Temporal Aggregation Bias in Stock-Flow Models

van Ours, J. C.; Burdett, K.; Coles, M.

Publication date:
1994

Citation for published version (APA):
van Ours, J. C., Burdett, K., & Coles, M. (1994). *Temporal Aggregation Bias in Stock-Flow Models*. (CEPR discussion paper; No. 967). CEPR.

ISSN 0265-8003

TEMPORAL AGGREGATION BIAS IN STOCK-FLOW MODELS

Kenneth Burdett, Melvyn Coles and Jan Van Ours

Discussion Paper No. 967
May 1994

Centre for Economic Policy Research
25-28 Old Burlington Street
London W1X 1LB
Tel: (44 71) 734 9110

This Discussion Paper is issued under the auspices of the Centre's research programme in **Human Resources**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

CEPR Discussion Paper No. 967

May 1994

ABSTRACT

Temporal Aggregation Bias in Stock-Flow Models*

The matching function describes the *flow* of job creation as a function of the *stocks* of unemployed and vacancies. Most empirical work tries to identify such a relationship by regressing the flow of matches (aggregated over the month) on the stocks of unemployment and vacancies measured at the beginning of that month. It is shown that estimates obtained using this procedure will be downward biased if unemployment and vacancies are mean-reverting processes. If the bias is small, the size of the bias is proportional to the length of the period interval. By further aggregating the data, say from monthly to quarterly data, the downward bias should triple. The resulting change in the parameter estimates can then be used to estimate the size of the original bias.

JEL classification: C13, J64
Keywords: matching, temporal aggregation bias

Kenneth Burdett and Melvyn Coles
Department of Economics
University of Essex
Colchester
CO4 3SQ
UK
Tel: (44 206) 873 333

Jan van Ours
Department of Economics
Vrije Universiteit
De Boelelaan 1105
1081 HV Amsterdam
THE NETHERLANDS

*This paper is produced as part of a CEPR research programme on *The UK Labour Market: Microeconomic Imperfections and Institutional Features*, supported by a grant from the UK Department of Employment (no. 4RP-154-90). The views expressed in this paper are those of the authors and do not necessarily reflect those of the Department or of CEPR.

Submitted 31 January 1994

NON-TECHNICAL SUMMARY

There is a large search and matching literature which provides an equilibrium explanation for unemployment in an economy. Much of this work uses the notion of a matching function, which describes the *flow* of job creation as a function of the *stocks* of unemployed and vacancies. Such a process is closely related to the production literature, where a production function is used to describe the flow of output from the stocks of capital and labour (and any other inputs). Identifying such functional forms is an empirical question. This paper points out that recent empirical work may be biased. Moreover, it states conditions when it will be biased downwards. This is particularly important in the matching function literature where researchers have been trying to identify whether there are constant or increasing returns to matching.

The basic problem confronting a researcher who wants to estimate a matching function is that the matching function describes a continuous time process. Given the 'inputs' of unemployment and vacancies at a point in time, the matching function gives the flow of matches at that point in time. Identifying such a relationship needs continuous time data, which is usually not available. In most of the empirical literature the data used is monthly. There, the flow of output (matches) is proxied by aggregating the total flow of output over the month. This paper analyses the bias caused by this temporal aggregation.

Given this aggregation procedure, most empirical work has focused on the simultaneity problem. The simultaneity problem is that an unusually high matching rate will cause the levels of unemployment and vacancies to fall, *ceteris paribus*. This feedback causes the stock terms (the conditioning variables) to be negatively correlated with the matching rate (the dependent variable). This potentially biases the results downwards. To avoid this problem, most work has regressed the flow of matches (aggregated over the month) on the stocks of unemployment and vacancies measured at the beginning of that month.

Although this procedure may avoid the simultaneity problem, this paper shows there is a different bias, which potentially biases the results downwards. If the researcher uses the stock of unemployment and vacancies as measured at the beginning of the period as the conditioning variables, the dependent variable ought to be the flow of matches at the point in time. By using the aggregated flow over the subsequent period, the dependent variable is mis-measured. Whether this biases the estimates or not depends on the time-series properties of the conditioning variable. It is shown that if these variables are mean reverting, the estimates are biased downwards. If the conditioning variables follow a trendless Brownian motion, there is no bias.

The intuition is straightforward. Suppose the conditioning variables are mean reverting and at the beginning of the period, their values are above the mean. Then at the start of the period, the flow of matches will be quite high. Within that period, however, the stock of vacancies and unemployment falls (mean reverts – especially given the high matching rate). Thus the matching rate will fall over this period. Since the data is time aggregated, the measure of flow of matches will be too low. Thus, when the conditioning variables are high, the measured flow will be too low. Conversely, if the conditioning variables are too low, the measured flow will be too high. This systematically biases the estimates downwards. It is shown that if there is no mean-reverting trend, there is no bias.

The paper then shows how to correct for this bias. If the bias is small, the size of the bias is proportional to the length of the period interval. Thus by aggregating the data from say monthly to quarterly data, the downward bias should triple. The change in the parameter estimates using this procedure can then be used to estimate the size of the original bias.

Introduction

Search theory provides a theory relating flow behaviour to stock variables. The classic example is that of the matching function which predicts how quickly workers find work, given the stock of unemployment and vacancies. Since several theoretical predictions rest on whether there are increasing returns or not, the applied literature has turned to estimating the returns to scale to this function. This paper considers the bias caused by estimating a flow equation on discrete data.

The basic problem occurs when we are interested in estimating a flow equation of the form

$$x = g(K;\alpha)$$

which relates x, a flow variable, to K (a stock variable?) and some parameters of interest α. Unfortunately, the available data is usually in a discrete time framework, such as monthly or quarterly. Hence data on x is not available, but $\tilde{x}_t = \int_t^{t+1} x(v)\, dv$ is.

In the past, many researchers have worried about the presence of simultaneity bias. If the dynamics of K are described by $\dot{K} = h(x,K,t)$, then x feeds back into the stock variable K. Thus the path of K(.) over [t,t+1] is not independent of $\tilde{x}_t$. To avoid the simultaneity problem, applied researchers might estimate α by using a regression equation of the form

$$\tilde{x}_t = g(K_t;\alpha) + u_t$$

But in this specification, the dependent variable $\tilde{x}_t$ mismeasures x_t. This paper examines the direction and magnitude of the bias caused by using the temporally aggregated variable $\tilde{x}_t$. This is particularly important in the matching function literature. There, the recent focus has been to determine whether there are increasing returns or not. This paper shows that there are theoretical grounds to expect previous empirical results are downward biassed. However, it seems plausible that the magnitude of the bias may be quite small.

The Framework

Consider the following matching (or production) process. Suppose there are U unemployed workers and V vacancies which are trying to match over time. In search theory, it is standard to assume there exists a matching function M(U,V) which describes the rate at which these two sets of agents match (in continuous time). In particular, most applied research has estimated a Cobb-Douglas matching function of the form

$$M(U,V) = k\, U^{\alpha_1}\, V^{\alpha_2} \qquad (1)$$

where α_1, α_2 are both positive constants.
For ease of exposition, suppose U = V throughout. Hence

$$M(.) = k\, U^{\alpha} \qquad (2)$$

where $\alpha = \alpha_1 + \alpha_2$, and testing for increasing returns involves testing whether $\alpha > 1$ or not. Typically, the researcher takes logs and estimates the log-linear form

$$\ln M(.) = \alpha_0 + \alpha \ln U \qquad (3)$$

Unfortunately, data on the flow rate M(.) is not usually available. Published data commonly reports total matches over a given fixed time period, such as monthly or quarterly periods. The aim of this section is to show how using time-aggregated data may bias the estimated value of α downwards.

For simplicity, suppose (2) is true and U(t) deviates about some long run constant trend $\bar{U}$, where

$$U(t) = \bar{U} + \varepsilon(t) \qquad (4)$$

Since we shall be using first order Taylor expansions, it is assumed that the expected value of $\varepsilon(t)/\bar{U}$ is small for all t. Clearly, to estimate (3), it is necessary that the variance of $\varepsilon(t)$ is strictly positive.

Since the economist only observes discrete data, let s denote the length of the period interval (e.g. monthly). Then the total measured matches over a given period (v,v+s) is

$$M_s(v) = \int_v^{v+s} k\, U(t)^{\alpha}\, dt \qquad (5)$$

Taylor expanding U(t) about $\bar{U}$, (5) implies

$$M_s(v) = ks[\bar{U}]^{\alpha} + \alpha k[\bar{U}]^{\alpha-1} \int_v^{v+s} \varepsilon(t)\, dt + O(s\varepsilon^2[\bar{U}]^{\alpha-2}) \qquad (6)$$

Taking logs and Taylor expanding again gives

$$\ln M_s(v) = \ln ks[\bar{U}]^{\alpha} + \alpha/[s\bar{U}] \int_v^{v+s} \varepsilon(t)\, dt + O(\varepsilon/\bar{U})^2 \qquad (7)$$

Taking logs of U(v) and Taylor expanding gives :

$$\ln U(v) = \ln \bar{U} + \varepsilon(v)/\bar{U} + O(\varepsilon/\bar{U})^2 \qquad (8)$$

The economist observes the stock of unemployed U(t) at intervals v, v+s, v+2s,..... . Examining the bias on α when estimating (3) using OLS, is reduced to finding how (7) and (8) covary over the observed data. In fact, these two equations reveal that the underlying problem is analogous to (though not the same as) the measurement error literature. (7) defines the true relationship between the dependent and the independent variables. There, $\ln M_s(v)$ varies with $(1/s) \int_v^{v+s} [\varepsilon(t)/\bar{U}]\, dt$ (because it is a time aggregated variable). However, the fitted regression equation incorporates $\ln U(v)$ as the right hand side variable, which varies with $\varepsilon(v)/\bar{U}$. The question is how does this mismeasurement affect the infered value of α.

In general, $\varepsilon(t)$ will be a complicated process as unemployment is an endogeneous process. If $\hat{\alpha}$ is estimated by using OLS, then its expected value is given by

$$E(\hat{\alpha}) = \text{cov}\ [\log M_s(t), \log U(t)]\ /\ \text{var}\ [\log U(t)]$$

assuming the covariance matrix is bounded. (7) and (8) imply

$$E(\hat{\alpha}) \quad = \alpha \frac{\text{cov}(\ \varepsilon(t),\ (1/s) \int_t^{t+s} \varepsilon(v)\, dv\ )}{\text{var}(\varepsilon(t))} \qquad + O(\varepsilon/\bar{U}) \qquad (9)$$

<u>Proposition 1</u>

Define $u_t = \varepsilon(t) - (1/s) \int_t^{t+s} \varepsilon(v)\, dv$

If u_t is independent of $\varepsilon(t)$, then OLS is unbiased.

<u>Proof</u> Trivial.

Proposition 1 shows that any temporal aggregation bias depends on the time series properties of $\varepsilon(t)$. A classic example of a process satisfying the conditions of proposition 1, is that $\varepsilon(t)$ follows a Brownian motion with zero trend. In that case, $\varepsilon(t)$ is an unbiased estimator of $(1/s)\int_{t}^{t+s}\varepsilon(v)dv$ and hence, OLS is unbiased.

To develop this intuition further, the analysis now considers two different cases. The first is when $\varepsilon(t)$ is a mean reverting Brownian motion. However, since assuming a mean reverting Brownian motion may be considered as too simplistic, the second case allows $\varepsilon(t)$ to follow <u>any</u> deterministic path over some period [0,T), and considers that as part of a stationary cycle of fixed period T.

<u>Proposition 2</u>

If $\varepsilon(t)$ is described by a mean reverting Wiener process

$$d\varepsilon_t = - \lambda \, \varepsilon_t \, dt + \sigma \, dz$$

then the bias through estimating (3) using OLS and time aggregated data is

$$\frac{\alpha - E(\hat{\alpha})}{\alpha} = \left[e^{-\lambda s} - (1 - \lambda s) \right]/(\lambda s)$$

<u>Proof in Appendix</u>

For λ,s small, the bias described by the proposition is approximated by $\lambda s/2$. With $\lambda = 0$, $\varepsilon(t)$ is an unbiased estimator of $(1/s) \int_{v}^{v+s} \varepsilon(t)\,dt$ and OLS is unbiased. However, with a positive mean reversion, $\lambda > 0$, then $\varepsilon(t)$ <u>overestimates</u> $(1/s) \int_{v}^{v+s} \varepsilon(t)\,dt$ and OLS compensates by choosing $\hat{\alpha} < \alpha$. Moreover, the magnitude of the bias increases with both λ and s. {Notice that this result is very different to the standard measurement error results. There, attenuation is obtained because of excess noise in the proxy variable. Here, we may even have an upward bias (in the case $\lambda < 0$).)

Given this, the immediate question is whether there are theoretical grounds for expecting mean reversion in unemployment. Indeed, the assumption that the matching function is increasing in U suggests that there will be some mean reversion. For example, suppose there is an increase in unemployment. With more unemployed workers looking for jobs, the existing vacancies should match more quickly so that the aggregate flow into employment should increase. Thus if unemployment increases, it should tend to decrease again (given V). By the above, such a reversion will cause $\hat{\alpha}$ to be downward biased.

An interesting result here is that if the bias is small, it is linear in s. Thus estimating the size of the bias can be obtained by doubling the size of the measuring interval and the estimated bias is then the observed fall in $\hat{\alpha}$.

Alternatively, suppose unemployment follows an arbitrary, but deterministic cycle of fixed period T, where $\bar{U}$ is defined such that

$$\int_0^T \varepsilon(t)\, dt = 0 \qquad (10)$$

With such an arbitrary cycle, consider how the data is generated by the sequence of observations $t = v, v+s, v+2s, \ldots$. With a cycle of fixed period T, these observations will pick out a sequence of points within the cycle $[0,T]$. For example, suppose $v = 0$ and $s = 2T/3$. Define the points in the cycle $t_0 = 0$, $t_1 = T/3$, $t_2 = 2T/3$. It follows that the sequence of observations $0, s, 2s, 3s, ..$ picks out the sequence $t_0, t_2, t_1, t_0, t_2, t_1, t_0, \ldots$ in the cycle. The point is that we would only observe 3 points in the cycle, where asymptotically, each point is observed at the same relative frequency (i.e. one third). In fact, this result is true for any proper fraction s/T (where s, T have no common factors). If for example $v = 0$ and $s/T = 17/387$, we will observe 386 points in the cycle, where $t_i = iT/387$, $i = 0$ to 386, and asymptotically, each point will be observed with the same frequency.

Now suppose s/T is an irrational number. It follows that no point in the cycle will be observed more than once. Suppose that asymptotically, the observed points will be uniformally distributed along the line [0,T]. Then asymptotically, (9) implies

$$\hat{\alpha} = \alpha \frac{\int_0^T \varepsilon(t) \left[(1/s) \int_t^{t+s} \varepsilon(v)\, dv \right] dt}{\int_0^T \varepsilon(t)^2\, dt} + O(\varepsilon/\bar{U}) \qquad (11)$$

since $\int_0^T \varepsilon(t)\, dt = 0$.

<u>Proposition 3</u>

If $\hat{\alpha}$ is defined by (11) and $\text{var}(\varepsilon) > 0$, then $\hat{\alpha} < \alpha$.

<u>Proof</u> in Appendix

To obtain some idea of the potential size of the bias with cyclical data, suppose $U(t) = \bar{U} + \varepsilon \sin(2\pi t/T)$, where T is the period of the cycle and $\varepsilon/\bar{U}$ is small. Then solving (11) implies

$$\hat{\alpha} = \alpha\, [\sin(2\pi s/T)]/(2\pi s/T) + O(\varepsilon/\bar{U})$$

The size of the bias depends on the size of s/T. The following diagram depicts this relationship

figure 1

Notice that because of aliasing type problems, there is not a monotonic relationship between $\hat{\alpha}$ and s/T. The diagram clearly shows that once s/T >

0.5, the estimaton procedure is producing nonsense results. However, for s/T small, it can be shown that $\hat{\alpha}$ is approximately quadratic in s/T.

The following table describes the magnitude of the bias (as a percentage of α), depending on whether the data is weekly, monthly etc, and on the period of the cycle (T).

Table 1 : Magnitude of Temporal Aggregation Bias (%)

		Cycle Period (Months)				
		1	3	6	12	60
Data Period	Weekly	36	4	1	0.3	0
	Monthly	100	59	17	4	0.2
	Quarterly	100	100	100	59	3
	Yearly	100	100	100	100	24

This table seems to suggest that the temporal aggregation bias will not be too important if the data collected is monthly (or less). If the data is quarterly or yearly, the size of the bias will depend crucially on the spectral frequency of the data.

The data of Blanchard and Diamond (1989) is monthly and has a cycle of approximately 60 periods. The corresponding bias of 0.2% is therefore statistically irrelevant.

APPENDIX

Proof to Proposition 2

Given the process describing $\varepsilon(.)$. Then given $\varepsilon(t)$,

$$\varepsilon(t+v) = e^{-\lambda v}\varepsilon(t) + \int_0^v \sigma\, e^{-\lambda(v-x)}\, dz(x)$$

Hence given $\varepsilon(t)$,

$$(1/s)\int_0^s \varepsilon(t+v)\, dv = (1/s)\int_0^s \left[e^{-\lambda v}\varepsilon(t) + \int_0^v \sigma\, e^{-\lambda(v-x)}\, dz(x) \right] dv$$

$$= \varepsilon(t)[1-e^{-\lambda s}]/(\lambda s) + \theta \qquad \text{(A1)}$$

where $\theta = (1/s)\int_0^s \int_0^v \sigma\, e^{-\lambda(v-x)}\, dz(x)\, dv$, and θ is a random variable which has zero expectation and is independent of $\varepsilon(t)$. Hence using (A1) :

$$\text{cov}(\ \varepsilon(t),\ (1/s)\int_t^{t+s} \varepsilon(v)\, dv) = \left[(1-e^{-\lambda s})/(\lambda s)\right] \text{var}(\varepsilon(t)) \qquad \text{(A2)}$$

Substituting (A2) into (9) and rearranging gives the result stated.

Proof to Proposition 3

From the text, we know

$$\hat{\alpha} = \alpha \frac{\int_0^T \varepsilon(t) \left[(1/s)\int_t^{t+s} \varepsilon(v)\, dv\ \right] dt}{\int_0^T \varepsilon(t)^2\, dt} + O(\varepsilon/\bar{U}) \qquad \text{(A3)}$$

The Proposition is proven if

$$\int_0^T \varepsilon(t) \left[(1/s)\int_t^{t+s} \varepsilon(v)\, dv\ \right] dt < \int_0^T \varepsilon(t)^2\, dt \qquad \text{(A4)}$$

Now the Cauchy-Schwarz inequality implies

$$\int_0^T \varepsilon(t) \left[(1/s)\int_t^{t+s} \varepsilon(v)\, dv\ \right] dt$$

$$\le \left[\int_0^T \left[(1/s)\int_t^{t+s} \varepsilon(v)\, dv\ \right]^2 dt \right]^{0.5} \left[\int_0^T \varepsilon(t)^2\, dt \right]^{0.5} \qquad \text{(A5)}$$

Also

$$\left[(1/s) \int_t^{t+s} \varepsilon(v)\, dv \right]^2 \le (1/s) \int_t^{t+s} \varepsilon(v)^2\, dv \tag{A6}$$

with strict inequality for some t if var(ε) > 0. Substituting (A6) into (A4) implies

$$\int_0^T \varepsilon(t) \left[(1/s) \int_t^{t+s} \varepsilon(v)\, dv \right] dt < \left[\int_0^T \left[(1/s) \int_t^{t+s} \varepsilon(v)^2\, dv \right] dt \right]^{0.5} \left[\int_0^T \varepsilon(t)^2\, dt \right]^{0.5} \tag{A7}$$

Finally, ε(t) a cycle of fixed period T implies

$$\int_0^T \left[(1/s) \int_t^{t+s} \varepsilon(v)^2\, dv \right] dt = \int_0^T \varepsilon(t)^2\, dt \tag{A8}$$

Using (A8) in (A7) implies (A4), which completes the proof.