Generalized Method of Moments Estimation: A Time Series Perspective

Lars Peter Hansen

(University of Chicago, Chicago, Illinois, USA) *

February 20, 2001

Abstract

This entry describes empirical methods for estimating dynamic economic systems using time-series data. By design, the methods target specific feature of the dynamic system and do not require a complete specification of the time-series evolution. The resulting generalized-method-of-moments estimation and inference methods use estimating equations implied by some components of a dynamic economic system. This entry describes the statistical methods and some applications of these methods.

1 Introduction

In many empirical investigations of dynamic economic systems, statistical analysis of a fully-specified stochastic process model of the time series evolution is too ambitious. Instead it is fruitful to focus on specific features of the time series without being compelled to provide a complete description. This leads to the investigation of partially-specified dynamic models. For instance, the linkages between consumption and asset returns can be investigated without a full description of production and capital accumulation. The behavior of monetary policy can be explored without a complete specification of the macroeconomic economy. Models of inventory behavior can be estimated and tested without requiring a complete characterization of input and output prices. All of these economic models imply moment conditions of the form:

$$Ef(x_t, \beta_o) = 0 \qquad (1)$$

where f is known *a priori*, x_t is an observed time series vector and β_o is an unknown parameter vector. These moment relations may fall short of providing a complete depiction

of the dynamic economic system. *GMM* estimation, presented in Hansen (1982), aims to estimate the unknown parameter vector β_o and test these moment relations in a computationally tractable way. This entry first compares the form of a *GMM* estimator to closely related estimators from the statistics literature. It then reviews applications of these estimators to partially-specified models of economic time series. Finally, it considers *GMM* related moment-matching problems in fully specified models economic dynamics.

2   Minimum Chi-square Estimation

To help place *GMM* estimation in a statistical context, I explore a closely related minimum chi-square estimation method. Statisticians developed minimum chi-square estimators to handle restricted models of multinomial data and a variety of generalizations. Neyman (1949) and Burankin and Gurland (1951), among others, aimed to produce statistically efficient and computationally tractable alternatives to maximum likelihood estimators. In the restricted multinomial model, estimators are constructed by forming empirical frequencies and minimizing Pearson's chi-square criterion or some modification of it.

The method has direct extensions to any moment-matching problem. Suppose that $\{x_t\}$ is a vector process, which temporarily is treated as being *iid*. Use a function ψ with n coordinates to define target moments associated with the vector x_t. A *model* takes the form:

$$E\left[\psi(x_t)\right] = \phi(\beta_o)$$

where β_o is an unknown parameter. The moment-matching problem is to estimate β_o by making the empirical average of $\{\psi(x_t)\}$ close to its population counterpart $\phi(\beta_o)$:

$$\min_{\beta} \frac{1}{T} \sum_{t=1}^{T} \left[\psi(x_t) - \phi(\beta)\right]' V \sum_{t=1}^{T} \left[\psi(x_t) - \phi(\beta)\right] \tag{2}$$

where V is the distance or weighting matrix. The distance matrix sometimes depends on the data and/or the candidate parameter vector β.[1]

The limiting distribution of the criterion (2) is chi-square distributed with n degrees of freedom if the parameter vector β_o is known, and if V is computed by forming the inverse of either the population or sample covariance matrix of $\psi(x)$. When β is unknown, estimates may be extracted by minimizing this chi-square criterion; hence the name. To preserve the chi-square property of the minimum (with an appropriate reduction in the degrees of freedom), we again form the inverse sample covariance matrix of $\psi(x)$, or form the inverse population covariance matrix for each value of β. The minimized chi-square property of the criterion may be exploited to build tests of over-identification and to construct confidence sets for parameter values. Results like these require extra regularity conditions, and this rigor is supplied in some of the cited papers.

---

[1]Our use of V to denote a weighting matrix that may actually depend on parameters or data is an abuse of notation. I use this simple notation because it is the probability limit of the weighting matrix evaluated at the parameter estimator that dictates the first-order asymptotic properties.

While the aim of this research was to form computationally tractable alternatives to maximum likelihood estimation, critical to statistical efficiency is the construction of a function ψ of the data that is a sufficient statistic for the parameter vector (see Burankin and Gurland (1951)).[2] Many distributions fail to have a finite number of sufficient statistics; but the minimum chi-square method continue to produce consistent, asymptotically normal estimators provided that identification can be established.

GMM estimators can have a structure very similar to the minimum chi-square estimators. Notice that the core ingredient to the moment-matching problem can be depicted as (1) with a separable function f:

$$f(x, \beta) = \psi(x) - \phi(\beta)$$

used in the chi-square criterion function. Target moments are one of many ways for economists to construct inputs into chi-square criteria, and it is important to relax this separability. Moreover, in *GMM* estimation, the emphasis on statistical efficiency is weakened in order to accommodate partially specified models. Finally, an explicit time series structure is added, when appropriate.

3   GMM Estimation

Our treatment of *GMM* estimation follows Hansen (1982), but it builds from Sargan (1958) and Sargan (1959) analyses of linear and nonlinear instrumental variables.[3] (See the entry on *Instrumental Variables in Economics and Statistics*.) *GMM* estimators are constructed in terms of a function f that satisfies (1) where f has more coordinates, say n, than there are components to the parameter vector β_o.[4] In later sections we will give examples of the construction of the f function, including ones that are not separable in x and β and ones for which there have more coordinates than parameter vectors. A minimum chi-square type criterion is often employed in *GMM* estimation. For instance, it is common to define the *GMM* estimator as the solution to:

$$b_T = \arg \min_{\beta} T g_T(\beta)' V g_T(\beta)$$

where

$$g_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} f(x_t, \beta)$$

---

[2]Berkson (1944) and Taylor (1953) generalize the minimum chi-square approach by taking a smooth one-to-one function h and building a quadratic form on $h\left[\frac{1}{T} \sum_t \phi(x_t)\right] - h[\psi(\beta)]$.

[3]See Ogaki (1993) for a valuable discussion of the practical implementation of *GMM* estimation methods.

[4]Another related estimation method is *M* estimation. *M*-estimation is a generalization of maximum likelihood and least squares estimation. *M*-estimators are typically designed to be less sensitive to specific distributional assumptions. (See the entry on *Robustness, in statistics*.) These estimators may be depicted as solving a sample counterpart to (1) with a function f that is nonseparable, but with the same number of moment conditions as parameter estimators.

and V is a positive definite weighting matrix. This quadratic form has the chi-square property provided that V is an estimator of the inverse of an appropriately chosen covariance matrix, one that accounts for temporal dependence.

The sections that follow survey some applications of *GMM* estimators to economic time series. A feature of many of these examples is that the parameter β by itself may not admit a full depiction of the stochastic process that generates data. *GMM* estimators are constructed to be to achieve *partial* identification of the stochastic evolution and to be robust to the remaining unmodeled components.

3.1   Time Series Central Limit Theory

Time series estimation problems must make appropriate adjustments for the serial correlation for the stochastic process $\{f(x_t, \beta_o)\}$. A key input into the large sample properties of *GMM* estimators is a central limit approximation:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(x_t, \beta_o) \Rightarrow \text{Normal}(0, \Sigma_o)$$

for an appropriately chosen covariance matrix Σ_o. An early example of such a result was supplied by Gordin (1969) who used martingale approximations for partial sums of stationary, ergodic processes.[5] The matrix Σ_o must include adjustments for temporal dependence:

$$\Sigma_o = \sum_{j=-\infty}^{+\infty} E f(x_t, \beta_o) f(x_{t-j}, \beta_o)',$$

which is the long-run notion of a covariance matrix that emerges from spectral analysis of time series. In many *GMM* applications, martingale arguments show that the formula for Σ_o simplifies to include only a small number of nonzero terms. It is the adjustment to the covariance matrix that makes the time series implementation differ from the *iid* implementation (Hansen (1982)).

Adapting the minimum chi-square apparatus to this environment requires that we estimate the covariance matrix Σ_o. Since f is not typically separable in x and β, an estimator of Σ_o requires an estimator of β_o. This problem is familiar to statisticians and econometricians through construction of feasible generalized least squares estimators. One approach is to form an initial estimator of β_o with an arbitrary nonsingular weighting matrix V and to use the initial estimator of β_o to construct an estimator of Σ_o. Hansen (1982), Newey and West (1987) and many others provide consistency results for the estimators of Σ_o. Another approach is to iterate back and forth between parameter estimation and weighting matrix estimation until a fixed point is reached, if it exists. A third approach is to construct an estimator of $\Sigma(\beta)$ and to replace V in the chi-square criterion by an estimator of $\Sigma(\beta)^{-1}$ constructed for each β. Given the partial specification of the model, it is not possible to

---

[5]See Hall and Heyde (1980) for an extensive discussion of this and related results.

construct $\Sigma(\beta)$ without use of the time series data. Long run covariance estimates, however, can be formed for the process $\{f(x_t, \beta)\}$ for each choice of β. Hansen, Heaton, and Yaron (1996) refer to this method as *GMM* estimation with a continuously-updated weighting matrix. Hansen, Heaton, and Yaron (1996), Newey and Smith (2000) and Stock and Wright (2000) describe advantages to using continuous-updating, and Sargan (1958) shows that in some special circumstances this method reproduces a quasi-maximum likelihood estimator.

3.2 Efficiency

Since the parameter vector β that enters moment condition (1) may not fully characterize the data evolution, direct efficiency comparisons of *GMM* estimators to parametric maximum likelihood are either not possible or not interesting. Efficiency statements can be made for narrower classes of estimators, however.

For the study of *GMM* efficiency, instead of beginning with a distance formulation, index a family of *GMM* estimators by the moment conditions used in estimation. Specifically, study

$$ag_T(b_T) = 0$$

where a is a k by n selection matrix. The selection matrix isolates which (linear combination of) moment conditions will be used in estimation and indexes alternative *GMM* estimators. Estimators with the same selection matrix have the same asymptotic efficiency.[6] In practice, the selection matrix can depend on data or even the parameter estimator provided that the selection matrix has a probability limit. As with weighting matrices for minimum chi-square criteria, we suppress the possible dependence of the selection matrix a on data or parameters for notational simplicity. The resulting *GMM* estimators are asymptotically equivalent (to possibly infeasible) estimators in which the selection matrix is replaced by its probability limit. As has been emphasized by Sargan (1958) and Sargan (1959) in his studies of instrumental variables estimators, estimation accuracy can be studied conveniently as a choice a^* of an efficient selection matrix. The link between a weighting matrix V and selection matrix a is seen in the first-order conditions:

$$a = V \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f(x_t, b_T)}{\partial \beta}$$

or their population counterpart:

$$a = Vd$$

where

$$d = E \left[\frac{\partial f(x_t, \beta_o)}{\partial \beta} \right].$$

---

[6]Without further normalizations, multiple indices imply the same estimator. Premultiplication of the selection matrix a by a nonsingular matrix e results in the same system of nonlinear equations.

Other distance measures including analogs to the ones studied by Berkson (1944) and Taylor (1953) can also be depicted as a selection matrix applied to the sample moment conditions $g_T(\beta)$. Moreover, *GMM* estimators that do not solve a single minimization problem may still be depicted conveniently in terms of selection matrices.[7]

Among the class of estimators indexed by a, the ones with the smallest asymptotic covariance matrix satisfy:

$$a = ed'\Sigma_o^{-1}$$

where e is any nonsingular matrix and the best covariance matrix is $d'(\Sigma_o)^{-1}d$. This may be shown by imitating the proof of the famed Gauss-Markov Theorem, which establishes that the ordinary least squares estimator is the best, linear unbiased estimator.

3.3   Semiparametric Efficiency

Many *GMM* applications, including ones that we describe subsequently, imply an extensive (infinite) collection of moment conditions. To apply the previous analysis requires an *ad hoc* choice of focal relations to use in estimation. Hansen (1985) extends this indexation approach to time-series problems with an infinite number of moment conditions. The efficiency bound for an infinite family of *GMM* estimators can be related to the efficiency of other estimators using a semiparametric notion of efficiency. A semiparametric notion is appropriate because an infinite-dimensional nuisance parameter is needed to specify fully the underlying model. It has been employed by Chamberlain (1987) and Hansen (1993) in their study of *GMM* estimators constructed from conditional moment restrictions.

4   Linear Models

Researchers in econometrics and statistics have long struggled with the idea of how to identify an unknown coefficient vector β_o in a linear model of the form:

$$\beta_o \cdot y_t = u_t$$

where y_t is a k-dimensional vector of variables observed by an econometrician. Least squares solves this problem by calling one of the variables, y_{1t}, the dependent variable and requiring the remaining variables, y_{2t}, to be orthogonal to the disturbance term:

$$E(u_t y_{2t}) = 0.$$

Alternatively, as suggested by Karl Pearson and others, when there is no natural choice of a left-hand side variable, we may identify β_o as the first principal component, the linear combination of y_t with maximal variance subject to the constraint $|\beta| = 1$.

---

[7]For example, see Heckman (1976) and Hansen (1982) for a discussion of recursive estimation problems, which require solving two minimization problems in sequence.

A third identification scheme exploits the time series structure and has an explicit economic motivation. It is a time-series analog to the instrumental variables and two-stage least squares estimators familiar to economists. Suppose that a linear combination of y_t cannot be predicted given data sufficiently far back into the past. That is,

$$E\left(\beta_o \cdot y_t | \mathcal{F}_{t-m}\right) = 0 \tag{3}$$

where $\mathcal{F}_t$ is a conditioning information set that contains at least current and past values of y_t. This conditional moment restriction can be used to identify the parameter vector β_o, up to scale. In this setup there may be no single variable to be designated as endogenous with the remainder being exogenous or even predetermined. Neither least squares nor principal components are appropriate for identifying β_o.

The model or the precise context of the application dictates the choice of lag m. For instance, restriction (3) for a specified value of m follows from martingale pricing relations for multiperiod securities, from Euler equations from the investment problems faced by decision-makers, or from the preference horizons of policy-makers.

To apply *GMM* to this problem, use (3) to deduce the matrix equation:

$$E\left(z_{t-m}y_t'\right)\beta_o = 0$$

where z_t is an n-dimensional vector of variables in the conditioning information set $\mathcal{F}_t$ and $n \geq k - 1$. By taking unconditional expectations we see that β_o is in the null space of the n by k matrix $E(z_{t-m}y_t')$. The model is over-identified if $n \geq k$. In this case the matrix $E(z_{t-m}y_t')$ must be of reduced rank and in this sense is special. This moment condition may be depicted as in (1) with:

$$f(x_t, \beta) = z_{t-m}y_t'\beta_o$$

where x_t is a vector containing the entries of z_{t-m} and y_t. The *GMM* test of over-identification, based on say a minimum chi-square objective, aims to detect this reduced rank. The *GMM* estimator of β_o seeks to exploit this reduced rank by approximating the direction of the null space.[8] Given that the null space of $Ez_{t-m}y_t'$ is not degenerate, it might have more dimensions than one. While the moment conditions are satisfied (the null space of $E(z_{t-m}y_t')$ is nondegenerate), the parameter vector itself is under-identified.[9]

As posed here the vector β_o is not identified but is at best identified up to scale. This problem is analogous to that of principal component analysis. At best we might hope to

---

[8]For the *GMM* applications in this and the next section, there is a related but independent statistics literature. Godambe and Heyde (1987) and others study efficiency criteria of martingale estimation equations for a fully-specified probability model. By contrast, our models are partially specified and have estimation equations, *e. g.* $\sum_{t=1}^{T} [z_{t-m}y_t']\beta_o$, that is a martingale only when $m = 1$. When the estimation equation is not a martingale, Hansen (1982) and Hansen (1985) construct an alternative martingale approximation to analyze statistical efficiency.

[9]A recent literature on *weak* instruments is aimed at blurring the notion of being underidentified. See Stock and Wright (2000) for an analysis of weak instruments in the context of *GMM* estimation.

identify a one-dimensional subspace. Perhaps a sensible estimation method should therefore locate a subspace rather than a parameter vector. Normalization should be inessential to the identification beyond mechanically selecting the parameter vector within this subspace. See Hansen, Heaton, and Yaron (1996) for a discussion of normalization invariant *GMM* estimators.[10]

There are some immediate extensions of the previous analysis. The underlying economic model may impose further, possibly nonlinear, restrictions on the linear subspace. Alternatively, the economic model might imply multiple conditional moment relations. Both situations are straightforward to handle. In the case of multiple equations, prior restrictions are needed to distinguish one equation from another, but this identification problem is well studied in the econometrics literature.

4.1   Applications

A variety of economic applications produce conditional moment restrictions of the form (3). For instance, Hansen and Hodrick (1980) studied the relation between forward exchange rates and future spot rates where m is the contract horizon. Conditional moment restrictions also appear extensively in the study of a linear-quadratic model of production smoothing and inventories (see West (1995) for a survey of this empirical literature). The first-order or Euler conditions for an optimizing firm may be written as (3) with $m = 2$. Hall (1988) and Hansen and Singleton (1996) also implement linear conditional-moment restrictions in the lognormal models of consumption and asset returns. These moment conditions are again derived from first-order conditions, but in this case for a utility-maximizing consumer/investor with access to security markets. Serial correlation due to time aggregation may cause m to be two instead of one. Finally, the linear conditional moment restrictions occur in the literature on monetary policy response functions when the monetary authority is forward-looking. Clarida, Gali, and Gertler (2000) estimate such models in which m is dictated by preference horizon of the Federal Reserve bank in targeting nominal interest rates.

4.2   Efficiency

The choice of z_{t-m} in applications is typically *ad hoc*. The one-dimensional conditional moment restriction (3) actually gives rise to an infinite number of unconditional moment restrictions through the choice of the vector z_{t-m} in the conditioning information set $\mathcal{F}_{t-m}$. Any notion of efficiency based on a preliminary choice z_{t-m} runs the risk of missing some potentially important information about the unknown parameter vector β_o. It is arguably more interesting to examine asymptotic efficiency across an infinite-dimensional family of *GMM* estimators indexed by feasible choices of z_{t-m}. This is the approach adopted by Hansen (1985), Hansen, Heaton, and Ogaki (1988) and West (2000). The efficient *GMM*

---

[10]This argument is harder to defend than it might seem at first blush. Prior information about the normalized parameter vector of interest may eliminate interest in an estimation method that is invariant to normalization.

estimator that emerges from this analysis is infeasible because it depends on details of the time series evolution. West and Wilcox (1996) and Hansen and Singleton (1996) construct feasible counterpart estimators based on approximating the time series evolution needed to construct the efficient z_{t-m}. In particular, West and Wilcox (1996) show important improvements in the efficiency and finite sample performance of the resulting estimators.

Efficient GMM estimators based on an infinite number of moment conditions often place an extra burden on the model-builder by requiring that a full dynamic model be specified. The costs of misspecification, however, is not so severe. While a mistaken approximation of the dynamics may cause an efficiency loss, by design it will not undermine the statistical consistency of the GMM estimator.

5   Models of Financial Markets

Models of well-functioning financial markets often take the form:

$$E\left(d_t z_t | \mathcal{F}_{t-m}\right) = q_{t-m} \tag{4}$$

where z_t is a vector of asset payoffs at date t, q_{t-m} is a vector of the market prices of those assets at date t and $\mathcal{F}_t$ is an information set available to economic investors at date t. By assumption, the price vector q_t is included in the information set $\mathcal{F}_t$. The random variable d_t is referred to as a *stochastic discount factor* between dates $t-m$ and date t. This discount factor varies with states of the world that are realized at date t and encodes risk adjustments in the security market prices. These risk adjustments are present because some states of the world are discounted more than others, which is then reflected in the prices. The existence of such a depiction of asset prices is well known since the work of Harrison and Kreps (1979), and its conditional moment form used here and elsewhere in the empirical asset pricing literature is justified in Hansen and Richard (1987). In asset pricing formula (4), m is the length of the financial contract, the number of time periods between purchase date and payoff date.

An economic model of financial markets is conveniently posed as a specification of the stochastic discount factor d_t. It is frequently modeled parametrically:

$$d_t = g(y_t, \beta_o) \tag{5}$$

where the function g is given *a priori* but the parameter β_o is unknown, and a target of estimation. Models of investor preferences may be written in this manner as can observable factor models in which d_t is a function (often linear) of vector y_t of *factors* observed by an econometrician. See Cochrane (2001) for examples and a discussion of how this approach is connected to other empirical methods in financial economics.

To apply GMM estimation, inference and testing to this problem we do two things. First we replace (4) by its unconditional counterpart:

$$E\left(d_t z_t - q_{t-m}\right) = 0,$$

which can be justified by applying the Law of Iterated Expectations. Second we substitute (5) into this unconditional moment implication:

$$E\left[g(y_t, \beta_o)z_t - q_{t-m}\right] = 0.$$

This may be depicted as (1) by writing:

$$f(x_t, \beta) = g(y_t, \beta)z_t - q_{t-m}$$

where x_t contains the entries of the *factors* y_t, the asset payoffs z_t and the corresponding prices q_{t-m}. As is evident from Hansen and Singleton (1982) and Kocherlakota (1996), *GMM*-based statistical tests of this model give rise to one characterization of what is commonly termed as the "equity-premium puzzle." This puzzle is a formal statement of how the observed risk-return tradeoff from financial market data is anomalous when viewed through the guises of many standard dynamic equilibrium models from the macroeconomics and finance literatures.

Studying unconditional rather than conditional moment relations may entail a loss of valuable information for estimation and testing. As emphasized by Hansen and Richard (1987), however, information in the set $\mathcal{F}_{t-m}$ may be used by an econometrician to form synthetic payoffs and prices. This information is available at the initiation date of the financial contract. While use of conditioning information to construct synthetic portfolios reduces the distinction between conditional and unconditional moment restrictions, it introduces additional challenges for estimation and inference that are being confronted in ongoing econometric research.

6　From Densities to Diffusions

A century ago Karl Pearson proposed a family of density functions and a moments-based approach to estimating a parameterized family of densities. The densities within this family have logarithmic derivatives that are the ratio of a first to a second-order polynomials. More recently, Wong (1964) provided scalar diffusion models with stationary distributions in the Pearson family. Diffusion models are commonly used in economic dynamics and finance.

A scalar diffusion is a solution to a stochastic differential equation:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dB_t$$

where μ is the local mean or *drift* for the diffusion, σ^2 is the local variance or *diffusion* coefficient and B_t is standard Brownian motion. I now revisit Pearson's estimation problem and method, but in the context of a data generated by a diffusion.

The stationary density q of a diffusion satisfies the integral equation:

$$\int_{\underline{r}}^{\bar{r}} \left(\mu\phi + \frac{\sigma^2}{2}\frac{d\phi}{dx}\right) qdx = 0 \tag{6}$$

for a rich class of ϕ's, referred to as test functions. This gives an extensive family of moment conditions for any candidate (μ, σ^2). In particular, moment conditions of the form (1) may be built by parameterizing μ and σ^2 and by using a vector of test functions ϕ. Pearson's moment recursions are of this form with test functions that are low-order polynomial.

Pearson's method of estimation was criticized by R. A. Fisher because it failed to be efficient for many members of the Pearson family of densities. Pearson and Fisher both presumed that the data generation is *iid*. To attain asymptotic efficiency with *iid* data, linear combinations of $f(x, \beta_o)$ should reproduce the score vector for an implied likelihood function. Low-order polynomials fail to accomplish this for many members of the Pearson family of densities, hence the loss in statistical efficiency.

When the data are generated by a diffusion, the analysis of efficiency is altered in a fundamental way. Parameterize the time series model via $(\mu_\beta, \sigma^2_\beta)$. There is no need restrict μ and σ^2 to be low-order polynomials. Let Φ be a vector of test functions with at least as many component functions as parameters to estimate. Associated with each Φ is a *GMM* estimator by forming:

$$f(x, \beta) = \mu_\beta(x)\Phi(x) + \frac{1}{2}\sigma^2_\beta(x)\frac{d\Phi}{dx}(x)$$

Then moment conditions (1) follow from (6).

Conley, Hansen, Luttmer, and Scheinkman (1997) calculate the efficiency bounds for this estimation problem using the method described in Hansen (1985). They perform this calculation under the simplifying fiction that a continuous-data record is available and show that the that an efficient choice of Φ is:

$$\Phi_\beta(x) = \frac{\partial}{\partial \beta}\left[\frac{2\mu_\beta(x) + d\sigma^2_\beta(x)/dx}{\sigma^2_\beta(x)}\right]. \tag{7}$$

This choice of test function turns out to be the *derivative* of the score vector with respect to the Markov state x.[11] The score function derivative is easier to compute than the score function itself because the constant of integration for implied density does not have to be evaluated for each β. While tractable, this efficient test function solution to the time series problem will typically not lead one to use low-order polynomials as test functions even for Pearson's parameterizations.

Pearson used moment recursions to construct tractable density estimates without numerical integration. Computational concerns have subsided, and the diffusion model restricts more than just the stationary density. Transition densities can also be inferred numerically for a given (μ, σ^2) pair. To motivate fitting only the stationary density, a model-builder must suppose potential model misspecification in the transition dynamics. One example of this form of misspecification is a subordinated diffusion model used to model financial time series

---

[11]This efficient choice of Φ depends on the true parameter vector β_o and hence is infeasible to implement. Conley, Hansen, Luttmer, and Scheinkman (1997) show that the same efficiency can be attained by a feasible estimator in which Φ is allowed to depend on β as in (7).

in which calendar time and a more relevant information-based notion of economic time are distinct.

Integral equation (6) is *localized* by using smooth test functions that concentrate their mass in the vicinity of given points in the state space. Banon (1978) used this insight to produce a nonparametric drift estimator using a locally constant parameterization of μ. Conley, Hansen, Luttmer, and Scheinkman (1997) justify a local linear version of the *GMM* test function estimator described in the previous subsection.

7   Related Approaches

In the last decade statisticians have explored empirical likelihood and other related methods. These methods are aimed at fitting empirical data distributions subject to *a priori* restrictions, including moment restrictions that depend on an unknown parameter. In particular, Qin and Lawless (1994) have shown how to use empirical likelihood methods to estimate parameters from moment restrictions like those given in (1) for *iid* data.[12] Baggerly (1998) describes a generalization of the method of empirical likelihood based on the Cressie-Read divergence criterion. Imbens, Spady, and Johnson (1998) and Bonnal and Renault (2000) use this generalization to unify the results of Qin and Lawless (1994), Imbens (1997), Kitamura and Stutzer (1997) and others and *GMM* estimation with a continuously-updated weighting matrix. (See Newey and Smith (2000) for a related discussion.) Within the *iid* framework, Bonnal and Renault (2000) show that the continuously-updated *GMM* estimator is a counterpart to empirical likelihood except that it uses Pearson's χ^2 criterion. The relative entropy or information-based estimation of Imbens (1997) and Kitamura and Stutzer (1997) is of the same type but based on an information criterion of fit for the empirical distribution.

Kitamura and Stutzer (1997) and others use clever blocking approaches for weakly dependent data to adapt these methods to time series estimation problems. Many *GMM* applications imply *conditional* moment restrictions where the conditioning information is lagged m time periods. Extensions of these empirical distribution methods to accommodate time series *conditional* moment implications is an important area of research.

8   Moment-Matching Reconsidered

Statistical methods for partially-specified models allow researchers to focus an empirical investigation and to understand sources of empirical anomalies. Nevertheless, the construction of fully-specified models is required to address many questions of interest to economists such as the effect of hypothetical interventions or policy changes.

Models of economic dynamical systems remain highly stylized, however; and they are not rich enough empirically to confront a full array of empirical inputs. Producing interesting comparative dynamic results for even highly stylized dynamic systems is often difficult, if not

---

[12]While Qin and Lawless (1994) use the statistics literature on estimation equations for motivation, they were apparently unaware of closely-related econometrics literature on *GMM* estimation.

impossible, without limiting substantially the range of parameter values that are considered. Since analytical solutions are typically not feasible, computational methods are required. As in other disciplines, this has led researchers to seek ways to calibrate models based on at least some empirical inputs.[13] The analysis of dynamical economic systems brings to the forefront both computational and conceptual problems.

The computational problems are reminiscent to those confronted by the inventors of minimum chi-square methods.[14] Two clever and valuable estimation methods closely related to moment matching have recently emerged from the econometrics literature. One method is called *indirect inference* and has been advanced by Smith (1993) and Gourieroux, Monfort, and Renault (1993). The idea is to fit a conveniently chosen, but misspecified approximating model that is easy to estimate. The empirical estimates from the approximating model are used as targets for the estimation of the dynamic economic model. These targets are *matched* using a minimum chi-square criteria.

This method is sometimes difficult to implement because the implied approximating models associated with the dynamic economic model of interest may be hard to compute. Gallant and Tauchen (1996) circumvent this difficulty by using the score function for the approximating model (evaluated at the empirical estimates) as targets of estimation. The computational burden is reduced to evaluating an empirical score expectation as a function of the underlying parameter of interest. This allows researchers to use an expanded set of approximating statistical models.

In both of these methods there are two models in play, an approximating statistical model and an underlying economic model. These methods address some of the computational hurdles in constructing parameter estimators through their use of convenient approximating statistical models. On the other hand, it is harder to defend their use when underlying dynamic economic model is itself misspecified. Gallant and Tauchen (1996), for instance, use statistical efficiency as their guide in choosing approximating statistical models. Better approximating models result in more accurate estimators of the parameters of interest. This justification, however, neglects the role of misspecification in the underlying dynamic economic model. For *indirect inference* it is often not evident how to construct approximating models that leave the estimates immune to the stylized nature of the underlying economic model. Thus it remains an important challenge for econometricians to devise methods for infusing empirical credibility into "highly stylized" models of dynamical economic systems. Dismissing this problem through advocating only the analysis of more complicated "empirically realistic" models will likely leave econometrics and statistics on the periphery of important applied research. Numerical characterizations of highly stylized models will continue with or without the aid of statistics and econometrics.

---

[13]See Hansen and Heckman (1996) for an analysis of this literature.

[14]Considerable progress has been made recently in the development of computationally tractable Bayesian alternatives that also avoid maximizing a likelihood function.

References

Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika 85*, 535–547.

Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM Journal of Control and Optimization 16*, 380–395.

Berkson, J. (1944). Application of the logistic function to bioassay. *Journal of the American Statistical Association 39*, 357–365.

Bonnal, H. and E. Renault (2000). A comparison of alternative moment matching estimators. CREST-INSEE.

Burankin, E. and J. Gurland (1951). On asymptotically normal estimators: I. *University of California Publications in Statistics 1*, 86–130.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics 34*, 305–334.

Clarida, R., J. Gali, and M. Gertler (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics 115*, 147–180.

Cochrane, J. (2001). *Asset Pricing.* Princeton, New Jersey: Princeton University Press.

Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies 10*, 525–577.

Gallant, A. R. and G. Tauchen (1996). Which moments to match. *Econometric Theory 12*, 657–681.

Godambe, V. P. and C. C. Heyde (1987). Quasi-likelihood and optimal estimation. *International Statistical Review 55*, 231–244.

Gordin, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl. 10*, 1174–1176.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application.* Boston: Academic Press.

Hall, R. E. (1988). Intertemporal substitution and consumption. *Journal of Political Economy 96*, 339–357.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hansen, L. P. (1985). A method for calculating bounds on asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics 30*, 203–238.

Hansen, L. P. (1993). Semiparametric efficiency bounds for linear time-series models. In P. C. B. Phillips (Ed.), *Models, Methods and Applications of Econometrics: Essays in Honor of A. R. Bergstrom*, Chapter 17, pp. 253–271. Cambridge, MA: Blackwell.

Hansen, L. P., J. C. Heaton, and M. Ogaki (1988). Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Society 83*, 863–871.

Hansen, L. P., J. C. Heaton, and A. Yaron (1996). Finite sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics 14*, 262–280.

Hansen, L. P. and J. J. Heckman (1996). The empirical foundations of calibration. *Journal of Economic Perspectives 10*, 87 – 104.

Hansen, L. P. and R. J. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates. *Journal of Political Economy 88*, 829–853.

Hansen, L. P. and S. F. Richard (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica 55*(3), 587–613.

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variables of nonlinear rational expectations models. *Econometrica 50*, 1269–1286.

Hansen, L. P. and K. J. Singleton (1996). Efficient estimation of linear asset pricing models with moving-average errors. *Journal of Business and Economic Statistics 14*, 53–68.

Harrison, J. M. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20*, 381–408.

Heckman, J. J. (1976). The common structure of statistical methods of truncation, sample selection, and limited dependent variable and a simple estimator for such models. *Annals of Economic and Social Measurement 5*, 475–492.

Imbens, G. W. (1997). One-step estimators for over-identified generalized method of moments models. *Review of Economic Studies 64*, 359–383.

Imbens, G. W., R. H. Spady, and P. Johnson (1998). Information theoretic approaches to inference in moment condition estimation. *Econometrica 66*, 333–357.

Kitamura, Y. and M. Stutzer (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica 65*(4), 861–874.

Kocherlakota, N. (1996). The equity premium: It's still a puzzle. *Journal of Economic Literature 34*, 42–71.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Newey, W. K. and R. J. Smith (2000). Asymptotic bias and equivalence of gmm and gel. manuscript.

Neyman, J. (1949). Contributions to the theory of the χ^2 test. *Proceedings of the Berkeley Symposium on Mathematical Statistics*. University of California Press.

Ogaki, M. (1993). Generalized method of moments: Econometric applications. In G. S. Maddala, C. R. Rao, and H. D. Vinod (Eds.), *Handbook of Statistics*, Volume 11, Chapter 17, pp. 455–486. Elsevier Science Publishers.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *Annals of Statistics 22*, 300–325.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*, 393–415.

Sargan, J. D. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society 21*, 91–105.

Smith, A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, 63–84.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*, 1055–1096.

Taylor, W. (1953). Distance functions and regular best aymptotically normal estimates. *Annals of Mathematical Statistics 24*, 85–92.

West, K. D. (1995). Inventory models. In *Handbook in Applied Econometrics - Volume I (Macroeconometrics)*, Volume I. Oxford: Basil Blackwell.

West, K. D. (2000). On optimal instrumental variables estimation of stationary time series models. *International Economic Review*. forthcoming.

West, K. D. and D. W. Wilcox (1996). A comparison of alternative instumental variables estimators of a dynamic linear model. *Journal of Business and Economic Statistics 14*, 281 – 293.

Wong, E. (1964). The construction of a class of stationary markoff processes. In R. Bellman (Ed.), *Sixteenth Symposium in Applied Mathematics - Stochastic Processes in Mathematical Physics and Engineering*, pp. 264 – 276. Providence, RI: American Mathematical Society.