Department of Economics

Working Paper No. 0210

http://www.fas.nus.edu.sg/ecs/pub/wp/wp0210.pdf

Impact of Systematic Sampling on Causality in the presence of Unit Roots

Rajaguru GULASEKARAN

Abstract: Quite contrary to the stationary case where systematic sampling preserves the direction of Granger causality, this paper shows that systematic sampling of integrated series may induce spurious causality, even if they are used in differenced form.

Key words: Systematic Sampling, Causality, Unit Roots, Cross covariance

JEL Classification: C15, C32, C43

 Correspondence: Department of Economics, National University of Singapore, 1 Arts Link, Singapore-117570, email: artp9449@nus.edu.sg, Tel : (+65) 68746132, Fax: (+65) 67752646.

1. Introduction

In applied econometric literature, the causal inferences are often made based on highly temporally aggregated or systematically sampled data. Marcelinno (1999) analysed the effects of temporal aggregation on time series properties such as exogenity, Granger causality, structural invariance, integration, cointegration, impulses responses, and measures of persistence and found that Granger causality is not invariant upon temporal aggregation. Cunningham and Vilasuso (1995), using Monte Carlo experiments, find that temporal aggregation is about two to ten times more likely to lead to false causal inference compared to systematic sampling. Abeysinghe and Gulasekaran (2000) have further confirmed the negative effects of temporal aggregation on causal inference. As opposed to temporal aggregation Wei (1982), using Geweke's linear decomposition, shows that systematic sampling preserves the causal direction. This result, however, applies only to stationary series. In this note we demonstrate that in the presence of unit roots systematic sampling may turn a unidirectional causality to a bi-directional one. We establish these results through a cross-covariance analysis and a Monte Carlo study.

2. Systematic Sampling and Granger Causality

Let z_t $(t=1,2,\ldots,n)$ be the equally spaced basic series. Systematic sampling (hereafter, s.sampling) means the construction of the series $Z_T=z_{mT}$ $(T=1,2,\ldots,N$ and $n=mN)$ by sampling from z_t at every mth interval (m is an integer). Let z_{1t}~I(d_1) and z_{2t} ~I(d_2) such that the differenced series $w_{1t} = (1-L)^{d_1} z_{1t}$ and $w_{2t} = (1-L)^{d_2} z_{2t}$ follow a weakly stationary process with mean zero and variance covariance matrix[1]

$$\Gamma^w(k) = \text{cov}(w_t, w_{t+k}) = E(w_t w'_{t+k}) = \begin{pmatrix} g_{11}^w(k) & g_{12}^w(k) \\ g_{21}^w(k) & g_{22}^w(k) \end{pmatrix}, \tag{1}$$

where $w_t=(w_{1t},w_{2t})'$. Let L′ be the backward shift operator on the s.sampled time unit T such that $L'Z_{1\,T} = Z_{1\,T-1}$ and $L'Z_{2\,T} = Z_{2\,T-1}$. Further let

$$W_{1T} = (1-L')^{d_1} Z_{1T} = (1-L^m)^{d_1} z_{1mT} = (1+L+.....+L^{m-1})^{d_1} w_{1\,mT}, \tag{2}$$

$$W_{2\,T} = (1-L')^{d_2} Z_{2\,T} = (1+L+....+L^{m-1})^{d_2} w_{2\,mT} \tag{3}$$

be the differenced series derived from the s.sampled series.

The cross covariance between $W_{1\,T}$ and $W_{2\,T+k}$ works out to be[2]

$$g_{12}^W(k) = Cov(W_{1\,T}, W_{2\,T+k}) = (1+L+......+L^{m-1})^{d_1+d_2} g_{12}^w(mk + d_1(m-1)) \forall k \geq 0 \tag{4}$$

$$g_{21}^W(k) = Cov(W_{2\,T}, W_{1T+k}) = (1+L+....+L^{m-1})^{d_1+d_2} g_{21}^w(mk + d_2(m-1)) \forall k \geq 0. \tag{5}$$

Here L operates on the index of $\gamma^w_{ij}(k)$ such that $L\gamma^w_{ij}(k)= \gamma^w_{ij}(k-1)$.

The expressions (4) and (5) show that the cross covariance of the s.sampled series is the weighted sum of the cross covariances of the basic series. In order to determine the consequences of s.sampling on Granger causality, we consider the following cases.

Case1: z_1 and z_2 are uncorrelated with each other

If the basic series z_1 and z_2 are uncorrelated with each other then $g^w_{ij}(k)=0$, for all k and i,j=1,2; i≠j. From (4) and (5) we can see that $g^W_{ij}(k)=0$, for all k and i,j=1,2; i≠j. Thus, the s.sampled process Z_1 and Z_2 also remains uncorrelated. This is true for all sampling intervals, m, and the orders of integrations, d_1 and d_2.

Case 2: Causality between z_1 and z_2 is one-sided

---

[1] $\gamma^w_{ii}(k)$ is the autocovariance of the i-th series, w_{it}, at lag k. When k=0, it is simply the variance of the series. $\gamma^w_{ij}(k)$ is the cross covariance between i-th and j-th series at lag k

[2] see appendix for the detail derivation

Without loss of generality assume that causality runs from z_1 to z_2 such that first K cross covariances are non-zero, i.e.

$$g_{12}^{w}(k) = \begin{cases} \neq 0 & if\ 1 \leq k \leq K \\ = 0 & otherwise. \end{cases}$$

S.sampling changes one-sided causality from z_1 to z_2 to two-sided causal relationship if and only if $g^{W}_{21}(k) \neq 0$ for some k>0. Since $g^{W}_{21}(k)$ is the weighted sum of the cross covariances of the basic process, $g^{w}_{21}(k)$, the bi-directional causality occurs iff $(1+L+L^2+\ldots+L^{m-1})^{d_1+d_2} g_{21}^{w}(mk+d_2(m-1)) \neq 0$ for some k>0. That is, a feedback relationship occurs due to s.sampling iff some of {mk+d_2(m-1), mk+d_2(m-1)-1,…., mk-d_1(m-1)} be negative for some k>0. This implies that the uni-directional causation becomes bi-directional iff at least mk-d_1(m-1)≤-1, or $d_1 \geq \frac{mk+1}{m-1}$. In particular, the cross covariance from Z_2 to Z_1 at lag 1 becomes non-zero when $d_1 \geq \frac{m+1}{m-1}$. This implies that, in general, feedback (spurious) relationship occurs due to s.sampling iff the order of integration of the series z_1 is such that, d_1≥1. The interesting feature of the above condition is that the spurious bi-directionality due to s.sampling is independent of the order of integration of the series z_2, d_2, when the true uni-directional causality runs from z_1 to z_2.

Case 3: Stationary Process

The result for the case when d_1=d_2=0 is consistent with the findings of Wei (1982) and Cunningham and Vilasuso (1995) that s.sampling preserves unidirectional causality.

3. Some Monte Carlo Results

It would be useful to examine how causal inferences are affected by s.sampling of non-stationary variables when m changes. We do this through a Monte Carlo study. We conducted an extensive Monte Carlo simulation based on the following process:

$$\begin{pmatrix} (1-L)^{d_1} z_{1t} \\ (1-L)^{d_2} z_{2t} \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} (1-L)^{d_1} z_{1t-1} \\ (1-L)^{d_2} z_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \qquad (6)$$

with $\mathrm{var}(e_t) = \mathrm{var}\left[\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}\right] = \begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{pmatrix} = \Sigma_e$.

The contemporaneous correlation between z_1 and z_2 is set to zero, i.e. $\sigma_{12}=0$, in order to assess how s.sampling affect contemporaneous correlation. We set $\varphi_{12}=0$ so that the causal relationship is unidirectional from z_1 to z_2. We set the sample size n=480 to represent 40 years of monthly data of z_1 and z_2. Then systematic samples of size N=160 (i.e. m=3, quarterly) and N=40 (m=12, annual) series were constructed from the basic series z_1 and z_2. We estimate the model (6), test for the significance of estimated parameters and test for the contemporaneous correlation[3] between the variables at each stage of s.sampling.

The exercise was replicated 1000 times. Two scenarios, stationary and non-stationary, were considered. In order to analyze the effect of s.sampling on unidirectional causality (z_1 to z_2) when the series are non-stationary, the following cases are considered: (1) $d_1=1$ and $d_2=0$, (2) $d_1=1$ and $d_2=1$ (3) $d_1=0$ and $d_2=2$. Differencing is done on the s.sampled series to induce stationarity. Since, our results confirm the finding that the s.sampling preserves the direction of causality when $d_1=d_2=0$, these results are not reported.

The Monte Carlo results for the non-stationary cases are reported in Table 1 (panel 1 through 3) for the sampling intervals m=3 and m=12. In all the cases, the parameters take

[3] The Lagrange Multiplier test developed by Breusch and Pagan (1980) is used to test the contemporaneous correlation between the residuals.

the values φ_{11}=0.9, φ_{22}=0.1, φ_{12}=0.0 and φ_{21} varies from 0.1 to 0.9. This allows us to examine the effect of s.sampling on causality when the true unidirectional causality changes from weaker (φ_{21} =0.1) to stronger (φ_{21} =0.9). The entries in Table 1 show the rejection frequencies (%) of the hypothesis that the parameter is zero at the 5% level of significance.

The table shows that When d_1=1 (see panels 1 and 2) the one-sided causality becomes a feedback system. This confirms our theoretical findings based on cross coavariances that when $d_1 \geq 1$ s.sampling introduces a spurious relationship. This effect becomes more prominent when both d_1 and d_2 are positive (integers) or the strength of the unidirectional causality from the basic non-stationary series becomes stronger or when m=3 or the both. Finally, the results reported in panel 3 for the case d_1=0 and d_2=2 corroborate our theoretical finding that spurious causal relationship due to s.sampling is independent of d_2. Thus if the uni-directional causality runs from a non-stationary series to a stationary or non-stationary series, there is a high likelihood of detecting spurious bi-direcional causality. Furthermore, in all cases s.sampling induces high contemporaneous correlation between the variables.

================

Table 1

================

4. Conclusion

This paper has documented the existence of Granger causality distortion due to s.sampling in the presence of unit roots. This distortion has been found to depend on strength of the causality and the sampling intervals. The spurious bi-directionality due to s.sampling is induced by the presence of unit roots in the causal variable rather than the effect variable. The cross covariance analysis reveals that the unrelated series remain unaltered due to s.sampling even if the series contain unit roots. But the absence of Granger causality in s.sampled series does not imply that the series are unrelated in nature. Granger causality testing with s.sampled as well as temporally aggregated data is less likely to resolve the debates on causal directions.

Acknowledgement

I wish to thank Tilak Abeysinghe and S.S.Pattnayak for helpful comments and suggestions.

Appendix

Proof of (4)

Define the forward shift operator $F=L^{-1}$ such that $Fw_t = w_{t+1}$ and $F\gamma_{ij}(k) = \gamma_{ij}(k+1)$.

$$g_{12}^{W}(k) = E[W_{1t}W_{2\,t+k}]$$

$$= E[(1+L+\cdots+L^{m-1})^{d_1} w_{1\,mt}(1+L+\cdots+L^{m-1})^{d_2} w_{2\,m(t+k)}]$$

$$= E[(c_0 w_{1\,mt} + c_1 w_{1\,mt-1} + \cdots c_{d_1(m-1)} w_{1\,mt-d_1(m-1)})$$
$$(e_0 w_{2\,mt+mk} + e_1 w_{2\,mt+mk-1} + \cdots e_{d_2(m-1)} w_{2\,mt+mk-d_2(m-1)})$$

$$= c_0[e_0 g_{12}^{w}(mk) + e_1 g_{12}^{w}(mk-1) + \cdots + e_{d_2(m-1)} g_{12}^{w}(mk - d_2(m-1))]$$
$$+ c_1[e_0 g_{12}^{w}(mk+1) + e_1 g_{12}^{w}(mk) + \cdots + e_{d_2(m-1)} g_{12}^{w}(mk - d_2(m-1)+1)]$$
$$+ \cdots\cdots$$
$$+ c_{d_1(m-1)}[e_0 g_{12}^{w}(mk + d_1(m-1)) + e_1 g_{12}^{w}(mk + d_1(m-1)-1) \cdots$$
$$+ e_{d_2(m-1)} g_{12}^{w}(mk - d_2(m-1) + d_1(m-1))]$$

$$= c_0[(1+L+\cdots L^{m-1})^{d2} g_{12}^{w}(mk)] + c_1[(1+L+\cdots L^{m-1})^{d2} g_{12}^{w}(mk+1)] + \cdots$$
$$+ c_{d_1(m-1)}[(1+L+\cdots L^{m-1})^{d_2} g_{12}^{w}(mk + d_1(m-1)]$$

$$= (1+L+\cdots+L^{m-1})^{d_2}[c_0 g_{12}^{w}(mk) + c_1 g_{12}^{w}(mk+1) + \cdots + c_{d_1(m-1)} g_{12}^{w}(mk + d_1(m-1)]$$

$$= (1++L\cdots+L^{m-1})^{d_2}(1+F+\cdots+F^{m-1})^{d_1} g_{12}^{w}(mk)$$

$$= (1+L+\cdots+L^{m-1})^{d_1+d_2} F^{d_1(m-1)} g_{12}^{w}(mk)$$

$$= (1+L+\cdots+L^{m-1})^{d_1+d_2} g_{12}^{w}(mk + d_1(m-1))$$

Thus,

$$g_{12}^{W}(k) = (1+L+\cdots+L^{m-1})^{d_1+d_2} g_{12}^{w}(mk + d_1(m-1))$$

Hence the result.

References:

Abeysinghe, T. and Gulasekaran, R., 2000. Impact of Temporal Aggregation on Causal Inferences, Working paper, National University of Singapore.

Cunningham, S.R., and Vilasuso, J. R.1995. Time Aggregation and Causality Tests: Results from a Monte Carlo Experiment, Applied Economics Letters 2, pp 403-405.

Marcellino, M. 1999. Some Consequences of Temporal Aggregation in Empirical Analysis, Journal of Business and Economic Statistics. 17, pp 129-36.

Pierse, R.G., and Snell, A.J. 1995. Temporal Aggregation and Power of Tests for a Unit Root, Journal of Econometrics, 65, pp. 333-45

Sims, C.A. 1971. Discrete Approximations to Continuous Time Distributed Lags in Econometrics, Econometrica 39, pp 545-63.

Wei, W.W.S. 1982. The effects of Systematic Sampling and Temporal Aggregation on Causality – A Cautionary Note, Journal of the American Statistical Association 77, pp. 316-19.

Wei, W.W.S. 1990. Time Series Analysis: Univariate and Multivariate Methods. Addison-Wesley, California.

Table 1: Granger Causality distortion (in %) due to systematic sampling

Panel 1: d_1=1 d_2=0,(j_{11}=0.9 j_{22}=0.1, j_{21}=0.1 to 0.9 and j_{12}=0.0)

		$\varphi_{21}\rightarrow$								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
m=3	φ_{12}	5.5	10.2	13.5	24.4	35.8	48.3	56.6	69.7	77.6
	φ_{21}	90.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ρ_{12}	33.7	86.6	99.2	99.6	100.0	100.0	100.0	100.0	100.0
m=12	φ_{12}	6.6	9.4	13.5	19.6	20.9	29.5	34.9	40.2	44.4
	φ_{21}	9.0	15.7	22.9	21.9	21.4	15.7	14.6	13.8	9.4
	ρ_{12}	30.2	72.9	94.8	99.6	99.8	100.0	100.0	100.0	100.0

Panel 2: d_1=d_2=1, (j_{11}=0.9 j_{22}=0.1, j_{21}=0.1 to 0.9 and j_{12}=0.0)

		$\varphi_{21}\rightarrow$								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
m=3	φ_{12}	6.3	16.6	28.0	40.6	54.9	64.2	73.0	79.0	85.2
	φ_{21}	96.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ρ_{12}	49.9	95.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0
m=12	φ_{12}	5.9	8.9	11.3	19.4	24.3	30.6	35.0	43.1	47.7
	φ_{21}	52.5	77.6	84.9	85.9	87.5	88.7	91.8	94.1	93.6
	ρ_{12}	78.3	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Panel 3: d_1=0 d_2=2, (j_{11}=0.9 j_{22}=0.1, j_{21}=0.1 to 0.9 and j_{12}=0.0)

		$\varphi_{21}\rightarrow$								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
m=3	φ_{12}	5.5	5.8	3.2	4.2	4.5	4.3	5.5	5.5	4.3
	φ_{21}	99.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ρ_{12}	11.2	21.2	39.4	58.0	77.5	88.4	93.9	97.5	98.9
m=12	φ_{12}	4.8	6.0	4.3	3.9	4.5	5.3	4.7	4.7	4.9
	φ_{21}	90.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ρ_{12}	12.0	26.1	49.5	69.7	70.1	79.0	81.8	85.6	87.3