UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 577

System GMM estimation of panel data models with time varying slope coefficients

Yoshihiro Sato and Måns Söderbom

December 2013

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se   info@handels.gu.se

System GMM estimation of panel data models with time varying slope coefficients*

Yoshihiro Sato† and Måns Söderbom‡

December 10, 2013

Abstract

We highlight the fact that the Sargan-Hansen test for GMM estimators applied to panel data is a joint test of valid orthogonality conditions and coefficient stability over time. A possible reason why the null hypothesis of valid orthogonality conditions is rejected is therefore that the slope coefficients vary over time. One solution is to estimate an empirical model where the coefficients are time specific. We apply this solution to the system GMM estimatior of the Cobb-Douglas production functions for a selection of Swedish industries, and find that relaxing the assumption that slope coefficients are constant over time results in considerably more satisfactory outcomes of the Sargan-Hansen test.

1 Introduction

The system GMM estimator, proposed by Arellano and Bover (1995) and Blundell and Bond (1998), has become a popular method for estimating panel data models.[1] In this paper we study this estimator in a setting where slope coefficients are time varying. The conventional system GMM estimator is based on the assumption that the slope coefficients are constant over time, a restriction that typically results in a large number of overidentifying restrictions. The null

*We thank Steve Bond for very helpful comments on an earlier draft of the paper. Funding from the Wallander foundation is gratefully acknowledged. All errors are our own.

†Department of Economics, University of Gothenburg, Sweden; and The European Institute of Japanese Studies, Stockholm School of Economics, Sweden. E-mail: Yoshihiro.Sato@economics.gu.se.

‡Department of Economics, University of Gothenburg, Sweden. Address: Department of Economics, University of Gothenburg. P.O. Box 640, SE 405 30 Gothenburg, Sweden. Tel.: +46 (0)31 786 4332. Fax: +46 (0)31 7861326. E-mail: mans.soderbom@economics.gu.se.

[1] There is a large number of recent applications of system GMM estimators. According to ideas.repec.org, Arellano and Bover (1995) has been cited in more than 1403 papers. Blundell and Bond (1998) has been cited in 2125 papers.

hypothesis underlying the Sargan-Hansen test is that all overidentifying restrictions, including those resulting from assuming time constant coefficients, are valid. Hence, if the slope coefficients are in fact time varying, so that (some of) the overidentifying restrictions do not hold, the Sargan-Hansen test will tend to indicate that the model is mis-specified.

We show how the system GMM model can be estimated while allowing the coefficients to be time varying. With this more general formulation of the model, the resulting Sargan-Hansen test has a more clear-cut interpretation, shedding light on whether the lagged values of the regressors are valid instruments or not. Stability of the coefficients over time is easily tested using a Wald test. When estimating Cobb-Douglas production functions using Swedish firm-level panel data on manufacturing industries with time constant coefficients imposed, we obtain plenty of evidence from the Sargan-Hansen test that the overidentifying restrictions do not hold. For a more general model with time varying slope coefficients, in contrast, all specification tests are satisfactory.

The rest of this study is as follows. Section 2 discusses the interpretation of the Sargan-Hansen test in the context of system GMM estimation with time constant slope coefficients imposed. Section 3 considers a more general model with time varying coefficients, and applies it to our empirical data. Section 4 concludes the study.

2 The basic problem

Consider the following linear panel data model with time varying slope coefficients:

$$y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta}_t + \epsilon_{i,t} \quad \text{for } i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T, \tag{1}$$

where $y_{i,t}$ is a dependent variable, $\mathbf{x}_{i,t}$ is a column vector of K regressors, and $\epsilon_{i,t}$ is the residual for firm i at period t. $\boldsymbol{\beta}_t$ is a column vector of K coefficients. Note that the suffix on $\boldsymbol{\beta}$ indicates that the coefficients are time varying and we therefore refer to the model as a time varying coefficient (TVC) model.

The conventional setup of a panel data model for GMM estimation is such that there is a separate set of instruments for each period. Such a framework enables the researcher to exploit more instruments over time. For simplicity, we focus on the case where there is one instrument for each regressor, resulting in orthogonality conditions of the following form:

$$E\left[\begin{pmatrix} \mathbf{z}'_{i,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}'_{i,2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}'_{i,T} \end{pmatrix} \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \vdots \\ \epsilon_{i,T} \end{pmatrix}\right] = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \tag{2}$$

which we write in more compact form as

$$E\left(\mathbf{Z}_i^{*\prime}\boldsymbol{\epsilon}_i\right) = \mathbf{0}, \tag{3}$$

where $\mathbf{Z}_i^*$ is a $TK \times T$ matrix of the instruments and $\boldsymbol{\epsilon}_i$ is a column vector of the residuals. This TVC model is just identified, i.e., we have TK unknown coefficients and TK orthogonality condition. Since no overidentifying restrictions are imposed, the Sargan-Hansen value associated with the GMM estimator for this model is exactly zero.

We now consider a restricted model specification, in which the coefficients are constant over time:

$$y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} \quad \text{for } i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T. \tag{4}$$

We refer to it as a constant coefficient (CC) model. GMM estimation based on the same set of instruments as in the TVC results in exactly the same orthogonality condition as eq.(3). In contrast to the CC model, however, this system of equations is clearly overidentified: there are K unknown coefficients and TK orthogonality conditions.

The CC model is very common in the literature that utilizes the system GMM model. Under the null hypothesis that all overidentifying restrictions are valid, the Sargan-Hansen test statistic has an asymptotic χ^2 distribution with $(T-1)\,K$ degrees of freedom. Comparing the unrestricted (eq.(1)) and the restricted (eq.(4)) specifications, the Sargan-Hansen test thus has a clearcut interpretation as a test of coefficient stability over time, i.e., $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_T$. Few if any papers in the applied literature advance this interpretation of the Sargan-Hansen test, however.

3 An empirical illustration

In this section, we estimate simple non-dynamic Cobb-Douglas production functions with and without time constant slope coefficients imposed. We use (unbalanced) panel data on Swedish manufacturing firms covering six industries (Chemicals, Motor vehicles, Pulp and paper, Wood products, Publishing and printing, and Machinery) for the period 1997–2006.[2] Defining $\mathbf{x}_{i,t} \equiv [l_{i,t}\ k_{i,t}]'$ and $\boldsymbol{\beta} = [\beta_l\ \beta_k]'$, we specify the production function with time constant slope coefficients (i.e. the CC model) as

$$y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta} + \delta_t D_t + (\eta_i + \epsilon_{i,t}) \quad \text{for } t = 3, 4, \ldots, T, \tag{5}$$

where $y_{i,t}$ denotes log value-added, $l_{i,t}$ is log employment, $k_{i,t}$ is log physical capital, D_t is year dummies, δ_t is year effects, η_i is time constant unobserved firm effects, and $\epsilon_{i,t}$ is a time varying residual. The differenced production function is expressed as

$$\Delta y_{i,t} = \Delta\mathbf{x}'_{i,t} \cdot \boldsymbol{\beta} + \delta_t D_t - \delta_{t-1} D_{t-1} + \Delta\epsilon_{i,t} \quad \text{for } t = 3, 4, \ldots, T. \tag{6}$$

GMM estimation of the system formed by eqs.(5) and (6) exploits the following orthogonality conditions: $E\left[\Delta\mathbf{x}_{i,t-1}\left(\eta_i + \epsilon_{i,t}\right)\right] = 0$, $E\left[\sum_{t=3}^{T} D_t\left(\eta_i + \epsilon_{i,t}\right)\right] = 0$, $E\left[\mathbf{x}_{i,t-2}\Delta\epsilon_{i,t}\right] = 0$, $E\left[\sum_{t=3}^{T} D_t \Delta\epsilon_{i,t}\right] = 0$, and $E\left[\sum_{t=3}^{T} D_{t-1}\Delta\epsilon_{i,t}\right] = 0$. Results for the two-step system GMM estimation are presented in the upper part of Table 1. The standard errors, presented in the parentheses, are robust and corrected according to Windmeijer (2005). Because of the two-step procedure, the Hansen test is an appropriate test for overidentifying restrictions. Results for the Hansen test, the difference-in-Hansen test and the Arellano-Bond autocorrelation test are also reported. All specifications include year effect dummies, but we refrain from reporting the estimated year effects in order to conserve space.

[Table 1: Estimation results for the time constant coefficient model]

The Hansen test is easily passed for Chemicals, Motor vehicles and Pulp and paper, while the overidentifying restrictions are rejected for Wood prod-

[2] The data is from the Structural Business Statistics from Statistics Sweden. The original database contains detailed information on the income statements, balance sheets, and physical investment of all firms active in Sweden, including private and public firms but not financial firms. Most of the data are obtained from registers at the Swedish national tax agency.

ucts, Publishing and printing, and Machinery. Clearly, one reason for the non-rejection in the first three industries may be the relatively small sample size for these industries. For the industries where there is evidence that the model is mis-specified, increasing the lag length for the instruments does not really help: results, shown in the lower part of Table 1, strongly indicate the overidentifying restrictions should be rejected in all of the three problematic industries when levels variables dated $t-3$ and differenced variables dated $t-2$ are used as instruments.

The specification of the production function with time varying slope coefficients (hereafter, the TVC model) is as follows. The levels equation is specified as

$$y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta}_t + \delta_t D_t + (\eta_i + \epsilon_{i,t}) \quad \text{for } t = 4, 5, \ldots, T \tag{7}$$

The differenced equation is then expressed as

$$\Delta y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta}_t - \mathbf{x}'_{i,t-1}\boldsymbol{\beta}_{t-1} + \delta_t D_t - \delta_{t-1} D_{t-1} + \Delta\epsilon_{i,t} \quad \text{for } t = 4, 5, \ldots, T \tag{8}$$

and the following orthogonality conditions are exploited: $E\left[\Delta\mathbf{x}_{i,t-1}\left(\eta_i + \epsilon_{i,t}\right)\right] = 0$, $E\left[\sum_{t=4}^{T} D_t\left(\eta_i + \epsilon_{i,t}\right)\right] = 0$, $E\left(\mathbf{x}_{i,t-2}\Delta\epsilon_{i,t}\right) = 0$, $E\left(\mathbf{x}_{i,t-3}\Delta\epsilon_{i,t}\right) = 0$, $E\left[\sum_{t=4}^{T} D_t\Delta\epsilon_{i,t}\right] = 0$, and $E\left[\sum_{t=4}^{T} D_{t-1}\Delta\epsilon_{i,t}\right] = 0$. Table 2 reports results. Stata code for estimation of the TVC model, and an example, can be found in Appendix 2. Again, all specifications include year effect dummies, but we refrain from reporting the estimated year effects in order to conserve space.

[Table 2: Estimation results for the time varying coefficient model]

For the three industries that were satisfactory in terms of the Hansen test in the CC model, the Hansen test in this TVC model is clearly passed. The Arellano-Bond AR(2) test and the difference-in-Hansen test are also satisfactory. For the Chemicals and the Motor vehicles, coefficient stability is accepted by the Wald test for both labor or capital at the 5 percent significance level, which is expected. Hence, for these industries, assuming time constant slope coefficients does not seem restrictive. In contrast, for the Pulp and paper, the joint Wald test rejects the null of coefficient stability. This result suggests that the Wald test may be more powerful than the Sargan-Hansen test for detecting time varying slope coefficients. This conjecture is supported by simulation results reported in Appendix 1.[3]

For the other three industries, where the Hansen test results led us to reject the null in the CC model, the TVC model provides satisfactory results in terms of the Sargan-Hansen test, the difference-in-Hansen test and the Arellano-Bond AR(2) test, except for the last industry where the dif-in-Hansen test rejects the null. The Wald tests all indicate that coefficient stability should be rejected, as expected. Hence, restricting the slope coefficients to be constant over time appears to be a crucial modeling issue in the present application.

Before attributing the rejection of the Hansen test of the overidentifying restrictions in the CC model to parameter instability over time, we check another possibility that can result in a rejection of the null. Suppose the residual $\epsilon_{i,t}$ follows an AR(1) process,

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + u_t, \quad 0 < \rho < 1,$$

which implies that the residual in period t is correlated with all past residuals. If, as is commonly suspected, the regressors are contemporaneously correlated with the residual, lagged regressors will generally not be valid instruments in this case. The Sargan-Hansen test should therefore reject the null hypothesis that the orthogonality conditions hold for the population. Monte Carlo simulations in Appendix 1 confirm that autocorrelation in the residual may result in a high frequency of rejections by the Sargan-Hansen test as well as by the the Arellano-Bond AR(2) test in the CC model. The simulation also confirms that, in this case, applying the TVC model does not solve the problem: the Sargan-Hansen test and the AR(2) test are still likely to reject the null. In contrast, the the simulation shows that, when a rejection of the overidentifying restrictions in a CC model is only attributed to parameter instability, estimation using the TVC model does yield estimates that pass both the Sargan-Hansen and the AR(2) tests. Hence, it appears that the conventional methods for testing will be helpful in enabling researchers to distinguish between autocorrelation in the error term and time varying coefficients as possible reasons why the overidentifying restrictions may be rejected.

When we go back to our empirical results for the TVC model, the AR(2)

[3] The Monte Carlo simulations in Appendix 1 show that the Wald test in CC models reject the null of parameter stability more often than the Sargan-Hansen test in TVC models does. This may explain why the null is accepted in the Sargan-Hansen test but not in the Wald test for the Pulp and paper industry.

tests for Wood products, Publishing and printing, and Machinery are all passed. For these industries, we thus attribute the rejection of the overidentifying restrictions in the CC model to parameter instability over time. We therefore prefer the estimates from the TVC model.

Presenting a full set of results for TVC models may be impractical for spatial reasons. In many cases, researchers are only interested in estimating the average effects of the regressors. We therefore present in Table 2 the unweighted average values of the estimated time varying coefficients.[4] For comparison, Table 3 reports results for reestimating the CC model using the same set of samples. Observe that the averages of the time varying coefficients sometimes differ quite substantially from the estimated time constant coefficients. Simulation results shown in Appendix 1 further indicate that incorrectly imposing time constant coefficients may result in biased estimates of period averages. In contrast, the time-averaged coefficient estimates in the TVC model are not significantly different from the average of the true coefficients.

[Table 3: Estimation results for the time varying coefficient model]

4 Conclusions

We have studied the system GMM estimator proposed by Arellano and Bover (1995) and Bond and Blundell (1998), focusing on the implications of time varying slope coefficients for the Sargan-Hansen specification test. We have pointed out that, given how the system GMM model is specified, time varying coefficients would violate the overidentifying restrictions underlying the estimator.

Generalizing the system GMM model to allow for time varying coefficients is reasonably straightforward. Using Swedish firm-level data, we report system GMM results which indicate that allowing for time varying slope coefficients can result in more satisfactory Sargan-Hansen test results. In particular, when we assume the coefficients to be constant over time, the Sargan-Hansen tests reject the null of valid orthogonality conditions for three industries. When we instead estimate an empirical model with time varying coefficients, the Sargan-Hansen test no longer rejects the null for these industries.

A common response by researchers to a Sargan-Hansen test result indicating that the overidentifying restrictions should be rejected is to modify the lag length for the instrument set. However, if coefficient instability is the source

[4] The standard errors are obtained by the delta method using the covariance matrix.

of the specification problem such a response will neither be appropriate nor effective. Our analysis shows that assuming time constant coefficients may be overly restrictive, that it is straightforward to relax this assumption, and that doing so can be an effective way of resolving the problem. Our results also show that standard testing methods can be effective in distinguishing between this type of specification problem and serial correlation in the error term.

References

[1] Arellano, Manuel and Bover, Olympia (1995), "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68(1), 29-51.

[2] Blundell, Richard and Bond, Stephen (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87(1), 115-43.

[3] Blundell, Richard and Bond, Stephen (2000), "GMM Estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19(3), 321-40.

[4] Windmeijer, Frank (2005), "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of Econometrics*, 126(1), 25-51.

Appendix 1: Monte Carlo simulations

We implement Monte Carlo simulations, under different assumptions of the underlying data generation process in terms of variation in the slope coefficient and serial correlation in the residual. Firstly, we are interested in to which extent deviations from a benchmark case (where the underlying data generation process has a constant coefficient and no autocorrelation in the residual) affect the probability of a rejection by the Hansen test, the AR(2) test and the Wald test. Secondly, we investigate whether incorrectly imposing time constant coefficients biases estimated coefficients.

Data generation process

We consider a simple univariate data generation process for $y_{i,t}$:

$$y_{i,t} = \beta_t x_{i,t} + \epsilon_{i,t} \tag{9}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. The residual $\epsilon_{i,t}$ may be serially correlated:

$$\epsilon_{i,t} = \rho \epsilon_{i,t-1} + v_{i,t},$$

where $v_{i,t} = N\left(0, \sigma_v^2\right)$ and $0 \leq \rho \leq 1$. The variable $x_{i,t}$ is generated by an AR(1) process:

$$x_{i,t} = \alpha + \lambda x_{i,t-1} + \delta \epsilon_{i,t} + u_{i,t} \tag{10}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where λ is an autocorrelation coefficient and $u_{i,t} = N\left(0, \sigma_u^2\right)$. The term $\delta \epsilon_{i,t}$ is added to create endogeneity between $x_{i,t}$ and $\epsilon_{i,t}$, with the covariance given by $Cov\left(x_{i,t}, \epsilon_{i,t}\right) = \delta$. The standard deviation of $x_{i,t}$ within i over time is expressed by $\sigma_x = \sigma_u / \sqrt{1-\lambda^2}$.

We define $x_{i,0} = 1$ and set the expected value of $x_{i,t}$ as $E\left[x_{i,t}\right] = 1$. Because $E\left[x_{i,t}\right]$ is expressed by $E\left[x_{i,t}\right] = \alpha / \left(1-\lambda\right)$, α is set to $\alpha = \left(1-\lambda\right) E\left[x_{i,t}\right] = 1 - \lambda$.

The coefficient β_t is drawn as mutually independent random variable, $\beta_t \sim N\left(1, \sigma_\beta^2\right)$. It changes over time when σ_β is non-zero, while it is constant when $\sigma_\beta = 0$.

We provide the following values to the parameters: $N = 1000$, $T = 10$, $\lambda = 0.9$, $\sigma_\epsilon = 0.2$, $\sigma_u = 0.2$, and $\delta = 0.2$. We assign different values to σ_β and ρ.

Because β_t is drawn randomly, the average and the standard deviation of the realized values may not be equal to the theoretical values of 1 and σ_β, respectively. For each sample, we adjust the realized values of β_t as

$$\tilde{\beta}_t = \frac{\sigma_\beta \left(\beta_t - \hat{\beta}\right)}{\hat{\sigma}_\beta} \tag{11}$$

where $\hat{\beta}$ and $\hat{\sigma}_\beta$ are the average and the standard deviation of the realized values of β_t, respectively. We then replace β_t with $\tilde{\beta}_t$. This adjustment guarantees that the average and the standard deviation of β_t for each sample are equal to 1 and σ_β, respectively.

Results

We implement 1,000 Monte Carlo replications with five sets of parameters:

$(\sigma_\beta,\ \rho) = (0,\ 0), (0.025,\ 0), (0.050,\ 0), (0,\ 0.25), (0,\ 0.5)$, and estimate the CC model and the TVC model using the two-step system GMM. Table 4 reports the probability of a rejection by different specification tests at the 5% significance level. In the benchmark case (the column [1]) where $\sigma_\beta = 0$ and $\rho = 0$, all probabilities are around 5%, which is in line with our expectations. As we increase the value of σ_β (the columns [2] and [3]), the probability of a rejection by the Hansen test and the dif-in-Hansen test increases. Because the residual and lagged regressors are uncorrelated in our data generation process, the rejection is solely attributed to parameter instability. It is also reflected in a higher frequency of rejections by the Wald test of stability of the estimated time varying coefficients in the TVC model. Note that the Wald test is slightly more likely to reject the null than the Hansen test in the CC model.

[Table 4: Results for Monte Carlo simulations]

The probabilities of a rejection by the Hansen test and the dif-in-Hansen test in the TVC model remain around 5%, which confirms our argument that these tests purely verify the validity of overidentifying restrictions and instruments.

Are the coefficient estimates biased when we incorrectly impose a restriction of a time constant coefficient? Table 4 also reports the probability of the estimated β being significantly different from the average of the true β.[5] It is shown that the probability increases as σ_β rises, suggesting that incorrectly imposing time constant coefficients may result in biased estimates. In contrast, the average of the estimated time varying coefficients in the TVC model does not significantly differ from the true coefficient average.

Autocorrelation in the residual may also results in a rejection of overidentifying restrictions by the Sargan-Hansen test. Columns [4] and [5] report results for the data generation processes with residual autocorrelation. It is shown that the probability of a rejection by the Hansen test increases when the autocorrelation coefficient becomes large. In contrast to the parameter instability cases, allowing for time varying slope coefficients does not solve the problem: the Hansen test and the AR(2) test are still likely to reject the null. This difference can be exploited in distinguishing a rejection in the Hansen test attributed to parameter instability and to residual autocorrelation.

[5] The average is taken from $\beta_3,\ \beta_4,\ \ldots,\ \beta_{10}$ because the CC model covers equations for $T \geq 3$.

Appendix 2: STATA instruction for estimation of the TVC model

This instruction describes how to estimate a dynamic panel model with time-varying coefficients using the command XTABOND2 in STATA.

We consider the following panel data model with time-varying coefficients β_t:

$$y_{i,t} = \beta_t x_{i,t} + \delta_t D_t + (\eta_i + \epsilon_{i,t}) \quad \text{for } t = 4, 5, \ldots, 10, \tag{12}$$

where $y_{i,t}$ is a dependent variable, $x_{i,t}$ is a regressor, δ_t is year specific effects, D_t is year dummies, η_i is unobserved firm-specific effects, and $\epsilon_{i,t}$ is a residual term that can potentially be correlated with $x_{i,t}$. We assume for convenience that the numbers of time periods is 10. Because we will use the second and the third lags of $x_{i,t}$ as instruments for the difference equations, the model is defined for $t \geq 4$. The level equation can in fact be defined even for $t = 3$ because the instrument used for levels equations is the lagged first difference. This, however, complicates the application of XTABOND2. We therefore define both the level and the difference equations for $t \geq 4$.

Eq.(12) constitutes the level equations. The instruments used for the estimation are $\Delta x_{i,t-1}$ for $x_{i,t}$, and 1 for D_t. The expression for each period is:

$$\begin{aligned}
y_{i,4} &= \beta_4 x_{i,4} + \delta_4 D_4 + (\eta_i + \epsilon_{i,4}) \quad \text{inst. } \Delta x_{i,3}, 1 && (13)\\
y_{i,5} &= \beta_5 x_{i,5} + \delta_5 D_5 + (\eta_i + \epsilon_{i,5}) \quad \text{inst. } \Delta x_{i,4}, 1 && (14)\\
&\vdots\\
y_{i,10} &= \beta_{10} x_{i,4} + \delta_{10} D_{10} + (\eta_i + \epsilon_{i,10}) \quad \text{inst. } \Delta x_{i,9}, 1 && (15)
\end{aligned}$$

The difference equations are expressed as

$$\Delta y_{i,t} = -\beta_{t-1} x_{i,t-1} + \beta_t x_{i,t} - \delta_{t-1} D_{t-1} + \delta_t D_t + \Delta\epsilon_{i,t} \quad \text{for } t = 4, 5, \ldots, 10 \tag{16}$$

The instrumentals used for the difference equations are $x_{i,t-2}$ and $x_{i,t-3}$ for

$x_{i,t-1}$ and $x_{i,t-2}$, 1 for D_{t-1} and D_t. The expression for each period is:

$$\Delta y_{i,4} = -\beta_3 x_{i,3} + \beta_4 x_{i,4} - \delta_3 D_3 + \delta_4 D_4 + \Delta\epsilon_{i,4} \quad \text{inst. } x_{i,1}, x_{i,2}, 1, 1 \tag{17}$$

$$\Delta y_{i,5} = -\beta_4 x_{i,4} + \beta_5 x_{i,5} - \delta_4 D_4 + \delta_5 D_5 + \Delta\epsilon_{i,5} \quad \text{inst. } x_{i,2}, x_{i,3}, 1, 1 \tag{18}$$

$$\vdots$$

$$\Delta y_{i,10} = -\beta_9 x_{i,9} + \beta_{10} x_{i,10} - \delta_9 D_9 + \delta_{10} D_{10} + \Delta\epsilon_{i,10} \quad \text{inst. } x_{i,7}, x_{i,8}, 1, 1 \tag{19}$$

I name the variables as follows in STATA:

dependent variable: `y`
regressor: `x`
year: `t`
year dummies: `t1, t2, ..., t10`

In addition, I generate the following terms. Firstly, the cross terms of `x` and time dummies:

```
"x_t3"  = x * t3
"x_t4"  = x * t4
"x_t5"  = x * t5
   ⋮
"x_t10" = x * t10
```

Secondly, the cross terms of the second lag of `x` and time dummies:

```
"L2x_t4"  = L2.x * t4
"L2x_t5"  = L2.x * t5
   ⋮
"L2x_t10" = L2.x * t10
```

Next, the cross terms of the third lag of `x` and time dummies:

```
"L3x_t4"  = L3.x * t4
"L3x_t5"  = L3.x * t5
   ⋮
"L3x_t10" = L3.x * t10
```

Lastly, the cross terms of the lagged difference of `x` and time dummies:

```
"LDx_t4"  = LD.x * t4
"LDx_t5"  = LD.x * t5
```

```
    ⋮
"LDx_t10" = LD.x * t10
```

The values of these cross terms at different periods are shown in the attached Table 5.

The STATA command used for the estimation is as follows:

```
XTABOND2 y x_t3-x_t10 t3-t10 if t>=4,
gmm(LDx_t4-LDx_t10, lag(0 0) equation(level) passthru)
iv(t3-t10, equation(level))
gmm(L2x_t4-L2x_t10 L3x_t4-L3x_t10, lag(0 0) equation(diff))
iv(t3-t10, equation(diff))
twostep robust noc
```

To easily understand, separate lines of `gmm`- and `iv`-instruments are presented for the levels equation (noted as `equation(level)`) and the difference equation (noted as `equation(diff)`). For the levels equations, the model specification `y x_t3-x_t10 t3-t10` implies that the level equation at, for instance, $t = 4$ corresponds to Eq.(13) because only `x_t4` and `t4` are nonzero. The instrument specification `LDx_t4-LDx_t10` implies that the instrument applied at $t = 4$ is $\Delta x_{i,3}$ because only `LDx_t4` is nonzero. Note that `passthru` tells STATA not to take the first difference of the instruments specified for the level equations as STATA otherwise does it. Note also that the `iv`-instruments specified for the level equations are not automatically first-differenced. The instruments applied at $t = 4$ are therefore 1 (for D_4).

For the difference equations, STATA automatically takes the first difference of the model specification. This implies the difference equation at $t = 4$ corresponds to Eq.(17) because the first difference of `x_t3` and that of `x_t4` at $t = 4$ are $-x_{i,3}$ and $x_{i,4}$, respectively, and other terms in `x_t3-x_t10` are zero (see Table 5). Similarly, the first difference of `t3` and that of `t4` at $t = 4$ are -1 and 1, respectively, and the other terms in `t3-t10` are zero. The `gmm`-instrument specification for the difference equations implies that the instruments applied at $t = 4$ are $x_{i,1}$ and $x_{i,2}$ because only `L2x_t4` and `L3x_t4` are nonzero. For the `iv`-instrument specification, STATA automatically takes the first difference. The `iv`-instruments applied at $t = 4$ become -1 and 1 (for D_3 and D_4).

The `if`-condition `t>=4` excludes the difference equation at $t = 3$. Without it, the equation $\Delta y_{i,3} = \beta_t x_{i,3} + \delta_3 D_3 + \Delta\epsilon_{i,3}$ is included in estimation, which is nonsense.

	Chemicals	Motor vehicles	Pulp and paper	Wood products	Publishing and printing	Machinery
System GMM estimation with the second lag of the regressors as instruments for the differenced equation						
Number of observations	2251	2476	1614	11319	15427	14995
Number of firms	404	449	278	1996	2787	2631
Variables						
$l_{i,t}$	1.001	0.952	0.866	1.064	0.925	0.993
	(0.068)	(0.058)	(0.094)	(0.047)	(0.055)	(0.032)
$k_{i,t}$	0.069	0.094	0.173	0.021	0.125	0.050
	(0.040)	(0.051)	(0.063)	(0.031)	(0.023)	(0.025)
Specification tests						
Arellano-Bond test for AR(2) (p-value)	0.091	0.127	0.891	0.013	0.647	0.398
Hansen test (p-value)	0.832	0.594	0.338	0.000	0.021	0.000
Dif-in-Hansen test (p-value)	0.638	0.588	0.957	0.065	0.198	0.000
System GMM estimation with the third lag of the regressors as instruments for the differenced equation						
Number of observations				9322	12640	12361
Number of firms				1759	2450	2360
Variables						
$l_{i,t}$				1.166	0.995	0.983
				(0.064)	(0.063)	(0.044)
$k_{i,t}$				-0.029	0.142	0.052
				(0.045)	(0.033)	(0.037)
Specification tests						
Arellano-Bond test for AR(2) (p-value)				0.045	0.342	0.211
Arellano-Bond test for AR(3) (p-value)				0.214	0.353	0.304
Hansen test (p-value)				0.000	0.003	0.002
Dif-in-Hansen test (p-value)				0.052	0.001	0.034

Table 1: Estimation results for the time constant slope coefficient model (the CC model)

	Chemicals	Motor vehicles	Pulp and paper	Wood products	Publishing and printing	Machinery
Number of observations	1847	2026	1336	9322	12640	12361
Number of firms	357	403	252	1759	2450	2360
Variables						
$l_{i,\ 1999}$	0.927	0.963	1.069	0.971	0.948	0.972
	(0.101)	(0.067)	(0.162)	(0.046)	(0.056)	(0.039)
$l_{i,\ 2000}$	0.940	0.967	1.109	0.999	0.960	0.963
	(0.103)	(0.076)	(0.166)	(0.048)	(0.056)	(0.041)
$l_{i,\ 2001}$	0.945	0.958	1.109	1.005	0.951	0.943
	(0.116)	(0.076)	(0.162)	(0.051)	(0.056)	(0.040)
$l_{i,\ 2002}$	0.924	0.960	1.205	0.972	0.957	0.950
	(0.132)	(0.076)	(0.163)	(0.054)	(0.056)	(0.041)
$l_{i,\ 2003}$	0.945	0.921	1.274	0.953	1.010	0.954
	(0.142)	(0.070)	(0.211)	(0.054)	(0.058)	(0.041)
$l_{i,\ 2004}$	0.969	0.930	1.264	0.969	0.989	0.959
	(0.154)	(0.079)	(0.198)	(0.054)	(0.058)	(0.042)
$l_{i,\ 2005}$	0.951	0.925	1.329	0.957	1.035	0.976
	(0.155)	(0.092)	(0.213)	(0.054)	(0.061)	(0.043)
$l_{i,\ 2006}$	0.983	0.965	1.264	0.983	1.058	1.005
	(0.170)	(0.099)	(0.217)	(0.056)	(0.062)	(0.044)
$k_{i,\ 1999}$	0.125	0.113	0.038	0.104	0.143	0.148
	(0.078)	(0.068)	(0.106)	(0.039)	(0.024)	(0.029)
$k_{i,\ 2000}$	0.111	0.102	0.041	0.106	0.137	0.171
	(0.079)	(0.074)	(0.108)	(0.043)	(0.024)	(0.032)
$k_{i,\ 2001}$	0.133	0.108	0.023	0.098	0.140	0.177
	(0.090)	(0.078)	(0.111)	(0.047)	(0.025)	(0.033)
$k_{i,\ 2002}$	0.151	0.092	-0.048	0.133	0.146	0.182
	(0.106)	(0.083)	(0.113)	(0.050)	(0.025)	(0.034)
$k_{i,\ 2003}$	0.141	0.130	-0.101	0.159	0.122	0.184
	(0.117)	(0.083)	(0.150)	(0.052)	(0.026)	(0.036)
$k_{i,\ 2004}$	0.128	0.145	-0.101	0.143	0.141	0.185
	(0.127)	(0.089)	(0.138)	(0.053)	(0.025)	(0.037)
$k_{i,\ 2005}$	0.140	0.135	-0.159	0.156	0.121	0.181
	(0.128)	(0.099)	(0.150)	(0.054)	(0.026)	(0.037)
$k_{i,\ 2006}$	0.125	0.106	-0.111	0.159	0.102	0.163
	(0.139)	(0.101)	(0.149)	(0.055)	(0.026)	(0.038)
Average effects						
$l_{i,t}$	0.948	0.949	1.203	0.976	0.989	0.965
	(0.129)	(0.075)	(0.183)	(0.050)	(0.056)	(0.040)
$k_{i,t}$	0.132	0.116	-0.052	0.132	0.132	0.174
	(0.105)	(0.082)	(0.125)	(0.048)	(0.023)	(0.034)
Specification tests						
Arellano-Bond test for AR(2) (p-value)	0.327	0.203	0.597	0.082	0.351	0.137
Hansen test (p-value)	0.518	0.291	0.543	0.104	0.206	0.208
Dif-in-Hansen test (p-value)	0.155	0.326	0.834	0.082	0.443	0.014
Wald test of stability of the coefficients for $l_{i,t}$ (p-value)	0.853	0.610	0.061	0.063	0.000	0.044
Wald test of stability of the coefficients for $k_{i,t}$ (p-value)	0.835	0.117	0.039	0.004	0.078	0.165
Joint test of the two Wald tests (p-value)	0.528	0.062	0.002	0.000	0.000	0.000

Table 2: Estimation results for the time varying slope coefficient model (the TVC model)

	Chemicals	Motor vehicles	Pulp and paper	Wood products	Publishing and printing	Machinery
System GMM estimation with the second lag of the regressors as instruments for the differenced equation						
Number of observations	1847	2026	1336	9322	12640	12361
Number of firms	357	403	252	1759	2450	2360
Variables						
$l_{i,t}$	0.894	0.943	1.094	1.070	0.934	0.986
	(0.130)	(0.070)	(0.106)	(0.051)	(0.064)	(0.046)
$k_{i,t}$	0.141	0.118	0.066	0.029	0.135	0.116
	(0.073)	(0.067)	(0.082)	(0.039)	(0.023)	(0.031)
Specification tests						
Arellano-Bond test for AR(2) (p-value)	0.446	0.166	0.617	0.058	0.379	0.174
Hansen test (p-value)	0.395	0.163	0.188	0.000	0.059	0.000
Dif-in-Hansen test (p-value)	0.350	0.750	0.800	0.012	0.155	0.000

Table 3: The CC model reestimated using the same set of samples as in Table 2

	[1]	[2]	[3]	[4]	[5]
	$\sigma_\beta = 0, \rho = 0$ (Benchmark)	$\sigma_\beta = 0.025, \rho = 0$	$\sigma_\beta = 0.050, \rho = 0$	$\sigma_\beta = 0, \rho = 0.25$	$\sigma_\beta = 0, \rho = 0.50$
Specification with time-constant coefficient (the CC model)					
Probability of a rejection by the Hansen test	0.048	0.654	1.000	0.306	0.898
Probability of a rejection by the dif-in-Hansen test	0.050	0.198	0.627	0.403	0.957
Probability of a rejection by AR(2) test	0.059	0.048	0.058	1.000	1.000
Probability of the estimated β being significantly different from the average of the true β	0.044	0.096	0.234	0.891	1.000
Specification with time-specific coefficient (the TVC model)					
Probability of a rejection by the Hansen test	0.049	0.049	0.043	0.227	0.666
Probability of a rejection by the dif-in-Hansen test	0.046	0.054	0.042	0.139	0.508
Probability of a rejection by AR(2) test	0.059	0.044	0.053	1.000	1.000
Probability of a rejection by the Wald test	0.056	0.776	1.000	0.134	0.472
Probability of the average of the estimated β being significantly different from the average of the true β	0.051	0.057	0.045	0.935	1.000

Table 4: Results for Monte Carlo simulations

	t = 3	t = 4	t = 5	⋯	t = 10
t3	1	0	0	⋯	0
t4	0	1	0	⋯	0
t5	0	0	1	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
t10	0	0	0	⋯	1
D(t3)	1	-1	0	⋯	0
D(t4)	0	1	-1	⋯	0
D(t5)	0	0	1	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
D(t9)	0	0	0	⋯	-1
D(t10)	0	0	0	⋯	1
x_t3	x(3)	0	0	⋯	0
x_t4	0	x(4)	0	⋯	0
x_t5	0	0	x(5)	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
x_t10	0	0	0	⋯	x(10)
D(x_t3)	x(3)	- x(3)	0	⋯	0
D(x_t4)	0	x(4)	- x(4)	⋯	0
D(x_t5)	0	0	x(5)	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
D(x_t9)	0	0	0	⋯	- x(9)
D(x_t10)	0	0	0	⋯	x(10)
L2x_t4	0	x(2)	0	⋯	0
L2x_t5	0	0	x(3)	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
L2x_t10	0	0	0	⋯	x(8)
L3x_t4	0	x(1)	0	⋯	0
L3x_t5	0	0	x(2)	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
L3x_t10	0	0	0	⋯	x(7)
LDx_t4	0	x(3)-x(2)	0	⋯	0
LDx_t5	0	0	x(4)-x(3)	⋯	0
⋮	⋮	⋮	⋮	⋱	⋮
LDx_t10	0	0	0	⋯	x(10)-x(9)

Table 5: Values of the variables at different periods