Estimating Long Memory Time-Series-Cross-Section Data

Wen-Jen Tsay*

July 31, 2008

The Institute of Economics, Academia Sinica, Taiwan

Abstract

This paper considers the estimation and inference problems of a general class of time-series-cross-section (TSCS) models consisting of stationary or nonstationary long memory regressors and errors, while allowing for cross-correlations and serial correlations in cross-section and time dimensions, respectively. Although the applicability of this class of TSCS models is far-reaching, we show that each regression coefficient of these models can be easily tested with the critical values from the standard normal distribution based on the approach proposed in this paper. Furthermore, our approach is built on Robinson's (1998) long-run variance estimator and thus does not involve the difficult problems of choosing a kernel function or a bandwidth parameter. We also demonstrate that, under various combinations of long memory processes and cross-section dimensions, the finite sample performance of our method for this class of long memory TSCS models is promising even though the time span is only 20. We then apply this method to re-examine the welfare spending studies of Hicks and Swank (1992). The testing results are different from the findings in Hicks and Swank (1992) and those in Beck and Katz (1995), because we find a weak but significant positive voter turnout effects when the number of differencing is equal to 1.

Key words: SURE model; TSCS model; Long memory; Spurious regression

*We would like to thank two anonymous referees and the Editor for their valuable comments and suggestions. We also thank Professor Duane Swank for generously providing the data for the empirical study conducted in this paper. Financial support from the NSC-94-2415-H-001-018 of Taiwan is gratefully acknowledged. Wen-Jen Tsay: The Institute of Economics, Academia Sinica. 128 Academia Road, Sec. 2, Taipei, Taiwan, R.O.C. Tel.: (886-2) 27822791-296. Fax: (886-2) 27853946. E-Mail: wtsay@econ.sinica.edu.tw

1 Introduction

This paper considers the estimation and inference problems associated with the temporally-dominated time-series-cross-section (TSCS) models where a fixed number of N units is observed for a period of time T. The TSCS model is a special case of the seemingly unrelated regressions (SURE) model proposed by Zellner (1962). The estimation of the SURE model has been considered in Parks (1967), Kmenta (1971), and Guilkey and Schmidt (1973), to name a few. Empirical studies based on the TSCS model are far-reaching in social science, including economics, sociology, and political science. Beck and Katz (1995) provide a list of 14 papers using TSCS data. Five of them are published in the *American Political Science Review*.

Before the publication of Beck and Katz (1995), the generalized least squares (GLS) method of Parks (1967) had been the most popular method for handling the TSCS model. Beck and Katz (1995) show that Parks' method can produce a misleading standard error that underestimates variability by 50% or more. Beck and Katz (1995) therefore suggest a panel-corrected standard errors (PCSE) estimator when the errors of the TSCS model are assumed to be contemporaneously correlated and panel heteroscedastic. As clearly spelled out in page 634 of their paper, Beck and Katz (1995) offer the PCSE estimator that is both "easier to implement and produces accurate standard errors", and their strategy is to retain the ordinary least squares (OLS) parameter estimates, but replace the OLS standard errors with PCSE.

The common feature shared by Parks (1967) and Beck and Katz (1995) is that they only consider the weakly dependent or $I(0)$ processes. When the data-generating processes (DGP) are the unit root, or $I(1)$, processes, there is no well accepted inference method for the TSCS model consisting of $I(1)$ processes. This leads Beck (2001, p. 280) to document: *But we know little about nonstationary TSCS data.*

The presence of the well-known autoregressive fractionally-integrated moving average (ARFIMA), or $I(d)$, process in many time series seems to incur an even more complicated structure within the TSCS models. The $I(d)$ process is first introduced by Granger (1980), Granger and Joyeux (1980), and Hosking (1981). Many data are found to be well described with the $I(d)$ processes, including the level of Nile River, *ex post* real interest rate, inflation, and forward premium. Baillie (1996) provides an overview about the empirical applications of the $I(d)$ process in the economics and finance literature.

The $I(d)$ process also has been observed in the political science time series. Box-Steffensmeier and Smith (1996, 1998) use the $I(d)$ process to explain the dynamics of aggregate partisanship and macroideology. Dolado et al. (2003) find that pure fractional process fits the Spanish opinion polls well. Moreover, Davidson et al. (2006) use a fractionally cointegrating vector error correction model to describe the poll data on approval of the performance of prime ministers and governments in the UK. Recognizing the increasing importance of the $I(d)$ process in political science, *Electoral Studies* (2000, vol. 19) devotes a special issue to address the recent advances in the analysis of political time series with $I(d)$ processes, including the dynamics of the aggregate political popularity studied by Byers et al. (2000) and the discussions in Box-Steffensmeier and Tomlinson (2000) about the potential of the $I(d)$ process to unify and simplify time series analysis. In that issue, Lebo et al. (2000, p. 32) point out that $I(d)$ processes "are by no means rare and need to be accounted for in their models." Indeed, many time series data of interest to political scientists can be characterized by $I(d)$ processes – in particular, the macropartisanship, presidential approval, the monthly and quarterly indices of consumer sentiment, and percentage liberalism in Supreme Court decision making. Furthermore, Lebo et al. (2000, p. 40) claim that *Robinson's and Sowell's estimator agree that fractional integration is extremely common among political time series.*

The distinguishing feature of the $I(d)$ process is that its autocovariance function declines at a slower hyperbolic rate, instead of the geometric rate found in conventional stationary and invertible ARMA processes. When $d > 0$, the $I(d)$ process is often called the long memory process. The $I(d)$ process can display nonstationary phenomenon when $d \geq 1/2$, otherwise, it is covariance stationary. As a consequence, with a continuum of possible values of d, both $I(0)$ and $I(1)$ processes can be viewed as special cases of the $I(d)$ process, where the differencing parameter d is 0 and 1, respectively. The $I(d)$ process therefore provides a unified and flexible modeling framework for empirical applications.

Given that fractional integration is extremely common among political time series and empirical studies based on the TSCS model are far-reaching in social science, we combine the long memory and TSCS literature into a unified framework. Clearly, the coverage of the long memory TSCS model includes the usual TSCS model with the AR(1) process as one of its special cases. The nonstationary TSCS processes discussed in Beck (2001) are also subsumed with the long memory TSCS data. To deal with the estimation and inference issues concerning the long memory TSCS model, we follow the spirit of Beck and Katz (1995, p. 634) to offer a new method that is both easier to implement and produces accurate standard errors for the long memory TSCS data. This method combines a nonparametric inference method based on the MD (multiple-differenced) estimator of Tsay (2007) and the long-run variance estimator of Robinson (1998). We thus name this procedure as the MD standard errors (MDSE) estimator. The MDSE method not only can deal with the aforementioned comment of Beck (2001, p. 280) about nonstationary TSCS data, but also can handle the complicated structure of TSCS data consisting of AR, MA, unit root, and stationary and nonstationary $I(d)$ processes in all the regressors and errors, provided that the regularity conditions outlined in the following Theorem 1 are satisfied. Consequently, the case of cointegration also can be taken care of with the MDSE estimator. On the other hand, the standard cointegration techniques cannot be applied to the long memory TSCS data easily due to the presence of $I(d)$ processes in the data. Even though we can generalize the standard cointegration techniques to the long memory TSCS data, the resulting OLS estimator generally has a non-standard distribution which is difficult to use in inference, because we need to simulate different sets of critical values for the complicated combinations of different $I(d)$ regressors and errors within and across cross units, and at various sample sizes under this circumstance.

The MD estimator of Tsay (2007) essentially first-differences both the dependent variable and the regressors M (a positive integer) times before running the subsequent OLS estimator in order to control the impacts of the fixed effect on the estimation results and to ensure that the resulting MD estimator is asymptotically normally distributed under suitable regularity conditions. The idea behind the MDSE is similar to the one used in Beck and Katz (1995) who retain the OLS parameter estimates, but replace the OLS standard errors with PCSE. The MDSE estimator retains the MD parameter estimates and applies the long-run variance estimator of Robinson (1998) to calculate the MDSE based on the residuals from the MD estimation. Most importantly, the use of multiple-differencing does not always result in the efficiency loss of the MD estimator as compared to the OLS counterpart as clearly investigated in Tsay (2007, p. 835 and p. 836), even though it can always eliminate the impacts of the fixed effect on the regression coefficient estimates.

There are at least four distinguished advantages of using the MDSE estimator for the long memory TSCS data. First of all, the use of multiple-differencing when implementing the MDSE estimator can resolve the spurious regression problems induced by the stationary or nonstationary long memory processes. The well-known spurious effect has been considered in Granger and Newbold (1974) and Phillips (1986) when considering the relationship between independent $I(1)$ processes. However, the spurious regression effect might happen more often than that documented in Phillips (1986), because Tsay and Chung (2000) find that when we regress a long memory $I(d_1)$ process on another independent long memory $I(d_2)$ process, no matter whether these processes are stationary or not, as long as their orders of integration sum up to a value greater than 1/2, the usual t-ratio statistic becomes divergent and spurious effects occur.

Second, the MDSE method can successfully handle the huge number of various $I(d)$ processes inherent in the TSCS data under suitable regularity conditions. In particular, when there are K stochastic regressors in each cross unit, the total number of differencing parameters in the regressors and the errors is $N \times (K+1)$. Moreover, the typical observations used for TSCS data are relatively small in time-dimension, indicating that the estimation of the AR, MA and the fractional differencing parameters cannot be accurate with the usual time-domain and frequency-domain estimation methods under this circumstance. However, this huge number of differencing parameters is not a problem to the MDSE method, because these differencing parameters are not estimated when calculating the MDSE. In addition, there is no problem concerning the selection of the number of AR and that of MA parameters with the MDSE method either, because these parameters are not estimated when implementing the MDSE estimator. Section 3 demonstrates the superior ability of the MDSE estimator in controlling the impacts of AR, MA, and fractional differencing parameters on the testing performance concerning the regression coefficients in the following Monte Carlo experiment even though the time span is only 20 and the regressors are nonstationary processes.

Third, the MDSE method is powerful enough to deal with the model allowing for cross-correlations and serial correlations in both cross-section and time dimensions. This is important in the TSCS literature, because it is well-known that many regressors in the TSCS model may be serially and contemporaneously correlated.

Fourth, the implementation of MDSE estimator is straightforward, and each regression coefficient of the long memory TSCS models can be tested with the critical values from the standard normal distribution. No restriction is imposed on the relative magnitude between N and T as long as the MD estimator for the long memory TSCS model can be implemented.

The remaining parts of this paper are arranged as follows: Section 2 presents the test statistics and the main results. In Section 3 the theoretical findings generated from the following Theorem 1 are verified through a Monte Carlo experiment. Section 4 applies the MDSE method to the welfare spending data of Hicks and Swank (1992). Section 5 provides a conclusion.

2 The models and main results

Consider the generic TSCS model as:

$$y_{i,t} = \mathbf{x}_{i,t}^{\top}\beta + u_{i,t}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where $y_{i,t}$ is a scalar observation on the i-th dependent variable at time t, $\mathbf{x}_{i,t}$ is a $(K \times 1)$ vector of observations for the stochastic regressors, β is a $(K \times 1)$ vector of unknown regression coefficients to be estimated and tested, and $u_{i,t}$ is a scalar random disturbance with mean zero. For expositional purposes, denote $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,T})^{\top}$ as the column vector containing the observations on the i-th dependent variable from time 1 to T, while $\mathbf{x}_i^{\top} = (\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \ldots, \mathbf{x}_{i,T})$ represents the corresponding $(K \times T)$ matrix of observations for the stochastic regressors, and $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,T})^{\top}$ is similarly defined. The observations for $\mathbf{y}_i$, $\mathbf{x}_i$, and $\mathbf{u}_i$ across i can be further stacked as:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}, \qquad \mathbf{U} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}, \qquad \text{and} \qquad \mathbf{Y} = \mathbf{X}\beta + \mathbf{U}. \tag{2}$$

When $\mathbf{x}_i$ is non-stochastic and $\mathbf{u}_i$ is contemporaneously correlated but not serially correlated, the method of Zellner (1962) can be used to estimate the model in (1). Under the same conditions, though $u_{i,t}$ in each equation is relaxed to be the AR(1) process, the GLS estimator of Parks (1967) had been widely adopted in the literature. Based on Monte Carlo simulations, Beck and Katz (1995) demonstrate that, under some circumstances, Parks' method might produce a misleading standard error that underestimates variability by 50% or more. Beck and Katz (1995) propose an alternative PCSE estimator for testing the value of β, the major interest of research, in (1) with the critical values from the standard normal distribution given that $\mathbf{u}_i$ are contemporaneously correlated and panel heteroscedastic. In other words, the PCSE estimator of Beck and Katz (1995) is not designed for the TSCS data consisting of stationary or nonstationary $I(d)$ processes, including the unit root which is the special case of the nonstationary $I(d)$ process.

This paper extends the MD-based method of Tsay (2007) to resolve the inference problems associated with the long memory TSCS data. Before illustrating the details of the proposed method, we define the underlying process clearly. Adapting the notation of Robinson (2005), we denote a short memory process $\{\eta_t, t = 0, \pm 1, \ldots\}$ as a zero-mean covariance stationary process with spectral density $f(\lambda)$, which is bounded and bounded away from zero. Thus, the short memory process η_t includes the stationary and invertible ARMA process as its special case.

With η_t and $d < 1/2$, we define the corresponding stationary $I(d)$ process ξ_t as:

$$\xi_t = \triangle^{-d}\eta_t, \quad t = 0, \pm 1, \ldots, \tag{3}$$

where $\triangle = 1 - L$, and L is the usual lag operator. The associated truncated ξ_t process is computed as:

$$\xi_t^{\#} = \xi_t \mathbf{1}(t \geq 1), \quad t = 0, \pm 1, \ldots, \tag{4}$$

where $\mathbf{1}$ is the indicator function. Based on the process $\xi_t^{\#}$ in (4), for $q \geq 0$, we define its corresponding Type I $I(q+d)$ process as:

$$W_t = \triangle^{-q}\xi_t^{\#}, \quad t = 0, \pm 1, \ldots. \tag{5}$$

This Type I $I(q+d)$ process has been used by Tsay (2000, 2007) and Velasco and Robinson (2000), among others. See Robinson (2005) about the comparison between Type I and Type II nonstationary processes.

Given the preceding $I(d)$ and its associated Type I nonstationary $I(q+d)$ $(q \geq 1)$ process, we generalize the model in (1) as:

$$\begin{cases} y_{i,t} = f_i + \mathbf{x}_{i,t}^{\top}\beta + u_{i,t}, \;\; x_{i,t}^h = I(d_{x,i,h}), \;\; u_{i,t} = I(d_{u,i}), \;\; \max\limits_{1 \leq h \leq K} d_{x,i,h} = d_{x,i}, \\ d_{x,i} > d_{u,i}, \quad \text{if } d_{x,i} \geq 1/2, \text{ for any } i = 1, 2, \ldots, N, \end{cases} \tag{6}$$

where f_i is the fixed-effect for unit i, and $x_{i,t}^h$ is the h-th element of $\mathbf{x}_{i,t}$. The first notable feature of the model is that it does not assume $\mathbf{x}_{i,t}$ to be non-stochastic as commonly adopted in the previous research done for the TSCS data, because it is natural and more reasonable to assume that the regressors are dependent observations when the errors are modeled to be serially correlated. For example, the regressors may contain the political popularity of the government which is frequently modeled as a random variable, including the research in Box-Steffensmeier and Smith (1996, 1998) and Byers et al. (2000). In so doing, the number of potentially different $I(d)$ processes in (6) can be as large as $N \times (K+1)$, and this model can be viewed as the direct extension of the model in (4) of Tsay (2007) to the TSCS context. See Tsay (2007) about detailed discussions on the restrictions imposed in (6) when $N = 1$.

We note here that no restriction is imposed on the relative magnitude of $d_{x,i}$ and $d_{u,i}$ in (6), if each component of $\mathbf{x}_{i,t}$ and $u_{i,t}$ is stationary for all $i = 1, 2, \ldots, N$. When the maximum value of the differencing parameter of the regressors $\mathbf{x}_{i,t}$, i.e., $d_{x,i}$, is greater than or equal to 1/2, then we only require that $d_{x,i} > d_{u,i}$ so as to ensure that the probability order of magnitude of $\mathbf{x}_{i,t}$ dominates that of $u_{i,t}$ for each $i = 1, 2, \ldots, N$. This condition is routinely imposed in the regression models involving nonstationary processes, indicating that the long memory TSCS data subsumes the standard cointegration model as one of its special cases.

Second, in sharp contrast with the typical TSCS model where $u_{i,t}$ is assumed to be the simple AR(1) process or its variants, $\mathbf{x}_{i,t}$ and $u_{i,t}$ in (6) belong to the more general $I(d)$ processes.

Third, the idiosyncratic errors $u_{i,t}$ are allowed to be serially and contemporaneously correlated. The same conditions also apply to the regressors in each cross unit. As a result, the applicability of the model in (6) is much broader than that considered in Zellner (1962), Parks (1967), Kmenta and Gilbert (1970), Guilkey and Schmidt (1973), and Beck and Katz (1995).

Despite the general framework covered by the long memory TSCS model in (6), we show that this model can be easily estimated and tested with the MDSE provided that we can employ the long-run variance estimator of Robinson (1998) in controlling the effects of nuisance parameters in the DGP on the inference problems associated with the parameter β. We thus require $\mathbf{x}_{i,t}$ and $u_{j,t}$ in (6) to satisfy the conditions in Assumption A of Robinson (1998) for all $i, j = 1, \ldots, N$. The most stringent condition imposed in Assumption A of Robinson (1998) is that $\mathbf{x}_{i,t}$ and $u_{j,t}$ must be uncorrelated at all leads and lags for all $i, j = 1, \ldots, N$. However, this condition is still much weaker than the one where $\mathbf{x}_{i,t}$ is fixed or strongly exogenous as frequently encountered in the previous TSCS literature.

One is aware that, when the regressors and errors are nonstationary, the OLS estimator can be consistent even though $\mathbf{x}_{i,t}$ and $u_{j,t}$ are correlated, but the MD estimator may result in an inconsistent estimate under this circumstance. Thus, similar to the standing point taken in Tsay (2007, p. 829), the objective of this paper is not to present a method which can "efficiently" estimate the long memory TSCS model in (6), but it proposes a unified testing methodology that can conveniently deal with the inference problems of the long memory TSCS model when the regularity conditions in the following Theorem 1 are fulfilled. This strategy also mimics the idea of Beck and Katz (1995, p. 634) who offer the PCSE estimator that is both "easier to implement and produces accurate standard errors".

The MDSE method consists of two steps. First, we multiple-difference the data, $y_{i,t}$ and $\mathbf{x}_{i,t}$, before we run the subsequent pooled OLS estimation for β, denoted as $\widehat{\beta}_{\mathrm{MD}}$, or MD estimator. In other words, we employ the multiple-differenced (M-th differenced as compared to the usual first-differenced) transformation:

$$\triangle^M y_{i,t} = \triangle^M \mathbf{x}_{i,t}^{\top}\beta + \triangle^M u_{i,t}, \quad t = M+1, M+2, \ldots, T, \tag{7}$$

where M must be a positive integer in order to eliminate the impacts of fixed effect f_i in (6) on the estimation of β. The fractional differencing parameter of $\triangle^M \mathbf{x}_{i,t}^h$ is $d_{x,i,h} - M$, while that of $\triangle^M u_{i,t}$ is $d_{u,i} - M$ under this circumstance. We then employ these M-th differenced dependent variables and M-th differenced regressors to estimate β with the following MD estimator:

$$\begin{aligned} \widehat{\beta}_{\mathrm{MD}} &= \left[\sum_{i=1}^{N} \sum_{t=M+1}^{T} (\triangle^M \mathbf{x}_{i,t} - \overline{\triangle^M \mathbf{X}})\, \triangle^M \mathbf{x}_{i,t}^{\top}\right]^{-1} \\ &\quad \sum_{i=1}^{N} \sum_{t=M+1}^{T} (\triangle^M \mathbf{x}_{i,t} - \overline{\triangle^M \mathbf{X}})\, \triangle^M y_{i,t}, \end{aligned} \tag{8}$$

where $\overline{S}$ denotes the sample average of the random vector $S_{i,t}$ across i and t as observed in the data. The associated residuals from the MD estimator are:

$$\begin{aligned} e_{i,t} &= (\triangle^M y_{i,t} - \overline{\triangle^M \mathbf{Y}}) - (\triangle^M \mathbf{x}_{i,t} - \overline{\triangle^M \mathbf{X}})^\top \widehat{\beta}_{\mathrm{MD}}, \\ & i = 1, \ldots, N, \;\; t = M+1, \ldots, T. \end{aligned} \tag{9}$$

Note that when $M = 0$, $\widehat{\beta}_{\mathrm{MD}}$ reduces to be the usual OLS estimator, but the choice of $M = 0$ will not be used in this paper, because of the presence of fixed effect f_i in (6). This also implies that the PCSE estimator of Beck and Katz (1995) should not be employed if there is a fixed effect inherent in the TSCS data.

We now present the criteria for choosing M in (7):

$$M = \begin{cases} g+1, & \text{if } d_{x,i,h} = d_{u,i} = 0, \;\; \text{for } i = 1, 2, \ldots, N, \; h = 1, 2, \ldots, K; \\ g+1, & \text{if } d_{x,i,h} < 0, \; d_{u,i} < 0, \;\; \text{for } i = 1, 2, \ldots, N, \; h = 1, 2, \ldots, K; \\ g+1+\max\left([d_{x,i}], [d_{u,i}]\right), & \text{otherwise}, \end{cases} \tag{10}$$

where g is a non-negative integer, and $[G]$ represents the integer part of G. The rule of choosing M is indeed identical to that outlined in Tsay (2007). The only difference is that $M \geq 1$ is always required in (10), because of the presence of fixed-effect f_i in (6). As will be argued more clearly in the Mathematical Appendix concerning the theoretical foundation of the choice of M, the message within (10) is that we simply pick a large enough M to ensure that the differencing parameters of the transformed regressors and those of the transformed error terms are all negative if we can only vaguely guess the range of the $N \times (K+1)$ differencing parameters. For example, if we "know" the regressors and the errors are $I(0)$ processes, then $M = 1$ can be used. When there is only one regressor but $0 < d_{x,i}, d_{u,i} < 1$ for all $i = 1, 2, \ldots, N$, then $M = 1$ can be chosen as well. Please note that we still can use $M = 2$ in this set-up and not necessarily increase a higher variance of the MD estimator as shown in Tsay (2007, p. 835 and p. 836). Another example is that, if we suspect some regressors are $I(1)$ processes, but the unit root or stationarity tests cannot tell us whether the differencing parameter of the data is 0 or 1, then we simply adopt $M = 2$ to implement our difference-based method if we are sure that $\max\limits_{i,j \in \{1,\ldots,N\}} ([d_{x,i}], [d_{u,j}])$ is less than 2. It is clear that the great flexibility enjoyed with our method in choosing M is invaluable in empirical applications.

Calculating the variance matrix estimator of $\widehat{\beta}_{\text{MD}}$, $\widehat{D}_{\text{MD}}$, serves as the second step of the MDSE method, i.e.:

$$\widehat{D}_{\text{MD}} = \widehat{M}_{\text{MD}}^{-1} \widehat{V}_{\text{MD}} \widehat{M}_{\text{MD}}^{-1}, \tag{11}$$

where

$$\widehat{M}_{\text{MD}} = \frac{1}{T-M} \sum_{i=1}^{N} \sum_{t=M+1}^{T} (\triangle^M \mathbf{x}_{i,t} - \overline{\triangle^M \mathbf{X}})\, \triangle^M \mathbf{x}_{i,t}^{\top}, \tag{12}$$

$$\widehat{V}_{\text{MD}} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{q=-(T-M)+1}^{(T-M)-1} \left(\widehat{c}_{i,j,q} \times \widehat{d}_{i,j,q} \right), \tag{13}$$

and

$$\begin{cases} \widehat{c}_{i,j,q} = (T-M)^{-1} \sum\limits_{M+1 \le t, t+q \le T} (\triangle^M \mathbf{x}_{i,t} - \overline{\triangle^M \mathbf{X}})(\triangle^M \mathbf{x}_{j,t+q} - \overline{\triangle^M \mathbf{X}})^{\top}, \\ \widehat{d}_{i,j,q} = (T-M)^{-1} \sum\limits_{M+1 \le t, t+q \le T} e_{i,t}\, e_{j,t+q}. \end{cases} \tag{14}$$

It is evident in (11), (12), (13), and (14) that the computation of $\widehat{D}_{\text{MD}}$ is extremely straightforward, because it is based on Robinson's (1998) elegant long-run variance estimator which completely avoids any choice of an autoregressive lag length, a kernel function, or a bandwidth parameter. Moreover, these formulae clearly show that we do not "estimate" the AR, MA, and fractional differencing parameters, but instead we employ the nonparametric estimator of Robinson (1998) to control the impacts of AR, MA, and fractional differencing parameters on the testing performance concerning the regression coefficients β. In the Mathematical Appendix we offer a more detailed explanation about the rational behind the computation of $\widehat{D}_{\text{MD}}$.

As will be shown in the following Theorem 1 and the simulations in Section 3, the use of $\widehat{V}_{\text{MD}}$ can effectively control the impacts of the nuisance parameters in the DGP of $\triangle^M \mathbf{x}_{i,t}$ and that of $\triangle^M u_{j,t}$ on the inference performance concerning β. The proof of Theorem 1 is omitted in that the results in Theorem 1 can be derived from extending the findings in Theorem 1 of Tsay (2007) or those of Theorem 5 of Robinson (1998).

Theorem 1. *Under M being defined in (10) and for every pair of $i, j \in \{1, 2, \ldots N\}$, $\triangle^M \mathbf{x}_{i,t}$ and $\triangle^M u_{j,t}$ satisfy the conditions in Assumption A of Robinson (1998), and the spectral density function of the product $\triangle^M \mathbf{x}_{i,t} \triangle^M u_{j,t}$ is finite and positive definite; as $T \to \infty$, the MD estimator for the model in (6) is asymptotically normally distributed and*

$$(T - M)^{1/2} \widehat{D}_{\mathrm{MD}}^{-1/2} \left(\widehat{\beta}_{\mathrm{MD}} - \beta \right) \xrightarrow{d} N(0, I_K),$$

where $\widehat{D}_{\mathrm{MD}}$ is defined in (11), $\xrightarrow{d}$ stands for convergence in distribution, and I_k is a $(K \times K)$ identity matrix.

Theorem 1 reveals that $\widehat{\beta}_{\mathrm{MD}}$ converges at the rate of $\sqrt{T}$, demonstrating the temporally-dominated nature of the long memory TSCS data. We thus predict that the precision of the MD estimator increases with the increasing value of T, other things being equal. Theorem 1 also indicates that the MD-based methodology provides a unified framework to handle the complicated combination of different orders of integration in the regressors and errors within and across unit i. Under the conditions in Theorem 1, each element of β can be tested with the usual t ratio statistic even though the DGP belong to the stationary or nonstationary long memory processes.

Denoting $\beta_{0,h}$ as the h-th element of a column vector β_0, the null hypothesis for each element of β can be tested as:

$$H_0 : \beta_h = \beta_{0,h}, \qquad h = 1, 2, \ldots, K. \tag{15}$$

These K individual t-ratio statistics are jointly named as the t_{MD} test, and the t-ratio statistic for the h-th regressor $x_{i,t}^h$ is:

$$t_{\mathrm{MD},h} \equiv \frac{\sqrt{T - M} \left(\widehat{\beta}_{\mathrm{MD},h} - \beta_{0,h} \right)}{\sqrt{\widehat{D}_{\mathrm{MD}}^{h,h}}}, \qquad \text{for} \qquad h = 1, 2, \ldots, K, \tag{16}$$

where $\widehat{\beta}_{\mathrm{MD},h}$ is the h-th element of the $\widehat{\beta}_{\mathrm{MD}}$ estimator, and $\widehat{D}_{\mathrm{MD}}^{h,h}$ denotes the row h, column h element of $\widehat{D}_{\mathrm{MD}}$ in Theorem 1. When the absolute value of $t_{\mathrm{MD},h}$ in (16) is greater or equal to 1.96, the null hypothesis $\beta_h = \beta_{0,h}$ is rejected at the 5% level of significance. Certainly, a Wald-type statistic can be used to test the value of β jointly.

Theorem 1 also shows that the convergence rate of $\widehat{\beta}_{\mathrm{MD}}$ does not depend on N, which is the number of cross units. This implies that pooling more cross units does not help achieve a more efficient regression coefficient estimator, and this prediction is clearly shown in the simulation results presented in the following Table 1. Theorem 1 also presents that the complicated structure induced by more cross units can be easily controlled with the variance estimator $\widehat{D}_{\mathrm{MD}}$ built on Robinson's (1998) estimator.

3 Monte Carlo experiment

In this section we assess the finite sample performance of the MDSE method for the long memory TSCS model. Following Beck and Katz (1995), we choose $K = 1$. In the context of a stochastic regressor framework, we generate 1,000 replications of $x_{i,t}$ and $u_{i,t}$ based on the following model:

$$y^l_{i,t} = \beta_0 + \beta_1 x^l_{i,t} + u^l_{i,t}, \quad i = 1, 2, \ldots, N, \;\; t = 1, 2, \ldots, T, \;\; l = 1, 2, \ldots, 1000, \qquad (17)$$

where l denotes the l-th replication of the data, β_0 is fixed at 1, and β_1 can be $1, 0.9$, or 0.8 for investigating the empirical powers of the MDSE estimator given that the value of $\beta_{0,h}$ in (15) is always set to be 1.

We first consider the cases where $x_{i,t}$ and $u_{i,t}$ are generated as the ARFIMA(0,d,0) processes. To demonstrate the coverage of Theorem 1, $x_{i,t}$ can be stationary or non-stationary, while $u_{i,t}$ are always set to be stationary, i.e., the differencing parameters of $u_{i,t}$ are less than 1/2 for all configurations considered in this section.

When $x_{i,t}$ and $u_{i,t}$ are stationary, they are generated as:

$$x_{i,t} = \triangle^{-d_{x,i}} v_{i,t}, \quad u_{i,t} = \triangle^{-d_{u,i}} w_{i,t}, \qquad (18)$$

where $d_{x,i} \in (0, 1/2)$, $d_{u,i} \in (0, 1/2)$, and $v_{i,t}$ and $w_{i,t}$ all are zero-mean normal and independently identically distributed (i.i.d.) processes with:

$$\mathrm{E}(v^2_{i,t}) = \sigma^2_{v,i}, \quad \mathrm{E}(w^2_{j,t}) = \sigma^2_{w,j}, \quad \mathrm{E}(v_{i,t} w_{j,s}) = 0, \quad \text{for all } \; i, j, t, \; \text{ and } \; s. \qquad (19)$$

To characterize the heterogeneity across the regressors, variances of the first half of the units are set to be 1, i.e., $\sigma^2_{v,i} = 1$ for the first half of the units, while the variances of the second half of the units are set to be 1/4. Moreover, $v_{i,t}$ are generated so that all pairs of units are equally cross-correlated with a value of 1/2. Likewise, $w_{i,t}$ is generated with the same cross-correlation characteristics.

When $x_{i,t}$ is modeled as a nonstationary $I(1+d)$ process, we just partial sum a corresponding stationary $I(d)$ process based on the definition of the Type I nonstationary process. Since the maximum value of $d_{x,i}$ considered in the Monte Carlo experiment is 1.4, $M = 2$ is chosen for the entire simulation studies, even though $M = 1$ is large enough to handle the cases where the regressor and error term are stationary. Thus, the flexibility of choosing M is clearly demonstrated with our experiment.

Table 1 contains the finite sample performance of the $\widehat{\beta}_{\mathrm{MD}}$ estimator when $\beta_1 = 1$ is used. For all specifications considered in Table 1, the simulations reveal that the pattern in the root-mean-squared error (RMSE) is consistent with the prediction made in Theorem 1, i.e., the precision of the MD estimator increases with the increasing value of T, other things being equal. In addition, Table 1 reveals that pooling more cross units does not necessarily achieve a lower value of RMSE. This finding is consistent with the analytical results presented in Theorem 1 in that the convergence rate of $\widehat{\beta}_{\mathrm{MD}}$ does not depend on N, which is the number of finite cross units.

Tables 2 and 3 illustrate the sampling properties of the t_{MD} test at the 5% level of significance. To better understand the finite sample performance of the t_{MD} test, we follow Davidson and MacKinnon (1985) and Tsay (2007) to assess the asymptotic local power of the t_{MD} test when the DGP is a sequence of the form:

$$\triangle^M y_{i,t} = \triangle^M \mathbf{x}_{i,t}^\top \beta_0 + \triangle^M \mathbf{x}_{i,t}^\top (\delta_0/\sqrt{T-M}) + \triangle^M u_{i,t}, \quad t = M+1, M+2, \ldots, T, \tag{20}$$

where δ_0 is a $(K \times 1)$ finite constant. Given that the true value of β_h is generated as $\beta_{0,h} + (T-M)^{-1/2}\delta_{0,h}$, then asymptotically:

$$t_{\mathrm{MD},h} \xrightarrow{d} N(D_{h,h}^{-1/2}\delta_{0,h}, 1), \tag{21}$$

where $D_{h,h}$ is the row h, column h element of the asymptotic value of $\widehat{D}_{\mathrm{MD}}$ in Theorem 1, and $\delta_{0,h}$ is the h-th element of δ_0. This implies that the power of the $t_{\mathrm{MD},h}$ test increases with the increasing value of $|\delta_{0,h}|$ for a fixed sample size T and a fixed $D_{h,h}$. When $\delta_{0,h} = 0$, the power of the $t_{\text{MD},h}$ test is equivalent to its own size.

Since $\beta_{0,h}$ in (15) is always set to be 1 in the experimental study, the simulation results correspond exactly to the size of the t_{MD} test when the data are generated with $\beta_1 = 1$. Table 2 and Table 3 show that the size control of the t_{MD} test is quite well even though the time span is only 20, because the worst distortion found within these tables is less than 7% (under the configuration $d_{x,i} = 0.2$, $d_{u,i} = 0.4$ in Table 2 with $N = 10$, $T = 20$). The size performance improves as T increases, clearly supporting the findings in Theorem 1 in that $\widehat{\beta}_{\text{MD}}$ is asymptotically normally distributed and converges at a rate of $\sqrt{T}$ under the conditions in Theorem 1.

We now consider the power performance of the $t_{\text{MD},h}$ test under two alternatives, i.e., $\beta_1 = 0.8$ and 0.9. According to the local power analysis of the $t_{\text{MD},h}$ test discussed above, we note that $(T - M)^{-1/2}\delta_{0,1} = -0.1$ when $\beta_1 = 0.9$, and $(T - M)^{-1/2}\delta_{0,1} = -0.2$ when $\beta_1 = 0.8$, Thus, for a fixed alternative ($\beta_1 = 0.9$, or 0.8), *ceteris paribus*, the value of $|\delta_{0,1}|$ increases with an increasing value of T, indicating that the power of the $t_{\text{MD},h}$ test improves as T becomes larger. This is basically what we observe in these tables. Indeed, this prediction is completely supported in Tables 2 and 3 for all N considered in the experiment when $\beta_1 = 0.8$ and $T \geq 20$.

To check the robustness of the preceding findings, we replicate the above simulations by generating $x_{i,t}$ and $u_{i,t}$ as the following ARFIMA(1,d,1) processes:

$$(1 - \phi_{x,i}L) \triangle^{d_{x,i}} x_{i,t} = (1 + \theta_{x,i}L)v_{i,t}, \quad (1 - \phi_{u,i}L) \triangle^{d_{u,i}} u_{i,t} = (1 + \theta_{u,i}L)w_{i,t}, \quad (22)$$

where

$$\phi_{x,i} = \theta_{x,i} = \phi_{u,i} = \theta_{u,i} = \frac{1}{2}, \quad i = 1, \ldots, N. \quad (23)$$

Tables 4 and 5 display the sampling properties of the t_{MD} test at the 5% level of significance with the aforementioned more general ARFIMA$(1, d, 1)$ processes. These tables show that the size control of the t_{MD} test is much better than those from the ARFIMA$(0, d, 0)$ processes, although the time span is only 20, because the worst distortion found within these tables is only 3% (under the configuration $d_{x,i} = 0.2$,

$d_{u,i} = 0.4$ in Table 4 with $N = 40$, $T = 20$). The size performance also improves as T increases and clearly confirms the findings in Theorem 1. We also find that the power of the $t_{\mathrm{MD},h}$ test improves as T becomes larger in Tables 4 and 5 for all N considered in the experiment when $\beta_1 = 0.8$ and $T \geq 20$. As a consequence, the superior ability of the MDSE estimator in controlling the impacts of nuisance parameters within the error terms and regressors on the testing performance of the t_{MD} test is clearly supported in Tables 2-5.

4 Empirical application

The methodology developed in this paper is motivated by combining the long memory and TSCS literature into a unified framework, because the applicability of the long memory TSCS model is enormous in social science. Due to the increasing importance of the $I(d)$ process in political science, we apply the MD-based approach to a real dataset from the political time series in this section.

We re-examine the data of Hicks and Swank (1992) about welfare spending in 18 advanced industrial countries for the period 1960-82, as Beck and Katz (1995) do. The dependent variable used in Hicks and Swank (1992) is welfare spending as a proportion of gross domestic product. Hicks and Swank (1992) use a variety of political, institutional, and economic independent variables, including electoral turnout, the natural log of gross domestic product, and the price level. The general findings from Hicks and Swank (1992) based on Parks' (1967) method are that most of the t-ratios are above 4 as has been pointed out in Beck and Katz (1995). When applying the PCSE estimator to re-examine the data of Hicks and Swank (1995), Beck and Katz (1995) find that the "evidence for positive voter turnout effects is pervasive and robust" in Hicks and Swank (1992) changes as being marginally insignificant.

This paper focuses on four independent variables in the analysis of Hicks and Swank (1992), i.e., the electoral turnout, the natural log of gross domestic product, the price level, and the proportion of the population that is elderly. These variables are chosen, because they are important in explaining the welfare spending study of Hicks and Swank (1992). The second reason is that their movements are more like smooth time series data as required in the conditions imposed in Theorem 1. Another reason is that we suspect some of these variables might be stationary or nonstationary $I(d)$ processes. For example, the price level is usually taken as a nonstationary process and has been tested to be an $I(d)$ process in Baum et al. (1999). This implies that the estimation results in both Hicks and Swank (1992) and Beck and Katz (1995) could be problematic if some regressors are really stochastic $I(d)$ processes. Importantly, with a short time span of 23, there is no way to test the exact order of integration of time series processes with confidence. This also explains why we propose the MDSE method for the inference of long memory TSCS models without exactly knowing whether the regressors or errors in each cross unit are stationary or not. We also document that what has been done here is not to replicate the work of Hicks and Swank (1992) or that of Beck and Katz (1995), but rather to signify that the potential presence of stationary or nonstationary $I(d)$ processes in the TSCS model is influential in affecting the testing results concerning welfare spending and other similar variables.

Following Hicks and Swank (1992), we control for a possible post-1973 shift in welfare effort. Our procedure is to regress the dependent variable and the four independent variables on a constant and a post-1973 dummy, respectively. The residuals from these five regressions are collected to implement the MDSE estimator. Note that, when $M = 0$, the post-1973 dummy equals 1 for 1973 and later, otherwise 0. To adapt to the use of first-differencing when $M = 1$, the post-1973 dummy equals 1 for 1973, otherwise 0.

Table 6 contains the estimation and testing results from two different values of M, i.e., $M = 0$, and 1. As shown previously, the presence of fixed effects prevents us from using $M = 0$ in implementing the MD estimator. Moreover, we highly suspect that the price level data are nonstationary processes. Thus, $M = 0$ should not be used in carrying out the MD estimator under this circumstance. The appearance of $M = 0$ in Table 6 is purely used as the benchmark to demonstrate the effects of ignoring the possible presence of stochastic $I(d)$ regressors on the testing results. We find that all these four independent variables in Table 6 are highly significant in explaining the movement of welfare spending when $M = 0$ is used.

The MD estimate for the variable *Turnout* is 4.1122 when $M = 1$. This estimate is almost identical to the OLS-AR1 estimate of 4.1 in Table 6 of Beck and Katz (1995). However, based on the MDSE estimator, the variable Turnout is significant at the 5% level in a one-tailed test when $M = 1$. Therefore, the results from using the MDSE estimator with $M = 1$ actually reveal a weak but significant positive effect of voter turnout on welfare spending. This finding does not strongly support the pervasive and robust positive voter turnout effects observed in Hicks and Swank (1992). It is also different from the observation in Beck and Katz (1995) who document that the evidence for positive voter turnout effects changes as being insignificant when using their PCSE method .

For the testing results concerning the remaining three variables, and not the variables of interest to Hicks and Swank (1992) and Beck and Katz (1995), we find that the absolute value of the t-ratios associated with these three variables greatly decreases as compared to those obtained with $M = 0$. This implies that these three variables might also be highly persistent time series or even nonstationary processes. Again, we cannot be sure about their orders of integration with the short time span available at hand, but we still can test the value of these regression coefficients with the MDSE method developed in this paper.

5 Conclusions

A general class of long memory TSCS models is suggested to combine long memory and TSCS models into one unified framework. Although the coverage of this class of long memory TSCS models is broad, we show that they still can be easily estimated with a unified procedure proposed in this paper. This approach is built on the MD-based method of Tsay (2007) in order to resolve the spurious effect induced by the long memory processes, and it eliminates the need to simulate different sets of critical values for the complicated combinations of different $I(d)$ regressors and errors within and across cross units, and at various sample sizes. This feature is especially valuable in empirical applications, because the fractional differencing parameters of the regressors and those of the error terms cannot be estimated without bias when the sample size is relatively small as found in the typical TSCS data.

The computational cost of the proposed procedure is extremely mild and can be conducted with standard statistics packages, because it is built on the easy-to-implement but powerful long-run variance estimator of Robinson (1998). Furthermore, there is no restriction imposed on the relative magnitude between N and T provided that the MD estimator for the TSCS model can be implemented.

The simulations conducted in this paper reveal that the size control of our method is very promising under various combinations of cross-section, time dimensions, and stationary and nonstationary long memory processes, although T is only 20. The pattern of the finite sample power performance of our method is highly consistent with the theoretical findings generated from an asymptotic local power analysis. Therefore, the proposed method has great potential to successfully deal with a general class of long memory TSCS data which have never been explicitly considered in the literature.

We also apply our method to re-examine the data of Hicks and Swank (1992) about the determinants of welfare spending, because some of the data used in Hicks and Swank (1992) might be $I(d)$ processes. The testing results from the MDSE method with $M = 1$ reveal a weak but significant positive effect of voter turnout on welfare spending. This finding is in the middle of the results in Hicks and Swank (1992) and those in Beck and Katz (1995). It also signifies that the potential presence of stationary or nonstationary $I(d)$ processes in the TSCS model is influential in affecting the testing results concerning welfare spending.

Mathematical Appendix

In this appendix we consider the theoretical foundation of choosing M for implementing the MD estimator and the associated modified long-run variance estimator of Robinson (1998).

Before discussing the method of choosing M in (10) for implementing the MD estimator, let us remark here that the choice of M must be big enough to ensure that the resulting transformed regressors and transformed error terms are covariance stationary in order to represent them with the following infinite-order MA processes:

$$\triangle^M \mathbf{x}_{i,t} = \sum_{n=0}^{\infty} \varphi_{i,n} \mathbf{a}_{i,t-n}, \qquad \triangle^M u_{j,t} = \sum_{n=0}^{\infty} \psi_{j,n} b_{j,t-n}, \qquad t = M+1, \ldots, T,$$

where $\mathbf{a}_{i,t}$ and $b_{j,t}$ satisfy the conditions in Assumption A of Robinson (1998) as required previously. Note that each element of $\mathbf{a}_{i,t}$ is allowed to be correlated with each other so as to capture the major feature of TSCS data. Moreover, $\mathbf{a}_{i,t}$ or $b_{i,t}$ can be also contemporaneously correlated across i, respectively, for the similar argument, as long as $\mathbf{x}_{i,t}$ and $u_{j,t}$ are uncorrelated at all leads and lags for all $i, j = 1, \ldots, N$ as discussed previously.

Given the above-mentioned restrictions imposed on $\triangle^M \mathbf{x}_{i,t}$ and $\triangle^M u_{j,t}$, Robinson (1998) shows that the long-run variance of $\triangle^M \mathbf{x}_{i,t} \, \triangle^M u_{j,t}$ can be consistently estimated if the fractional differencing parameter of $\triangle^M \mathbf{x}_{i,t}$ and that of $\triangle^M u_{j,t}$ are all greater than or equal to zero and $0 \leq (d_{x,i,h} - M) + (d_{u,j} - M) < 1/2$ for all $i, j = 1, \ldots, N$ and $h = 1, \ldots, K$. Theoretically, the finding concerning Robinson's (1998) long-run variance estimator is of great importance to the inference of β based on the MD estimator defined in (8). One drawback, empirically, is that the possibility that the conditions $0 \leq (d_{x,i,h} - M) + (d_{u,j} - M) < 1/2$ are satisfied for all $i, j = 1, \ldots, N$ and $h = 1, \ldots, K$ could be very low, especially when the values of N and K are relatively large. If we further consider the difficulties that the integration orders of $\triangle^M \mathbf{x}_{i,t}$ and that of $\triangle^M u_{j,t}$ are simply unknown in reality, and the differencing parameters of the transformed regressors $\triangle^M \mathbf{x}_{i,t}$ and that of transformed error term $\triangle^M u_{j,t}$ just cannot be estimated without bias when the sample size is relatively moderate, then we quickly realize that it may be risky to apply Robinson's (1998) estimator to the long memory TSCS data without modifications, not to mention that the total differencing parameters to be estimated are $N \times (K+1)$.

The major advantages of using the choice of M in (10) is that: Even though we are not sure about the range and the combinations of the $N \times (K+1)$ differencing parameters inherent in (6), we still can apply the methodology developed in Robinson (1998) to test the value of β in the long memory TSCS model, provided that we are willing to pick a larger value of M to ensure that the differencing parameters of the transformed regressors and those of the transformed error terms are all negative. There are two reasons behind this choice of M. First, under suitable regularity conditions, we prove that the asymptotic covariance matrix of $\widehat{\beta}_{\mathrm{MD}}$ can be consistently estimated with the help of Robinson's (1998) method under this circumstance. It follows that our choice of M is theoretically justified. Furthermore, if the differencing parameter of $\triangle^M x_{i,t}^h$ and that of $\triangle^M u_{j,t}$ have opposite signs, then the spectral density function of $\triangle^M x_{i,t}^h \triangle^M u_{j,t}$ may degenerate to be zero based on Lemma A.1 of Tsay (2000), implying that the $\sqrt{T}$ convergence rate of the MD estimator presented in Theorem 1 may not hold, and the theoretical foundation of using the critical values from the standard normal distribution for testing the value of β is problematic under this situation. This motivates us to select a larger value of M to ensure that the differencing parameters of $\triangle^M \mathbf{x}_{i,t}$ and that of $\triangle^M u_{j,t}$ are all negative in order to result in an asymtotically normally distributed MD estimator defined in (8).

The second reason for the choice of M is more empirically oriented, i.e., we may not be sure about the exact value of the fractional differencing parameter of the time series under consideration. However, it is much easier for us to pick up a large enough M to ensure that the differencing parameters of the transformed regressors and those of the transformed error terms are all negative with the help of the semiparametric estimators developed in Robinson (1995). These arguments together explain why we want to and can propose a unified testing procedure for the long memory TSCS model.

We now discuss the rational behind the computation of $\widehat{D}_{\mathrm{MD}}$ in (11). Under M being defined in (10), the model in (6) becomes the one in (7), because $M \geq 1$ is imposed in (10). For simplicity and without loss of generality, assuming that we know the mean of $\triangle^M \mathbf{x}_{i,t}^\top$ and that of $\triangle^M u_{i,t}$ are all zero, we can represent the associated pooled OLS estimator for β as:

$$\widehat{\beta}_{\mathrm{MD}} - \beta = \left[\sum_{i=1}^{N} \sum_{t=M+1}^{T} \triangle^M \mathbf{x}_{i,t} \, \triangle^M \mathbf{x}_{i,t}^\top\right]^{-1} \sum_{i=1}^{N} \sum_{t=M+1}^{T} \triangle^M \mathbf{x}_{i,t} \, \triangle^M u_{i,t}. \tag{8$'$}$$

Here, $\widehat{V}_{\mathrm{MD}}$ in (13) is employed to consistently estimate the N^2 values of the long-run variances of $\triangle^M \mathbf{x}_{i,t} \triangle^M u_{j,t}$ from the value of $\sum\limits_{i=1}^{N} \sum\limits_{t=M+1}^{T} \triangle^M \mathbf{x}_{i,t} \, \triangle^M u_{i,t}$ in (8$'$). The design of $\widehat{V}_{\mathrm{MD}}$ reflects the defining feature of Robinson's estimator and that of TSCS data. Since the condition that N is finite is always satisfied for the basic set-up of TSCS data, the pooling of the finite N cross units will not change the convergence rate of the MD estimator in (8). Moreover, $\widehat{V}_{\mathrm{MD}}$ in (13) reduces to be the corresponding counterpart considered in Tsay (2007) where $N = 1$.

REFERENCES

1. Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5-59.

2. Baum, C.F., Barkoulas, J.T., Caglayan, M., 1999. Persistence in international inflation rates. Southern Economic Journal 65, 900-913.

3. Beck, N., 2001. Time-series-cross-section data: What have we learned in the past few years? Annual Review of Political Science 4, 271-293.

4. Beck N., Katz, J.N., 1995. What to do (and not to do) with Time-Series Cross-Section Data. American Political Science Review 89, 634-647.

5. Box-Steffensmeier, J.M., Smith, R.M., 1996. The dynamics of aggregate partisanship. American Political Science Review 90, 567-580.

6. Box-Steffensmeier, J.M., Smith, R.M., 1998. Investigating political dynamics using fractional integration methods. American Journal of Political Science 42, 661-689.

7. Box-Steffensmeier, J.M., Tomlinson, A.R., 2000. Fractional integration methods in political science. Electoral Studies 19, 63-76.

8. Byers, D., Davidson, J., Peel, D.A., 2000. The dynamics of aggregate political popularity: Evidence from eight countries. Electoral Studies 19, 49-62.

9. Davidson, J.E.H, Peel, D.A., Byers, J.D., 2006. Support for governments and leaders: Fractional cointegration analysis of poll evidence from the UK, 1960-2004. Studies in Nonlinear Dynamics & Econometrics 10, No. 1, Article 3.

10. Davidson, R., MacKinnon, J.G., 1985. The interpretation of test statistics. Canadian Journal of Economics 18, 38-57.

11. Dolado, J.J., Gonzalo, J., Mayoral L., 2003. Long-range dependence in Spanish political opinion poll series. Journal of Applied Econometrics 18, 137-155.

12. Granger, C.W.J., 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227-238.

13. Granger, C.W.J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15-29.

14. Granger, C.W.J., Newbold, P., 1974. Spurious regression in econometrics. Journal of Econometrics 2, 111-120.

15. Guilkey, D.K., Schmidt, P., 1973. Estimation of seemingly unrelated regressions with vector autoregressive errors. Journal of the American Statistical Association 68, 642-647.

16. Hicks, A., Swank, D., 1992. Politics, institutions and welfare spending in industrialized democraticies, 1960-1982. American Political Science Review 86, 658-674.

17. Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165-176.

18. Kmenta, J., 1971. Elements of Econometrics. Macmillan, New York.

19. Kmenta, J., Gilbert, R.F., 1970. Estimation of seemingly unrelated regressions with autoregressive disturbances. Journal of the American Statistical Association 65, 186-197.

20. Lebo, M.J., Walker, R.W., Clarke, H.D., 2000. You must remember this: Dealing with long memory in political analyses. Electoral Studies 19, 31-48.

21. Parks, R., 1967. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. Journal of the American Statistical Association 62, 500-509.

22. Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. Journal of Econometrics 33, 311-340.

23. Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23, 1630-1661.

24. Robinson, P.M., 1998. Inference-without-smoothing in the presence of nonparametric autocorrelation. Econometrica 66, 1163-1182.

25. Robinson, P.M., 2005. The distance between rival nonstationary fractional processes. Journal of Econometrics 128, 283-300.

26. Tsay, W.J., 2000. Estimating trending variables in the presence of fractionally integrated errors. Econometric Theory 16, 324-346.

27. Tsay, W.J., 2007. Using difference-based methods for inference in regression with fractionally-integrated processes. Journal of Time Series Analysis 28, 827-843.

28. Tsay, W.J., Chung, C.F., 2000. The spurious regression of fractionally integrated processes. Journal of Econometrics 96, 155-182.

29. Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95, 1229-1243.

30. Zellner, A., 1962. An efficient method for estimating seemingly unrelated regressions and tests for aggregate bias. Journal of the American Statistical Association 57, 348-368.

Table 1. RMSE of the MD Estimator for the Long Memory TSCS Data

T	10	20	30	40	50	60
		$d_{x,i} = 0.4,\ \ d_{u,i} = 0.2$ for all $i = 1, 2, \ldots, N$				
$N = 10$	0.2565	0.1672	0.1381	0.1194	0.1044	0.0982
$N = 20$	0.2214	0.1527	0.1278	0.1099	0.0993	0.0884
$N = 30$	0.2241	0.1479	0.1282	0.1082	0.0971	0.0890
$N = 40$	0.2137	0.1420	0.1251	0.1040	0.0962	0.0894
		$d_{x,i} = 0.2,\ \ d_{u,i} = 0.4$ for all $i = 1, 2, \ldots, N$				
$N = 10$	0.2316	0.1521	0.1253	0.1081	0.0942	0.0888
$N = 20$	0.1988	0.1373	0.1160	0.0994	0.0898	0.0798
$N = 30$	0.2017	0.1340	0.1154	0.0974	0.0881	0.0804
$N = 40$	0.1926	0.1284	0.1135	0.0941	0.0869	0.0808
		$d_{x,i} = 1.4,\ \ d_{u,i} = 0.2$ for all $i = 1, 2, \ldots, N$				
$N = 10$	0.3882	0.2431	0.2029	0.1764	0.1558	0.1461
$N = 20$	0.3443	0.2318	0.1862	0.1626	0.1487	0.1320
$N = 30$	0.3425	0.2210	0.1944	0.1650	0.1425	0.1328
$N = 40$	0.3302	0.2098	0.1808	0.1535	0.1426	0.1325
		$d_{x,i} = 1.2,\ \ d_{u,i} = 0.4$ for all $i = 1, 2, \ldots, N$				
$N = 10$	0.3644	0.2303	0.1919	0.1669	0.1472	0.1381
$N = 20$	0.3214	0.2181	0.1765	0.1537	0.1401	0.1247
$N = 30$	0.3214	0.2080	0.1829	0.1550	0.1350	0.1253
$N = 40$	0.3079	0.1986	0.1719	0.1453	0.1348	0.1253

Notes: The results are all based on 1,000 replications with $M = 2$. The data are generated based on (17), (18), (19), and $\beta_0 = \beta_1 = 1$. RMSE denotes the root of mean-squared-errors.

Table 2. Empirical Power of the t_{MD} Test at the 5% Level of Significance with ARFIMA$(0, d, 0)$ Processes and Stationary Regressor

T	Alternatives	10	20	30	40	50	60
		$d_{x,i} = 0.4$, $d_{u,i} = 0.2$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	20.40	11.00	8.80	8.00	7.30	6.40
	$\beta_1 = 0.9$	23.40	18.30	17.80	19.50	19.40	22.10
	$\beta_1 = 0.8$	31.60	33.10	42.40	46.30	52.40	58.30
$N = 20$	$\beta_1 = 1.0$	19.60	10.00	9.80	8.00	7.20	7.20
	$\beta_1 = 0.9$	22.10	17.10	18.30	20.90	22.10	24.70
	$\beta_1 = 0.8$	32.80	37.60	44.60	52.80	56.60	62.40
$N = 30$	$\beta_1 = 1.0$	20.10	10.80	9.20	8.40	7.60	6.50
	$\beta_1 = 0.9$	22.00	18.10	17.80	22.60	22.00	26.40
	$\beta_1 = 0.8$	34.60	35.70	44.50	51.50	61.60	66.20
$N = 40$	$\beta_1 = 1.0$	19.80	11.60	9.40	8.20	7.10	8.20
	$\beta_1 = 0.9$	26.40	20.04	20.08	22.80	23.50	25.60
	$\beta_1 = 0.8$	36.10	37.90	43.90	53.40	60.10	70.20
		$d_{x,i} = 0.2$, $d_{u,i} = 0.4$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	21.10	11.90	8.70	8.50	7.20	6.80
	$\beta_1 = 0.9$	24.30	19.70	19.60	21.70	21.70	24.20
	$\beta_1 = 0.8$	33.50	36.90	46.70	52.60	60.50	66.70
$N = 20$	$\beta_1 = 1.0$	20.00	9.70	10.40	8.00	7.20	6.70
	$\beta_1 = 0.9$	24.00	18.50	21.50	23.40	25.70	28.40
	$\beta_1 = 0.8$	37.60	42.90	50.20	59.90	64.10	69.90
$N = 30$	$\beta_1 = 1.0$	19.80	10.80	8.90	8.50	7.90	7.00
	$\beta_1 = 0.9$	24.20	20.60	20.40	25.60	25.90	30.70
	$\beta_1 = 0.8$	39.20	41.10	52.60	59.10	67.90	75.70
$N = 40$	$\beta_1 = 1.0$	20.50	11.40	9.60	7.90	7.00	8.10
	$\beta_1 = 0.9$	27.50	23.80	22.80	25.60	26.80	28.90
	$\beta_1 = 0.8$	40.10	43.70	52.10	62.00	67.80	77.40

Notes: The results are all based on 1,000 replications with $M = 2$. The data are generated based on (17), (18), and (19). The null hypotheses in (15) are $\beta_{0,1} = 1$. The rejection percentages for the alternative, $\beta_1 = 1$, represent the empirical size of the t_{MD} tests.

Table 3. Empirical Power of the t_{MD} Test at the 5% Level of Significance with ARFIMA$(0, d, 0)$ Processes and Nonstationary Regressor

T	Alternatives	10	20	30	40	50	60
		$d_{x,i} = 1.4, \quad d_{u,i} = 0.2$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	16.80	8.20	8.10	7.00	7.00	6.70
	$\beta_1 = 0.9$	17.50	12.10	11.90	13.90	12.00	13.80
	$\beta_1 = 0.8$	20.60	20.70	24.60	26.40	29.40	31.90
$N = 20$	$\beta_1 = 1.0$	17.00	9.50	7.00	7.00	7.10	6.00
	$\beta_1 = 0.9$	16.20	11.30	12.10	14.10	12.90	15.60
	$\beta_1 = 0.8$	21.60	22.60	26.50	29.20	32.60	34.00
$N = 30$	$\beta_1 = 1.0$	16.70	8.30	8.70	7.00	6.10	6.40
	$\beta_1 = 0.9$	17.10	13.40	12.30	12.30	13.10	15.70
	$\beta_1 = 0.8$	20.30	21.60	25.00	29.50	33.80	37.10
$N = 40$	$\beta_1 = 1.0$	17.40	9.00	8.10	7.10	7.10	7.80
	$\beta_1 = 0.9$	19.80	15.90	12.90	13.30	14.10	14.90
	$\beta_1 = 0.8$	24.80	22.60	24.00	31.70	33.80	39.30
		$d_{x,i} = 1.2, \quad d_{u,i} = 0.4$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	18.20	9.50	7.80	7.10	6.80	7.00
	$\beta_1 = 0.9$	19.00	13.30	13.20	15.50	13.30	14.70
	$\beta_1 = 0.8$	22.40	22.40	26.30	29.20	31.30	35.60
$N = 20$	$\beta_1 = 1.0$	17.30	10.00	7.20	7.00	7.50	6.00
	$\beta_1 = 0.9$	17.20	12.10	12.50	14.90	14.10	16.50
	$\beta_1 = 0.8$	22.60	24.70	29.20	32.20	35.10	38.00
$N = 30$	$\beta_1 = 1.0$	17.70	8.80	8.70	7.20	6.20	6.10
	$\beta_1 = 0.9$	18.70	14.40	12.90	14.00	13.90	16.80
	$\beta_1 = 0.8$	21.90	23.30	27.10	32.40	37.20	39.70
$N = 40$	$\beta_1 = 1.0$	17.60	9.20	8.40	7.60	7.40	7.80
	$\beta_1 = 0.9$	21.10	16.20	13.60	15.10	14.50	15.90
	$\beta_1 = 0.8$	26.90	24.60	26.10	34.70	36.70	43.30

Notes: The results are all based on 1,000 replications with $M = 2$. The data are generated based on (17), (18), and (19). The null hypotheses in (15) are $\beta_{0,1} = 1$. The rejection percentages for the alternative, $\beta_1 = 1$, represent the empirical size of the t_{MD} tests.

Table 4. Empirical Power of the t_{MD} Test at the 5% Level of Significance with ARFIMA$(1, d, 1)$ Processes and Stationary Regressor

T	Alternatives	10	20	30	40	50	60
		$d_{x,i} = 0.4$, $d_{u,i} = 0.2$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	13.10	7.00	6.70	7.00	5.70	5.70
	$\beta_1 = 0.9$	14.70	12.10	13.50	15.60	15.10	17.20
	$\beta_1 = 0.8$	20.10	26.30	32.60	36.70	41.30	50.50
$N = 20$	$\beta_1 = 1.0$	13.40	7.30	5.40	6.20	6.80	6.70
	$\beta_1 = 0.9$	12.70	11.80	14.20	15.90	17.00	19.70
	$\beta_1 = 0.8$	21.90	29.50	35.00	42.10	46.70	53.80
$N = 30$	$\beta_1 = 1.0$	12.20	7.90	7.90	7.20	5.50	5.10
	$\beta_1 = 0.9$	14.70	13.50	14.50	17.30	18.70	21.10
	$\beta_1 = 0.8$	22.00	28.70	35.00	41.50	51.20	55.70
$N = 40$	$\beta_1 = 1.0$	13.20	6.90	6.10	6.80	7.30	6.80
	$\beta_1 = 0.9$	16.90	15.80	13.70	17.50	18.20	21.70
	$\beta_1 = 0.8$	25.70	29.40	34.20	47.00	49.80	59.70
		$d_{x,i} = 0.2$, $d_{u,i} = 0.4$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	13.50	7.20	7.10	7.00	5.80	5.40
	$\beta_1 = 0.9$	15.70	12.80	13.60	16.50	16.20	18.00
	$\beta_1 = 0.8$	22.00	28.70	33.90	38.70	43.50	52.80
$N = 20$	$\beta_1 = 1.0$	13.70	7.40	5.80	6.50	7.10	6.50
	$\beta_1 = 0.9$	13.70	12.20	14.20	16.90	17.70	19.80
	$\beta_1 = 0.8$	23.70	31.30	37.10	44.30	48.60	55.70
$N = 30$	$\beta_1 = 1.0$	12.40	7.90	7.30	7.00	6.10	5.30
	$\beta_1 = 0.9$	15.70	14.00	15.30	17.70	18.70	22.90
	$\beta_1 = 0.8$	23.10	31.20	36.40	44.00	53.20	57.60
$N = 40$	$\beta_1 = 1.0$	14.00	8.00	6.20	6.40	7.50	7.30
	$\beta_1 = 0.9$	18.20	17.40	15.00	19.10	19.40	22.10
	$\beta_1 = 0.8$	26.20	31.60	36.10	47.40	52.50	62.60

Notes: The results are all based on 1,000 replications with $M = 2$. The data are generated based on (17), (19), and (22). The null hypotheses in (15) are $\beta_{0,1} = 1$. The rejection percentages for the alternative, $\beta_1 = 1$, represent the empirical size of the t_{MD} tests.

Table 5. Empirical Power of the t_{MD} Test at the 5% Level of Significance with ARFIMA$(1, d, 1)$ Processes and Nonstationary Regressor

T	Alternatives	10	20	30	40	50	60
		$d_{x,i} = 1.4$, $d_{u,i} = 0.2$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	6.60	4.30	5.40	4.70	4.30	4.80
	$\beta_1 = 0.9$	8.60	10.70	15.50	15.70	19.90	22.40
	$\beta_1 = 0.8$	16.00	30.20	40.70	50.00	55.10	63.60
$N = 20$	$\beta_1 = 1.0$	7.50	4.30	4.20	4.90	4.70	4.70
	$\beta_1 = 0.9$	9.20	12.50	14.20	19.80	24.20	27.40
	$\beta_1 = 0.8$	19.20	32.90	44.20	53.60	60.80	70.90
$N = 30$	$\beta_1 = 1.0$	6.90	4.50	6.10	6.40	3.50	3.90
	$\beta_1 = 0.9$	11.10	14.80	15.20	21.00	23.70	27.40
	$\beta_1 = 0.8$	20.60	34.90	45.50	57.50	64.10	72.40
$N = 40$	$\beta_1 = 1.0$	7.50	4.30	3.40	4.00	4.60	5.70
	$\beta_1 = 0.9$	11.60	16.90	18.60	21.20	22.10	28.50
	$\beta_1 = 0.8$	23.00	35.30	43.80	59.70	65.30	74.60
		$d_{x,i} = 1.2$, $d_{u,i} = 0.4$ for all $i = 1, 2, \ldots, N$					
$N = 10$	$\beta_1 = 1.0$	7.70	5.20	5.50	5.40	4.30	4.80
	$\beta_1 = 0.9$	9.30	10.40	14.40	14.30	16.30	19.10
	$\beta_1 = 0.8$	15.60	27.80	35.00	42.00	47.00	55.20
$N = 20$	$\beta_1 = 1.0$	8.40	5.00	4.40	4.30	5.10	5.30
	$\beta_1 = 0.9$	9.90	12.30	13.50	17.20	21.60	23.10
	$\beta_1 = 0.8$	19.30	29.70	38.80	46.90	53.60	61.70
$N = 30$	$\beta_1 = 1.0$	8.50	5.80	7.60	6.90	4.50	4.90
	$\beta_1 = 0.9$	11.40	13.30	14.90	18.50	19.70	23.30
	$\beta_1 = 0.8$	19.70	29.70	38.40	48.90	57.20	63.00
$N = 40$	$\beta_1 = 1.0$	8.50	5.20	4.20	5.00	5.50	6.10
	$\beta_1 = 0.9$	12.60	15.00	16.00	18.60	18.80	23.70
	$\beta_1 = 0.8$	22.40	31.20	36.90	51.40	57.60	66.60

Notes: The results are all based on 1,000 replications with $M = 2$. The data are generated based on (17), (19), and (22). The null hypotheses in (15) are $\beta_{0,1} = 1$. The rejection percentages for the alternative, $\beta_1 = 1$, represent the empirical size of the t_{MD} tests.

Table 6. Estimates of the Hicks and Swank Data with the MDSE Method

	$M = 0$		$M = 1$	
Variables	Estimate	t-ratio	Estimate	t-ratio
Turnout	19.5944	12.3269***	4.1122	1.7104*
lnGDP	1.5000	2.4337**	0.3549	0.3146
Price level $\times 100$	0.0583	4.3347***	0.0545	2.7440***
Aged share of population	113.8592	10.4161***	82.8911	2.5354**

Notes: The results are based on the data of Hicks and Swank (1992) about welfare spending in 18 advanced industrial countries for the period 1960-82. Turnout is the electoral turnout, lnGDP is the natural log of gross domestic product, Price level is the consumer price index, and Aged share of population is the proportion of the population who are elderly. See Hicks and Swank (1992) about the detailed definition of these independent variables. ***, **, and * denote significance at the 1%, 5%, and 10% levels in a two-tailed test, respectively.