Title
Panel data and Euler equations

Author
Ligon, Ethan

Publication Date
1997-05-01

PANEL DATA AND EULER EQUATIONS

ETHAN LIGON

1. Introduction

Dynamic rational expectations models featuring agents with additively time-separable utility functions typically yield some sort of martingale restriction which may be used for either testing the model or estimating its parameters. While different models may yield a variety of these sorts of restrictions, the one most commonly observed is the Euler equation.

Martingale restrictions are exceptionally useful in a time series context, since the martingale property delivers a very useful sort of independence. However, as noted by Chamberlain (1984), this independence property does not extend to the analysis of panel data. Most panels have many agents observed over a small number of time periods. For these data, the most natural asymptotic theory would hold the number of time periods (T) fixed, while letting the number of agents (N) approach infinity. However, when working with the Euler equation or similar restrictions, this procedure will yield inconsistent estimators whenever there are important aggregate components to the innovations observed by each agent (Pakes 1994).

This paper develops a characterization of estimators that rely on $N \to \infty$, but which hold T fixed. In particular, we show that limiting estimator is a random variable, and show how to calculate its distribution when there are overidentifying restrictions. When the distribution of the limiting estimator is nondegenerate, the estimator cannot be consistent, but knowledge of this distribution permits us to engage in the usual sort of inference and hypothesis testing.

2. Model

Suppose that we have a dataset of observations on N agents over the course of T periods. An economic model implies that

$$y_{it} = f(x_{it}, b_0) + u_{it},$$

Date: May, 1997.

where $i = 1, 2, \ldots, N$ indexes agents, $t = 1, 2, \ldots, T$ indexes time, and b_0 is a vector of unknown parameters, assumed to lie in some compact parameter space $B \subseteq \mathbb{R}^k$. Observations on a single agent may be written as a vector $w_i = (y_{i1}, \ldots, y_{iT}, x'_{i1}, \ldots, x'_{iT}, z'_{i0}, \ldots, z'_{iT-1})$, where the z_{it}'s are q-vectors of instrumental variables, which generate the sequence of σ-algebras $\{\mathcal{F}_{it}\}_{t=0}^{T-1}$. We assume that the u_{it} are stationary.

We wish to test the hypothesis that

$$\mathrm{E}(u_{it}|\mathcal{F}_{it-1}) = 0. \tag{1}$$

Such conditional moment restrictions arise from a wide variety of dynamic economic models. In particular, the Euler equation found in work on the permanent income hypothesis yields precisely this sort of restriction, where the u_{it} may be regarded as agents' forecast errors.

Define $h_{it}(b) = (y_{it} - f(x_i t, b))z_{it-1}$; note that $h_{it}(b_0) = u_{it}z_{it-1}$. Then an implication of (1) is

$$\mathrm{E}h_{it}(b_0) = 0. \tag{2}$$

In order to estimate b_0, it seems natural to proceed by constructing a sample counterpart to (2). In thinking about asymptotic properties there are two possible limits we can think of taking, in N and in T, rather than the usual one. Accordingly, we construct our sample moments in two steps. First, we define a set of T functions $h_t(b)$ by

$$h_t(b) = \operatorname*{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} h_{it}(b).$$

We assume that $h_t(b)$ exists and is finite for all $b \in B$. Furthermore, we assume that the conditions of some central limit theorem are satisfied (cf. White (1984)), and that

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} h_{it}(b_0) - h_t(b_0)\right) \Rightarrow N(0, \Sigma_t(b_0)). \tag{3}$$

The second step in the construction of our sample moment involves averaging over time. Define

$$g_T(b) = \frac{1}{T}\sum_{t=1}^{T} h_t(b). \tag{4}$$

Since the one-sided conditional moment restrictions (1) imply that $\{h_{it}(b_0), \mathcal{F}_{it}\}$ is a martingale difference sequence, we can apply Chow's law of large numbers to show that as $T \to \infty$, $g_T(b_0)$ converges to zero with probability one. As a consequence generalized method of moments (GMM) estimators will be consistent as T grows large.

However, it is often the case that panel datasets have N considerably larger than T. In these circumstances, it would be nice to have an estimator which was consistent as $N \to \infty$, holding T fixed. Unfortunately, theory often suggests that we may observe considerable cross-sectional dependence (Chamberlain 1984). For example, macroeconomic innovations or the use of consumption insurance (Mace 1991) would be expected to induce dependence across agents' forecast errors. Accordingly, for any particular t, $h_t(b_0)$ may be quite different from zero, even though $\mathrm{E}h_t(b_0) = 0$; as a consequence, estimators based on this moment condition will not be consistent in general.[1]

3. Example

In this section we develop an example inspired by Hansen and Singleton (1982). Suppose that we wish to estimate risk aversion using a panel of N agents observed over six periods. We assume that each agent i has preferences over own consumption given by

$$\mathrm{E}\left[\sum_{t=0}^{\infty} \beta^t U(c_{it}) \middle| \mathcal{F}_{i0}\right]$$

where β is the discount factor, and where agents' momentary utility from consumption, $U(c)$, is of the CES form

$$U(c) = \frac{c^{1-b_0} - 1}{1 - b_0}.$$

Here c is consumption and b_0 is the coefficient of relative risk aversion. To make our example simple, we assume that all agents hold some asset with return R_t at time t; we might think of this asset as being some sort of market portfolio.[2] Accordingly, we expect agents' consumptions to satisfy the Euler equation

$$\mathrm{E}\left[\beta R_{t+1}\left(\frac{c_{it}}{c_{it+1}}\right)^{b_0} \middle| \mathcal{F}_{it}\right] - 1 = 0$$

[1]Not all cross-sectional dependence produces inconsistent estimators, of course. A sufficient condition for consistency is that the data generating process for $\{h_{it}\}$ be stationary and ergodic. We might imagine sorts of 'local' dependence—perhaps due to weather shocks or regional price variation—which while inducing some cross-sectional dependence would still satisfy these conditions; see Conley (1996). On the other hand, the existence of markets available to all of the agents in an economy is apt to yield data generating processes which are *both* non-stationary (e.g., agent's marginal utilities of consumption may obey a martingale) and non-ergodic.

[2]We also assume that if R_t is the return on the the market portfolio, then the returns on the primitive assets do not span the set of possible states— we want our agents to face some idiosyncratic as well as aggregate uncertainty.

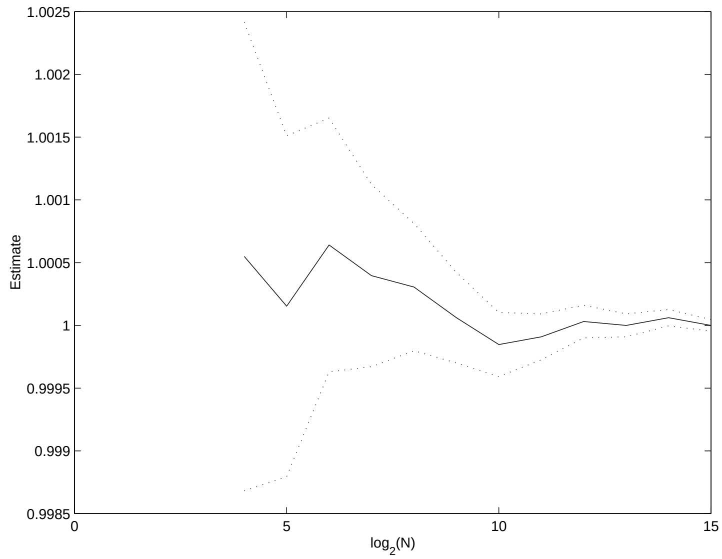

FIGURE 1. Sample path of GMM estimator with no aggregate shocks and $b_0 = 1$. The dotted lines are conventionally estimated 95 per cent confidence intervals.

[cf. Hansen and Singleton (1982)]. We choose as instruments q linearly independent, $\mathcal{F}_{it}$-measurable, covariance stationary variables z^j_{it-1}, $j = 1, \dots, q$. Then define

$$h_{it}(b) = \begin{bmatrix} \left[\beta R_t \left(\frac{c_{it-1}}{c_{it}}\right)^b - 1\right] z^1_{it-1} \\ \vdots \\ \left[\beta R_t \left(\frac{c_{it-1}}{c_{it}}\right)^b - 1\right] z^q_{it-1} \end{bmatrix}$$

so that $\mathrm{E}h_{it}(b_0) = 0$ is a pair of unconditional moment conditions. In the absence of any sort of aggregate shocks, $g_T(b_0) = \frac{1}{T}\sum_{t=1}^T h_t(b_0) = 0$, so the usual GMM estimator is consistent. Let $g_{NT}(b) = \frac{1}{NT}\sum_{t=1}^T \sum_{i=1}^N h_{it}(b)$ be our estimator of the moment conditions, and let b_{NT} denote our estimator of b_0.; Figure 1 displays a particular sample path for the estimator when $b_0 = 1$. Here the solid line is the sample path of the estimator, while the dotted lines describe a conventionally calculated 95 per cent confidence interval.[3]

[3]The confidence interval is constructed by assuming that the estimator b_{NT} is normally distributed with $\mathrm{E}b_{NT} = b_0$, and estimated variance $\frac{1}{NT}\left[\frac{\partial g_{NT}(b_{NT})}{\partial b'} V_{NT}(b_{NT})^{-1} \frac{\partial g_{NT}(b_{NT})}{\partial b}\right]^{-1}$.

However, let us suppose that aggregate shocks matter in this economy. In particular, let c_t denote per capita consumption at time t in the economy as a whole, and assume that $\beta R_t \left(\frac{c_{t-1}}{c_t}\right)^{b_0} - 1$ is normally distributed with mean zero and variance σ_0^2.[4] Furthermore, we assume that the sampling process is such that $\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^N c_{it} = c_t$; this is sufficient for $\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^N h_{it}(b) = h_t(b)$. It follows that $h_t(b_0)$ is normally distributed with mean zero and covariance matrix $V = \sigma_0^2 \mathrm{E} z_{it-1} z_{it-1}'$.

A particular draw from $N(0, V/T)$ using a (pseudo-) random number generator yields the following values for the aggregate shocks:

$$\begin{aligned}&\{h_1(b_0), h_2(b_0), \ldots, h_T(b_0)\}\\ &= \left\{ \begin{pmatrix}-0.0361\\-0.0361\end{pmatrix}, \begin{pmatrix}0.1453\\0.0819\end{pmatrix}, \begin{pmatrix}0.0944\\0.0113\end{pmatrix}, \begin{pmatrix}-0.0054\\-0.0001\end{pmatrix}, \begin{pmatrix}0.0134\\0.0001\end{pmatrix} \right\}.\end{aligned}$$

The global mean of these shocks is $g_T(b_0) = \left(\begin{smallmatrix}0.0423\\0.0114\end{smallmatrix}\right)$. While one might think that this isn't too different from zero,[5] there's considerable variation in the $h_t(b_0)$; a different draw might easily have yielded $g_T(b_0)$ quite different from zero. Indeed, even in the present case, the limiting estimator as $N \to \infty$ is decidedly inconsistent, as Figure 2 shows. In fact, the estimated confidence interval never includes $b_0 = 1$, and never overlaps with the confidence interval estimated in the absence of aggregate shocks.

4. Inference When N is Large

It is likely that estimators that rely on sequential moment restrictions such as (2) will not be consistent when T is fixed. In this setting $h_t(b)$ is a random variable; in order to obtain estimators with desirable properties we must place some additional structure on this random variable. Most earlier work on this problem has assumed that $h_t(b)$ is in fact a constant. Here we permit $h_t(b)$ to be a random variable with a non-degenerate distribution. While this is not enough to deliver a consistent estimator, it does add enough structure to the problem that we can conduct meaningful inference.

Assume for now that we observe $\{h_t(b)\}$, and that $h_t(b_0)$ is distributed $N(0, V(b_0))$. Let $V_T(b) = \frac{1}{T}\sum_{t=1}^T h_t(b)h_t(b)' - g_T(b)g_T(b)'$ be

[4] A complete description of the data generating process used in this example, along with code implementing the dgp, is available from the author.

[5] A χ^2 test of the hypothesis that $\mathrm{E}g_T(b_0) = 0$ yields a p-value of 0.18, so we cannot reject the null hypothesis at conventional levels of significance.

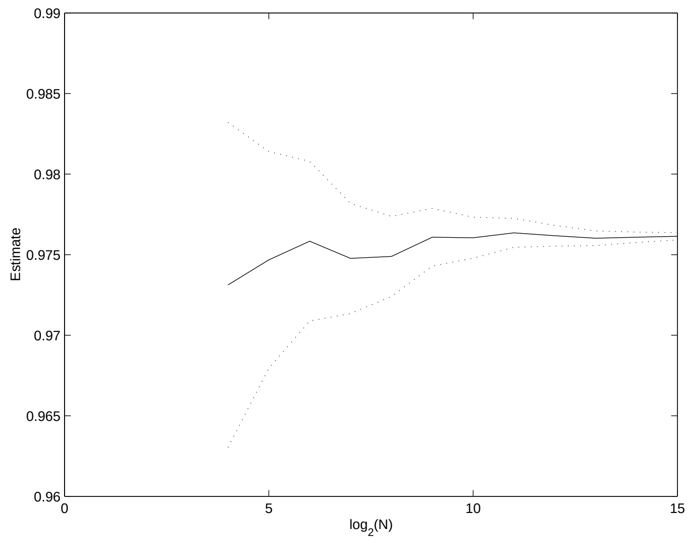

FIGURE 2. Sample path of GMM estimator with aggregate shocks and $b_0 = 1$. The dotted lines are conventionally estimated 95% confidence intervals.

an estimator of $V(b)$.[6] Accordingly, $V_T(b)/T$ is an estimator of the covariance matrix of $g_T(b)$. Since we know that $g_T(b_0)$ is distributed exactly $N(0, V(b_0)/T)$, we can write the maximum likelihood estimator of b_0 as

$$b_T^* = \operatorname*{argmin}_{b \in B} |V_T(b)|^{1/q} g_T(b)'[V_T(b)]^{-1} g_T(b). \tag{5}$$

Note that the maximum likelihood criterion function (minus one times the concentrated log likelihood, *modulo* a constant) differs from the natural GMM criterion only by the factor $|V_T(b)|^{1/q}$.

Call the GMM criterion

$$J_T(b) = g_T(b)'[V_T(b)]^{-1} g_T(b).$$

Since $h_t(b_0) \sim N(0, V(b_0))$, it follows that $TV_T(b_0)$ obeys the Wishart distribution $W(T, V(b_0))$ (Morrison 1962), and that $TJ_T(b_0)$ is drawn from Hotelling's generalized $\mathbb{T}^2$ distribution. Accordingly,

$$\left(\frac{T - q + k + 1}{q - k}\right) J_T(b_0) \sim F(q - k, T - q + k + 1) \tag{6}$$

[6]This estimator of $V(b)$ is appropriate when $h_t(b)$ is serially independent. An alternative estimator to account for possible serial dependence could be used instead (e.g., Newey and West (1987)); however, when T is quite small, such adjustments are likely to be unimportant.

(Rao (1973), pp. 541–2). As in the usual GMM setting, we know that $TJ_T(b_0)$ is distributed asymptotically χ^2_{q-k} as $T \to \infty$. However, here we know the exact distribution of $TJ_T(b_0)$ even when T is small; thus, when N is large, we can use this fact to construct exact tests of overidentifying restrictions.

Of course, the statistic given by (6) requires that we know b_0. We'd prefer to be able to work with a test statistic that depends on an estimate of b_0,

$$S_T(b) = \left(\frac{T-q+k+1}{q-k}\right) J_T(b). \tag{7}$$

This test statistic is appropriate if we wish to test the null hypothesis that $\mathrm{E}g_T(b) = 0$, or equivalently the joint hypothesis that $\mathrm{E}g_t(b_0) = 0$ and $b_0 = b$. Furthermore, since we know the pdf of $F(q-k, T-q+k+1)$, we can calculate the pdf of the estimator so long as g_T satisfies some regularity conditions (Billingsley (1986), Thm. 19.3).

Returning the example we began in Section 3, Figure 3 displays an estimate of the pdf of the GMM estimator b_T. Here we've used

$$\varphi(b) = \left|\frac{\partial S_T(b)}{\partial b}\right| f(\,S_T(b) - S_T(b_T)|\, q-k, T-q+k+1)$$

to estimate the pdf, where $f(\cdot|\gamma_1, \gamma_2)$ is the pdf for the F distribution, with (γ_1, γ_2) degrees of freedom. Note the correction to the argument in the F pdf. In this context b_T is the GMM point estimate, so $S_T(b_T) = \min_{b \in B} S_T(b_T)$. Clearly $\varphi(b_T) = 0$. Since b_T has measure zero in B in this example, this is harmless. As a practical matter, however, the limits of machine precision cause the pdf of b_T to be poorly estimated in some neighborhood of the GMM point estimate. One might correct for this error by omitting this neighborhood, and interpolating in order to the pdf instead. With this (possibly corrected) pdf in hand, we can go about inference in the usual way. For example, using the pdf in Figure 3, we can construct a confidence interval for b_0; in this example a 95 per cent confidence interval is $[0.9302, 1.0059]$, which includes b_0.

5. Inference When N is Small

When N is small, we don't observe the probability limits $\{h_t(b)\}$ which we exploited in the previous section. We can construct estimates of these by averaging over agents at each time period, but as a consequence we need to adjust our procedures for inference by accounting for sampling error.

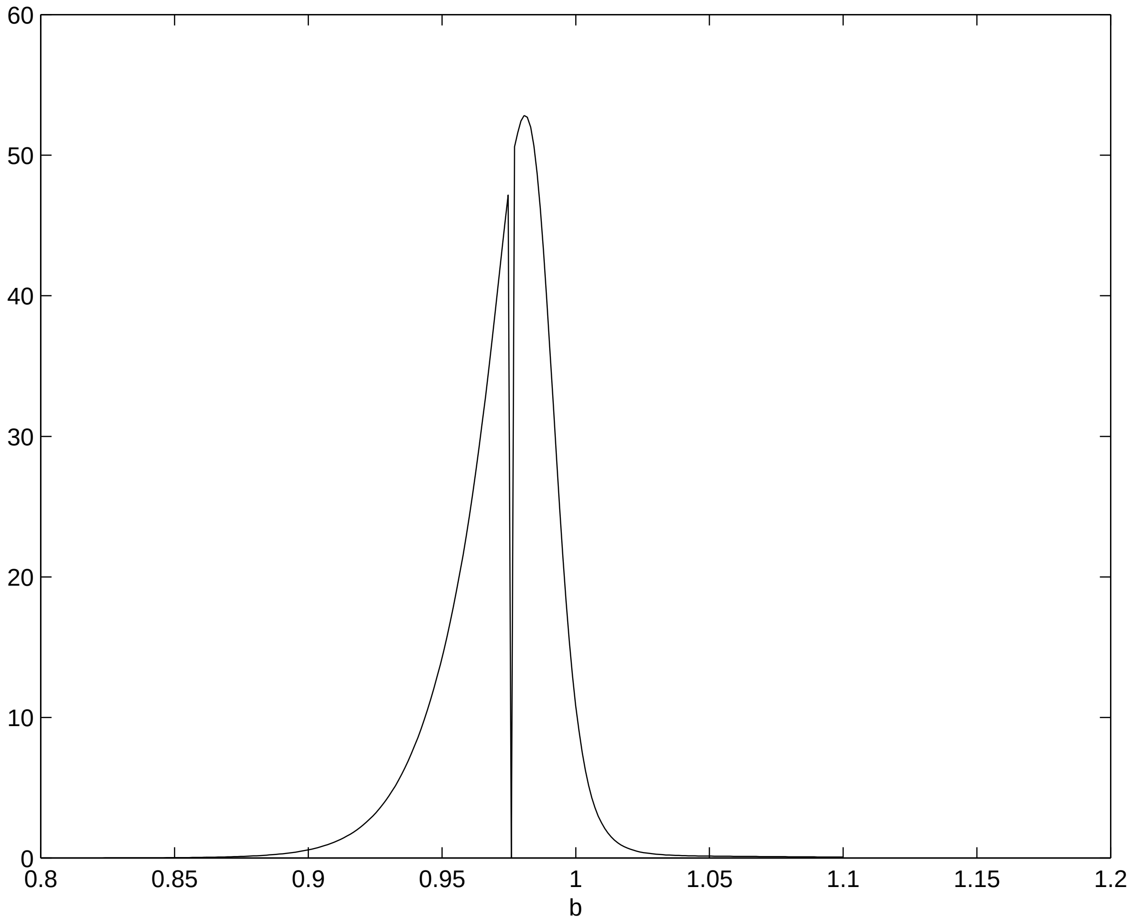

FIGURE 3. Density of the GMM estimator b_T.

Let us begin by defining

$$g_{NT}(b) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} h_{it}(b).$$

Using a similar notation, define $h_{Nt}(b) = \frac{1}{N}\sum_{i=1}^{N} h_{it}(b)$, and note that $\operatorname{plim}_{N\to\infty} h_{Nt}(b) = h_t(b)$. Then

$$g_{NT}(b) = \left[\frac{1}{T}\sum_{t=1}^{T} h_t(b)\right] + \left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(h_{it}(b) - h_t(b)\right)\right]. \tag{8}$$

We can write this as $g_{NT}(b) = g_T^1(b) + \bar{g}_{NT}(b)$, where the first term, $g_T(b)$ corresponds to the first bracketed term on the right hand side of (8), and $\bar{g}_{NT}(b)$ corresponds to the second.

From the assumption (3) it follows that

$$\sqrt{N}\bar{g}_{NT}(b_0) \Rightarrow N\left(0, \frac{1}{T^2}\sum_{t=1}^{T}\Sigma_t(b_0)\right),$$

and so we write $\bar{g}_{NT}(b_0) \overset{a}{\sim} N(0, \Sigma_T(b_0)/(NT^2))$, where $\Sigma_T(b_0) = \sum_{t=1}^{T}\Sigma_t(b_0)$, to indicate that $\bar{g}_{NT}(b_0)$ is approximately distributed $N\left(0, \frac{\Sigma_T(b_0)}{NT^2}\right)$. Note that this central limit result relies only on N being reasonably large, regardless of T.

Now recall that $g_T(b_0)$ is distributed $N(0, V(b_0)/T)$ by assumption. Accordingly, $g_{NT}(b_0) \overset{a}{\sim} N(0, T^{-1}\Omega_{NT})$, where $\Omega_{NT} = V(b_0) + \frac{1}{NT}\Sigma_T(b_0)$. Of course, we wouldn't ordinarily observe either $V(b_0)$ or $\Sigma_t(b_0)$. However, we can estimate Ω_{NT} by using

$$V_{NT}(b) = \frac{1}{T}\sum_{t=1}^{T} h_{Nt}(b)h_{Nt}(b)' - g_{NT}(b)g_{NT}(b)..$$

Note that $\mathrm{E}V_{NT}(b_0) = \Omega_{NT}$, and that $V_{NT}(b_0)$ is a consistent estimator of $V(b_0) = \operatorname{plim}_{N\to\infty}\Omega_{NT}$ as $N \to \infty$. With this covariance matrix estimator in hand, we might estimate b_0 via GMM, with a criterion function

$$J_{NT}(b) = g_{NT}(b)'[V_{NT}(b)]^{-1}g_{NT}(b),$$

or perhaps via quasi-maximum likelihood,[7] with a criterion function

$$|V_{NT}(b)|^{1/q}J_{NT}(b).$$

Reprising an argument from the previous section, since $h_{Nt}(b_0) \overset{a}{\sim} N(0, \Omega_{NT})$, it follows that $TV_{NT}(b_0)$ is approximately distributed $W(T, \Omega_{NT})$, and that

$$\left(\frac{T-q+k+1}{q-k}\right) J_{NT}(b_0) \overset{a}{\sim} F(q-k, T-q+k+1).$$

As before, we can't consistently estimate b_0; however, we can conduct inference using the statistic

$$S_{NT}(b) = \left(\frac{T-q+k+1}{q-k}\right) J_{NT}(b).$$

While we only know the approximate distribution of this statistic, we can conduct inference as if the distribution were exact; in fact, the inference will be asymptotically correct as $N \to \infty$.

[7]We term this estimator a *quasi*-maximum likelihood estimator because g_{NT} may be only approximately normally distributed.

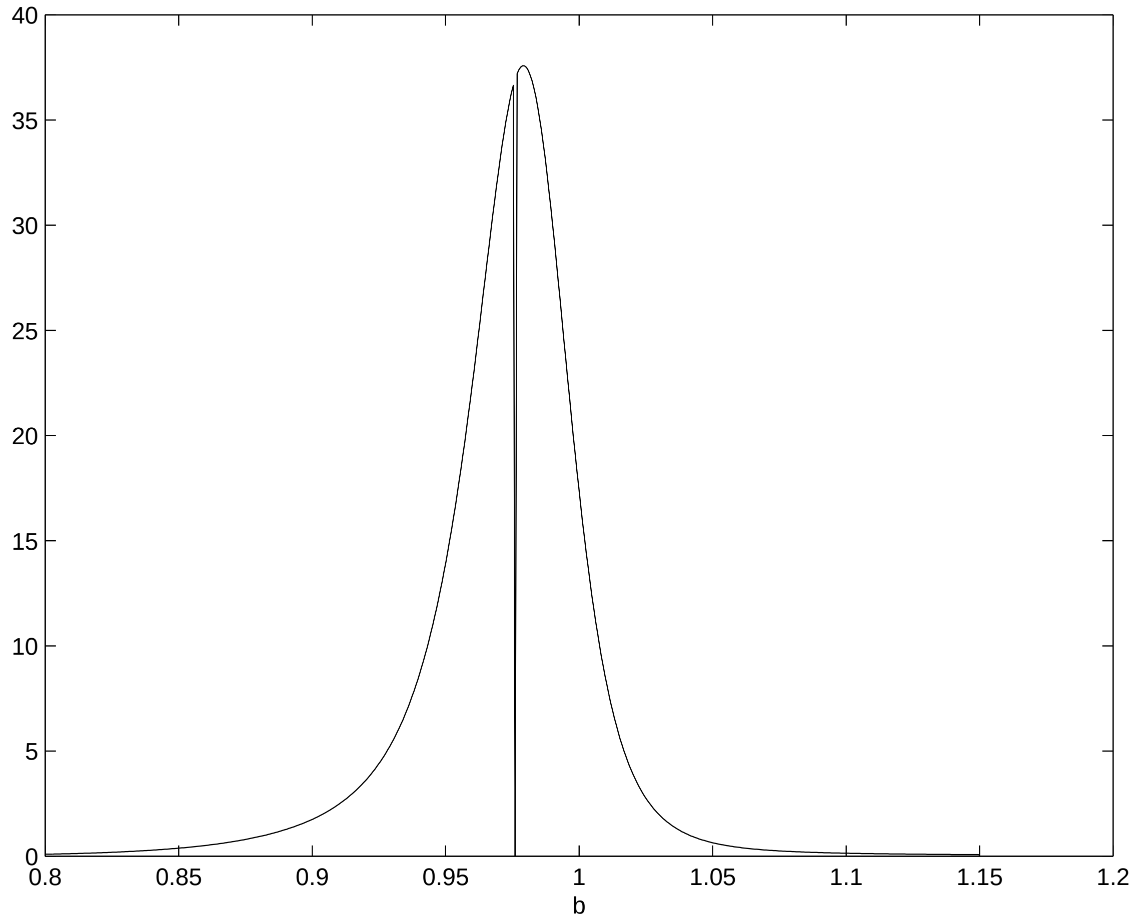

FIGURE 4. Density of the GMM estimator b_{NT}.

Figure 4 displays the pdf for the GMM estimator b_{NT} when $N = 1024$; a 95 per cent confidence interval for this estimator is $[0.9017, 1.0308]$.

6. CONCLUSION

When confronted with a panel dataset, researchers should ask themselves whether or not aggregate shocks are likely to influence the process which generated the data. If not, then moment conditions shouldn't depend on the date, which can be used as the basis for a simple test of the null hypothesis of no aggregate shocks. If this test rejects the null hypothesis, then estimators which rely on a growing cross-sectional dimension should be presumed to be inconsistent.

One possible response to this problem is to develop a complete characterization of the data-generating process, and proceed to use a maximum likelihood estimator. The need to completely specify the economic environment may often be a serious drawback here; if it were possible to completely specify the data-generating process, then it would probably have been wise to use maximum likelihood to begin with, since by doing soone could realize gains in efficiency relative to GMM. The appeal of method of moments estimators is that the researcher is permitted to remain agnostic with respect to many aspects of the economic environment.

This paper has developed a sort of middle ground. By imposing (or, better, deriving from the model) some additional structure on the aggregate shock process, one can often proceed to do the usual sorts of hypothesis testing and inference, despite the inconsistency of the estimator. In this paper we've focused on the case in which aggregate shocks are normally distributed; however, extensions to other distributions are straightforward.

References

Billingsley, P. (1986). *Probability and Measure* (Second ed.). John Wiley & Sons.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*. North-Holland.

Conley, T. G. (1996). Econometric modelling of cross sectional dependence. Manuscript, Northwestern University.

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

Mace, B. J. (1991). Full insurance in the presence of aggregate uncertainty. *Journal of Political Economy 99*, 928–956.

Morrison, D. F. (1962). On the distribution of sums of squares and cross products of normal variates in the presence of intra-class correlation. *Annals of Mathematical Statistics 33*, 1461–1463.

Newey, W. K. and K. D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review 28*(3), 777–787.

Pakes, A. (1994). Dynamic structural models: Problems and prospects. In C. A. Sims (Ed.), *Advances in Econometrics: Sixth World Congress*, Volume 1, Chapter 5. Cambridge: Cambridge University Press.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (Second ed.). John Wiley & Sons.

White, H. (1984). *Asymptotic Theory for Econometricians*. Orlando: Academic Press.