Generalized method of moments

Khanh Duong

Personal repository

June 11, 2024

Introduction

The modelling equation can be rewritten as

$$Y_{(i,t)} = \theta_0 + \theta_1 Y_{(i,t-1)} + \theta_2 X_{(i,t)} + \mu_i + \epsilon_{(i,t)} \tag{1}$$

where $i = [1, N]$ represents many **cross-sectional units** and $t = [1, T]$ represents few **time periods**, with **country-specific effects** μ_i and the idiosyncratic error term $\epsilon_{(i,t)}$. The regressors $X_{(i,t)}$ can have different properties with respect to their correlation with the error term $\epsilon_{(i,t)}$.

Classification of regressors

The choice of whether to treat the regressors as **strictly exogenous**, **predetermined**, or **exogenous** depends on the nature of the data and the research question being addressed.

Strictly exogenous regressors

- If the regressors are strictly exogenous, it means that the variables X are uncorrelated with **past**, **present**, and **future** values of the error term.
- In other words, $E[X_{it}\epsilon_{is}] = 0$ for all t and s.
- This implies that the values of X do not depend on the values of the error term at any point in time.

Predetermined regressors

- If the regressors are predetermined, it means that the variables X are correlated with **past** values of the error term but **not present nor future** values.
- In other words, $E[X_{it}\epsilon_{is}] \neq 0$ for $s < t$, but $E[X_{it}\epsilon_{is}] = 0$ for $s \geq t$.

Endogenous regressors

- If the regressors are endogenous, it means that the variables X are correlated with **past** and **present** values of the error term but **not future** values.
- In other words, $E[X_{it}\epsilon_{is}] \neq 0$ for $s < t$, but $E[X_{it}\epsilon_{is}] = 0$ for $s \geq t$.

Country-specific effects

To account for the **country-specific effects** μ_i, we employ the first difference transformation to Equation 1 into

$$\Delta Y_{(i,t)} = \theta_1 \Delta Y_{(i,t-1)} + \theta_2 \Delta X_{(i,t)} + \Delta \epsilon_{(i,t)}$$

The method of moments (1/2)

- Afterwards, we generate **instruments** for the lagged dependent variable by using the second and third lags of Y, either as differences or lagged levels.
- Utilizing the **Arellano–Bond approach**[1], we construct instrumental variables Z_i by imposing these moment conditions on the first difference model: $E[Y_{i,t-s}\Delta\epsilon_{it}] = 0$ for $s = 2, t$.
- When the individual effect term exhibits high variance across individual observations, the Arellano-Bond estimator may yield **poor performance** in finite samples.
- This occurs because the lagged dependent variables become **weak instruments** under such circumstances.

[1] Arellano, M. and Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, 58(2), pp.277-297.

The method of moments (2/2)

- **Blundell and Bond**[2] derived a condition that allows for the utilization of an additional set of moment conditions for the level model: $E[\Delta Y_{i,t-1} \Delta \epsilon_{it}] = 0$ for $t = 2, T$.
- Similarly, depending on the nature of the variables X, other instruments for those variables are incorporated as well.
- Overall, we assemble the stacked moment conditions as: $E[Z_i' \Delta \epsilon_i] = 0$, where the instrumental variables Z_i are constructed from values of X and Y.
- Hence, we refer to this approach as **the method of moments**.

[2] Blundell, R. and Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1), pp.115-143.

The GMM estimator (1/2)

- The moment conditions form a system of equations with unknown coefficients θ: $E[Z_i' \Delta\epsilon_i] = E[m_i(\theta)] = 0$.
- It is evident that the vector of moment conditions (m_i) is larger than the vector of coefficients θ, meaning that $E[m_i(\theta)]$ cannot simultaneously satisfy the condition of being equal to zero.
- Therefore, we aim to minimize the squared distance between the sample moment conditions and zero, which can be represented as $\|\hat{m}_i(\theta)\|_W^2 = \hat{m}_i(\theta)^T W \hat{m}_i(\theta) = f(\theta)$, where W is **the weight matrix of moments** and $\hat{m}^T$ denotes transposition.
- By using a generalized metric for moment conditions $f(\theta)$, this method is referred to as the **generalized method of moments**.

The GMM estimator (2/2)

- The minimal value of $f(\theta)$ occurs when its derivative with respect to θ is equal to zero, i.e., $\frac{df}{d\theta} = 0$. Obviously, the GMM estimator depends on the choice of the weight matrix W.
- A commonly used proposal for the weight matrix is $\hat{W} = \left(\frac{1}{N} Z' H Z\right)^{-1}$, where Z is the instrument matrix. Under the Blundell and Bond approach for the system GMM estimator, H is equal to the identity matrix (I).
- The estimation generated by this method is called the **one-step system** GMM estimator, and its weighting matrix is denoted as $\hat{W}_1$.
- The **two-step estimator** utilizes $\hat{W}_2 = \left(\frac{1}{N} Z' \hat{s}_1 \hat{s}_1' Z\right)^{-1}$, where $\hat{s}_1$ represents the residuals obtained from the one-step estimation.

Robustness (1/4)

- The absence of higher-order serial correlation in $\Delta\epsilon_{it}$ is crucial for the validity of using $Y_{i,t-2}$, $Y_{i,t-3}$, and other variables as instruments in the GMM framework.
- Similarly, it is important for the instruments of predetermined and endogenous variables.
- To test for this, the **Arellano-Bond serial-correlation test** should be conducted. The test statistic, following an asymptotic $N(0, 1)$ distribution, examines the null hypothesis $H_0 : \mathrm{Corr}(\Delta\epsilon_{it}, \Delta\epsilon_{it-j}) = 0$ for $j > 0$.
- If the null hypothesis is rejected for $j = 1$ but not rejected for $j > 1$, it suggests that the model passes this specification test.

Robustness (2/4)

- Additionally, in overidentified models (e.g. GMM), where the number of moment conditions L exceeds the number of unknown coefficients K, it is important to test the validity of the $L-K$ overidentifying restrictions. .
- These tests assume that at least K instruments are valid.
- The **Sargan overidentification test** is commonly used for this purpose.
- The test statistic follows an asymptotic χ^2 distribution with df degrees of freedom, where df is equal to $L-K$.
- The Sargan overidentification test statistic $J(\hat{\theta}, W)$ is calculated as: $\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_i(\hat{\theta})\right)' W \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_i(\hat{\theta})\right)$, where N represents the number of observations or individuals in the sample.

Robustness (3/4)

- In the presence of **heteroskedasticity**, panel-robust standard errors can be computed using system GMM estimation.
- In this case, the one-step GMM estimator remains consistent under heteroskedasticity but is no longer efficient.
- The two-step standard errors are biased in finite samples, so the Windmeijer finite-sample correction should be applied.
- The corrected standard errors are still biased but less severely.
- However, when using panel-robust standard errors, the Sargan overidentification test cannot be computed because the asymptotic distribution is unknown.

Robustness (4/4)

- It should be noted that **multicollinearity** is not a problem when using instrumental analysis, which isolates the effect of explanatory variables from group effects and other variable effects.
- In instrumental analysis, the focus is on the strength of the instruments rather than the correlation among independent variables.
- Multicollinearity among exogenous independent variables (i.e., variables not affected by endogeneity) generally does not affect the validity of the instruments or the identification strategy used in instrumental analysis.
- However, perfect multicollinearity among the instruments themselves can weaken their ability to address endogeneity, and system GMM will **drop variables with perfect multicollinearity**.