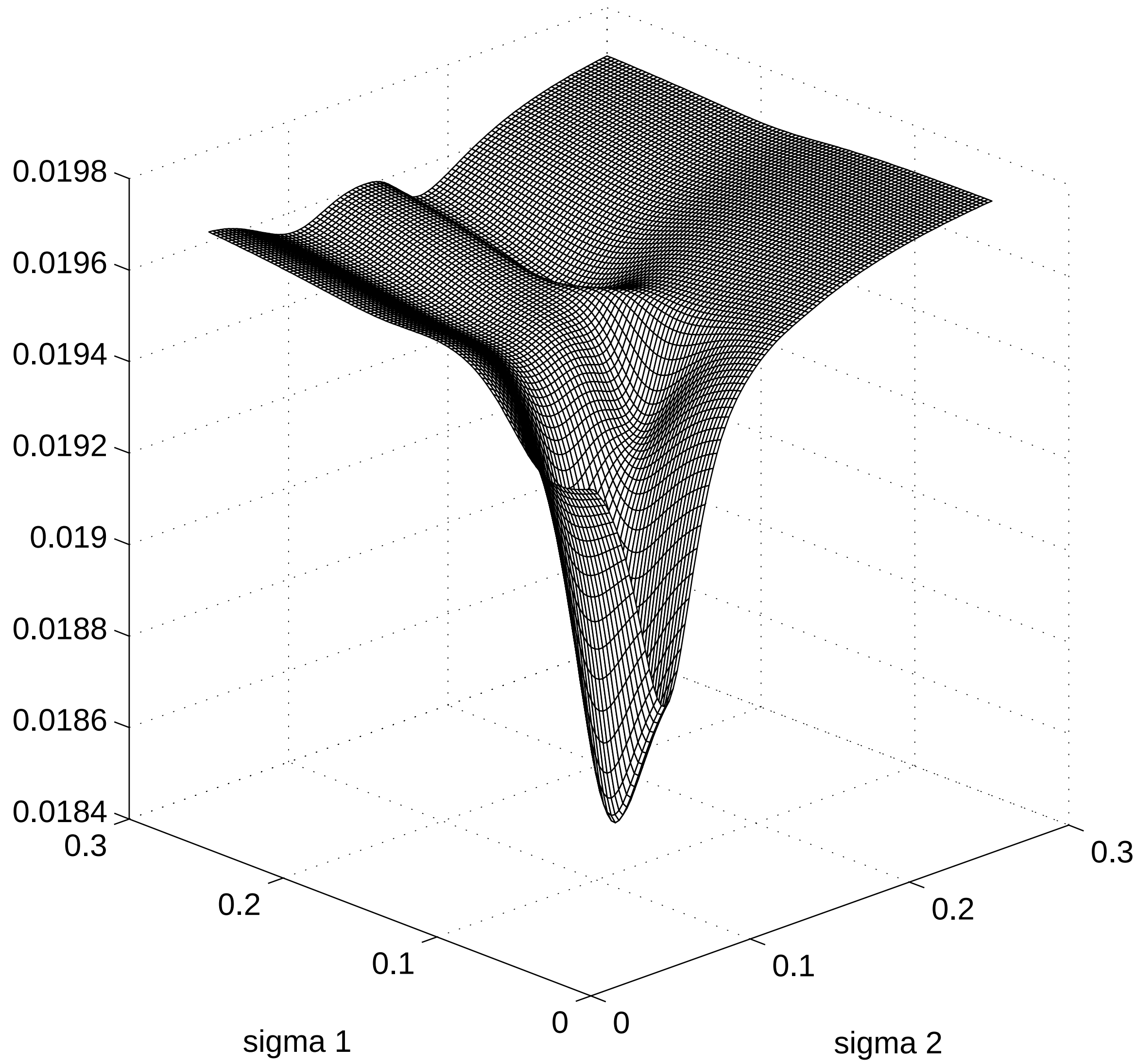

Figure 2: GMM Objective Function (Over-Identified Case, Absolute Pricing Error Moments).
The vertical axis is the objective function $g'Wg$ when we restrict the implied volatilities to be flat across moneyness for each maturity class. A simple grid search confirms the optimum reported in Table VII: $\hat{\sigma}_1 = 8.38\%$, and $\hat{\sigma}_2 = 9.57\%$. There are 2529 CBOE SPX option price observations drawn at 15 minute intervals from Jan 03 1995 to May 18 1995 for each of six moneyness/maturity classes.

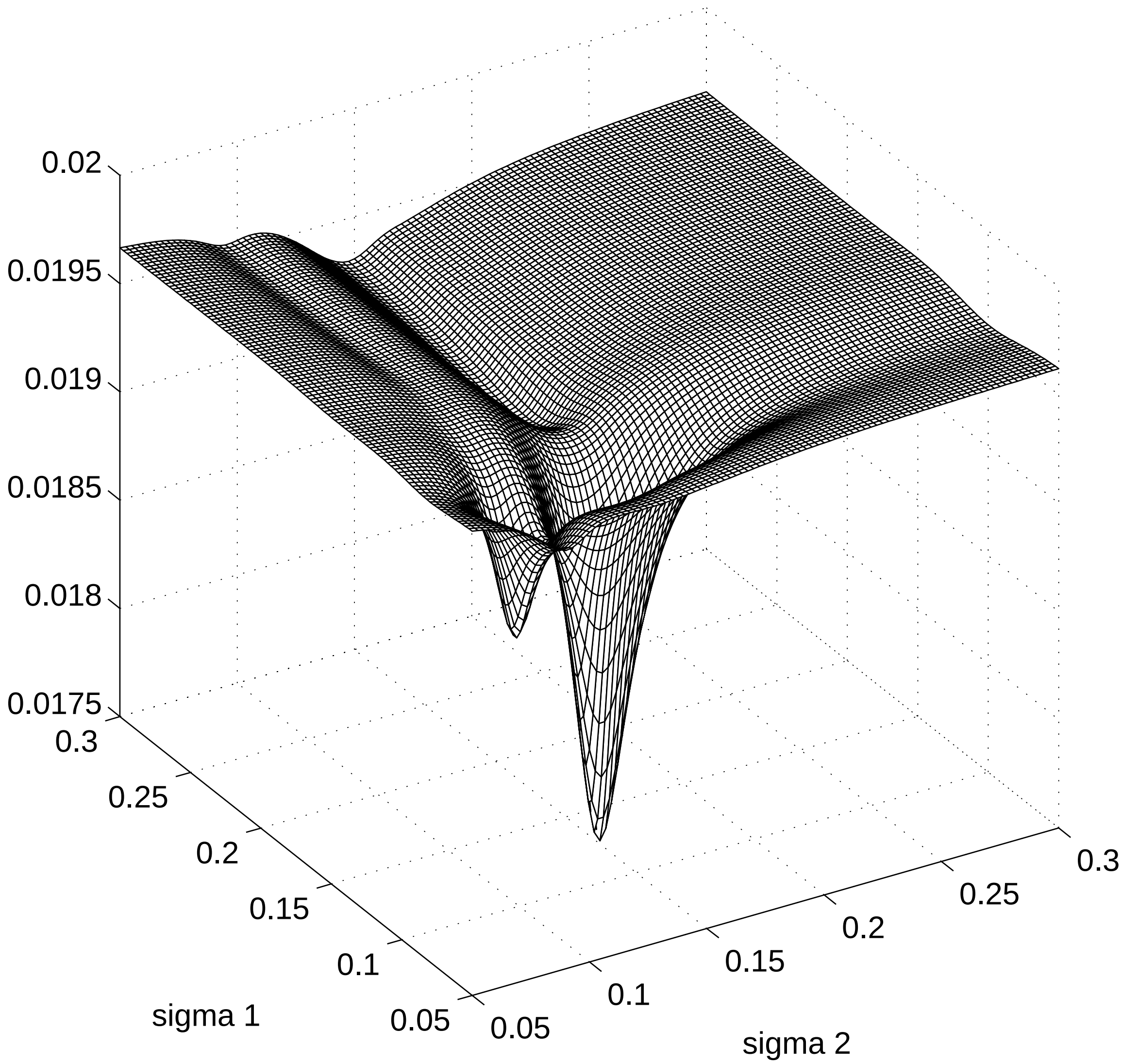

Figure 3: GMM Objective Function (Over-Identified Case, Relative Pricing Error Moments).
The vertical axis is the objective function $g'Wg$ when we restrict the implied volatilities to be flat across moneyness for each maturity class. A simple grid search confirms the optimum reported in Table XI: $\hat{\sigma}_1 = 10.57\%$, and $\hat{\sigma}_2 = 13.77\%$. There are 2529 CBOE SPX option price observations drawn at 15 minute intervals from Jan 03 1995 to May 18 1995 for each of six moneyness/maturity classes.

Summary Statistics for Call Option Premia (Dollars)			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Minimum	37.5	10.4	0.4
Maximum	179.5	201.5	48.8
Median	121.5	38.8	12.1
Mean	**121.6**	**38.1**	**12.0**
Std. Dev.	40.4	5.8	3.4
# Obs. (15 min. windows)	2529	2529	2529
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Minimum	42.5	16.9	0.2
Maximum	182.0	80.1	110.3
Median	137.1	42.8	19.0
Mean	**126.2**	**42.0**	**16.7**
Std. Dev.	38.9	5.1	7.1
# Obs. (15 min. windows)	2529	2529	2529

Table II: Summary Statistics for Call Option Premia

These summary statistics are for observations of call option premia (bid-ask quotes midpoints in dollars) for the six moneyness/maturity classes observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. We cut at moneyness of $\frac{X}{F} = 0.92$ and $\frac{X}{F} = 0.98$, where $F = Se^{(r-q)\tau}$ is the forward price. Means are in bold for easy reference.

Summary Statistics for Call Option Term to Maturity (Calendar Days)			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Minimum	2.0	2.0	2.0
Maximum	87.0	87.0	87.0
Median	61.0	58.0	53.0
Mean	**56.3**	**51.1**	**50.8**
Std. Dev.	22.3	25.0	22.4
# Obs. (15 min. windows)	2529	2529	2529
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Minimum	92.0	88.0	88.0
Maximum	347.0	340.0	347.0
Median	220.0	205.0	221.0
Mean	**201.5**	**191.4**	**211.5**
Std. Dev.	63.8	60.6	65.6
# Obs. (15 min. windows)	2529	2529	2529

Table III: Summary Statistics for Call Option Term to Maturity

These summary statistics are for observations of call option maturities (measured in calendar days) for the six moneyness/maturity classes observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. We cut at moneyness of $\frac{X}{F} = 0.92$ and $\frac{X}{F} = 0.98$, where $F = Se^{(r-q)\tau}$ is the forward price. Means are in bold for easy reference.

Summary Statistics for Call Option Moneyness $\left(\frac{X}{F},\ F = Se^{(r-q)\tau}\right)$			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Minimum	0.66	0.92	0.98
Maximum	0.92	0.98	1.09
Median	0.74	0.92	0.98
Mean	**0.75**	**0.93**	**0.99**
Std. Dev.	0.08	0.01	0.01
# Obs. (15 min. windows)	2529	2529	2529
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Minimum	0.65	0.92	0.98
Maximum	0.92	0.98	1.22
Median	0.72	0.93	0.99
Mean	**0.74**	**0.93**	**1.01**
Std. Dev.	0.08	0.01	0.03
# Obs. (15 min. windows)	2529	2529	2529

Table IV: Summary Statistics for Call Option Moneyness

These summary statistics are for observations of call option moneyness (measured as $\frac{X}{F}$ where $F = Se^{(r-q)\tau}$) for the six moneyness/maturity classes observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. We cut at moneyness of $\frac{X}{F} = 0.92$ and $\frac{X}{F} = 0.98$. Means are in bold for easy reference.

	In-the-money	Near-the-money	Out-of-the-money
Short Maturity	0.1686 (0.0109)	0.1593 (0.0033)	0.1127 (0.0013)
Long Maturity	0.1629 (0.0043)	0.1428 (0.0010)	0.1162 (0.0010)
$test_{smile} = 1540.16$, *p-value*=0			
$test_{term} = 62.36$, *p-value*=0			

Table V: Implied Volatilities (Exactly-Identified Case, Absolute Pricing Error Moments).
For absolute pricing error moments (i.e. $f = c^{(\mathrm{mkt})} - c^{(\mathrm{BS})}$) these are implied volatilities on options observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. We cut at moneyness of $\frac{X}{F} = 0.92$ and $\frac{X}{F} = 0.98$, where $F = Se^{(r-q)\tau}$ is the forward price. Newey-West standard errors with 50 lags are in parentheses (more lags had little effect). To form test statistics based on the above numbers, we set

$$\hat{\theta} = [0.1686\ 0.1593\ 0.1127\ 0.1629\ 0.1428\ 0.1162]',$$

and form restriction matrices

$$R_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \text{and} \quad R_2 = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

We test for the volatility smile using

$$test_{smile} = (R_1\hat{\theta})'[R_1\widehat{V_{GMM}}R_1'/T]^{-1}(R_1\hat{\theta}).$$

We test for the term structure of volatility using

$$test_{term} = (R_2\hat{\theta})'[R_2\widehat{V_{GMM}}R_2'/T]^{-1}(R_2\hat{\theta}).$$

In each case $\widehat{V_{GMM}}$ is the estimated asymptotic VCV of $\hat{\theta}$ (hence our division by T). The test statistics are Wald tests asymptotically chi-squared with four and three degrees of freedom respectively.

Average Pricing Error for Exactly-Identified GMM Estimation			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.35	\$0.67	\$0.62
Average Relative Error	0.4%	1.7%	5.4%
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.59	\$0.73	\$1.19
Average Relative Error	0.6%	1.8%	15.7%

Table VI: Average Pricing Errors (Exactly-Identified Case, Absolute Pricing Error Moments)
These are the average dollar and percentage option pricing errors for the GMM estimation in the exactly-identified case where the moments use absolute pricing errors (i.e. $f = c^{(\mathrm{mkt})} - c^{(\mathrm{BS})}$). More formally, "Average Absolute Error" is the mean of the absolute value of the moments: $\frac{1}{T}\sum |f_t|$ (the mean of the raw moments is zero in the exactly-identified case). "Average Relative Error" is the mean of the absolute value of the moments that have been scaled by the option premia: $\frac{1}{T}\sum \left|\frac{f_t}{c_t}\right|$.

	All Moneyness Classes
Short Maturity	0.0838 (0.0012)
Long Maturity	0.0957 (0.0010)
$test_{term} = 93.17$, *p-value*=0	
$\chi^2 = Tg'\Omega^{-1}g = 46.92$, *p-value*=0	

Table VII: Implied Volatilities (Over-Identified Case, Absolute Pricing Error Moments).
For absolute pricing error moments (i.e. $f = c^{(\mathrm{mkt})} - c^{(\mathrm{BS})}$) these are implied volatilities on options observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. Newey-West standard errors with 50 lags are in parentheses (more lags had little effect). The test statistic $test_{term}$ is chi-squared with one degree of freedom under the null. The chi-squared test of over-identifying restrictions is chi-squared with 4 degrees of freedom.

Average Pricing Error for Over-Identified GMM Estimation			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.36	\$1.42	\$1.85
Average Relative Error	0.4%	3.6%	14.4%
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.89	\$4.18	\$2.93
Average Relative Error	1.1%	9.8%	17.7%

Table VIII: Average Pricing Errors (Over-Identified Case, Absolute Pricing Error Moments)

These are the average dollar and percentage option pricing errors for the GMM estimation in the over-identified case where the moments use absolute pricing errors (i.e. $f = c^{(\mathrm{mkt})} - c^{(\mathrm{BS})}$). More formally, "Average Absolute Error" is the mean of the absolute value of the moments: $\frac{1}{T}\sum|f_t|$ (the mean of the raw moments is zero in the exactly-identified case). "Average Relative Error" is the mean of the absolute value of the moments that have been scaled by the option premia: $\frac{1}{T}\sum\left|\frac{f_t}{c_t}\right|$.

	In-the-money	Near-the-money	Out-of-the-money
Short Maturity	0.1604 (0.0069)	0.1540 (0.0022)	0.1097 (0.0010)
Long Maturity	0.1584 (0.0034)	0.1420 (0.0011)	0.1075 (0.0014)
$test_{smile} = 1833.65$, *p-value*=0			
$test_{term} = 36.50$, *p-value*=0			

Table IX: Implied Volatilities (Exactly-Identified Case, Relative Pricing Error Moments).

For relative pricing error moments (i.e. $f = (c^{(\mathrm{mkt})} - c^{(\mathrm{BS})})/c^{(\mathrm{mkt})}$) these are implied volatilities on options observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. We cut at moneyness of $\frac{X}{F} = 0.92$ and $\frac{X}{F} = 0.98$, where $F = Se^{(r-q)\tau}$ is the forward price. Newey-West standard errors with 50 lags are in parentheses (more lags had little effect).

Average Pricing Error for Exactly-Identified GMM Estimation			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.34	\$0.64	\$0.62
Average Relative Error	0.34%	1.6%	5.14%
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.37	\$0.73	\$1.71
Average Relative Error	0.56%	1.84%	14.36%

Table X: Average Pricing Errors (Exactly-Identified Case, Relative Pricing Error Moments)
These are the average dollar and percentage option pricing errors for the GMM estimation in the exactly-identified case where the moments use relative pricing errors (i.e. $f = (c^{(\mathrm{mkt})} - c^{(\mathrm{BS})})/c^{(\mathrm{mkt})}$). More formally, "Average Absolute Error" is the mean of the absolute value of the moments scaled by the option premia: $\frac{1}{T}\sum|f_t \times c_t|$ (the mean of the raw moments is zero in the exactly-identified case). "Average Relative Error" is the mean of the absolute value of the moments: $\frac{1}{T}\sum|f_t|$.

	All Moneyness Classes
Short Maturity	0.1057 (0.0008)
Long Maturity	0.1377 (0.0009)
$test_{term} = 1415.90$, *p-value*=0	
$\chi^2 = Tg'\Omega^{-1}g = \;\; 44.65$, *p-value*=0	

Table XI: Implied Volatilities (Over-Identified Case, Relative Pricing Error Moments).
For relative pricing error moments (i.e. $f = (c^{(\mathrm{mkt})} - c^{(\mathrm{BS})})/c^{(\mathrm{mkt})}$) these are implied volatilities on options observed during 2592 15 minute windows over 96 days from Jan 03 1995 to May 18 1995. 63 of these 15 minute windows are dropped because of missing observations. This gives a total sample size of 2529 15 minute windows. The maturity class is cut at 88 days. Newey-West standard errors with 50 lags are in parentheses (more lags had little effect).

Average Pricing Error for Over-Identified GMM Estimation			
Short-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.36	\$1.22	\$0.69
Average Relative Error	0.4%	3.0%	5.4%
Long-Maturity	In-the-money	Near-the-money	Out-of-the-money
Average Absolute Error	\$0.64	\$0.87	\$3.20
Average Relative Error	0.7%	2.1%	37.6%

Table XII: Average Pricing Errors (Over-Identified Case, Relative Pricing Error Moments)

These are the average dollar and percentage option pricing errors for the GMM estimation in the over-identified case where the moments use relative pricing errors (i.e. $f = (c^{(\mathrm{mkt})} - c^{(\mathrm{BS})})/c^{(\mathrm{mkt})}$). More formally, "Average Absolute Error" is the mean of the absolute value of the moments scaled by the option premia: $\frac{1}{T}\sum|f_t \times c_t|$ (the mean of the raw moments is zero in the exactly-identified case). "Average Relative Error" is the mean of the absolute value of the moments: $\frac{1}{T}\sum|f_t|$.

Hints for Handling Local Extrema
• Use many starting points
• Upon convergence restart routines like DFP or BFGS that start with $W = I$ and update W.
• Allow for unorthodox step lengths that look far beyond nearby extrema during line searches.
• We prefer a simple Newton routine with unorthodox step lengths (slow but sure) to a canned higher-tech BFGS (which is fast but finds local extrema).

Table XIII: Advice for Handling Local Extrema in Numerical Optimization of Objective Functions

In the table "DFP" is the Davidon-Fletcher-Powell hill climber and "BFGS" is the Broyden-Fletcher-Goldfarb-Shanno hill climber (Press (1996, pp425-428)).

Why we Prefer Exactly- to Over-Identified Estimation
• Exactly-identified estimation yields more information (e.g. a $\sigma_{i,j}$ for each moneyness/maturity class).
• Exactly-identified estimation allows tests of many different specific hypotheses (e.g. $H_0: \ \sigma_{i,1} = \sigma_{i,2} = \sigma_{i,3}$).
• Over-identified estimation produces parameters that are not necessarily economically meaningful if model rejects, and this goes hand-in-hand with the existence of multiple local minima in the objective function.
• Exactly-identified estimation converges very quickly using any hill climber. Over-identified estimation converges very slowly with all hill climbers.
• Exactly-identified estimation can be run with $W = I$ because solution is independent of weighting matrix. You save time by not re-computing W via Newey-West at each iteration. Also decreases complexity of code.
• You know if solution to exactly-identified estimation is global minimum because you know *a priori* that objective function is zero at global optimum – though uniqueness is not guaranteed.
• You cannot game the exactly-identified estimation as you can the over-identified estimation.

Table XIV: Why we Prefer Exactly- to Over-Identified Estimation
This is a summary of why we prefer exactly- to over-identified estimation in an option pricing context. In the table "W" is the GMM weighting matrix. The null hypothesis $H_0: \ \sigma_{i,1} = \sigma_{i,2} = \sigma_{i,3}$ states that the volatility smile is flat. That is, $\sigma_{i,j}$ is constant across moneyness classes j for each maturity class i. Two strikes against exactly-identified estimation is that it eats degrees of freedom and that over-identifying may increase power by optimally combining multiple moments to estimate a parameter. The former is not a big problem because GMM uses asymptotic results so you need plenty of data anyway.