Problem Set 5 Solution

May 14th, 2009 by Yang

1. Adding More Instruments in 2SLS Estimation

(i) Since Z is an $n \times n$ matrix, the projection matrix becomes an identity matrix.

$$P_z := Z(Z'Z)^{-1}Z' = ZZ^{-1}(Z')^{-1}Z' = I_n$$

Therefore,

$$\widehat{\beta}_{2SLS} = (X'P_zX)^{-1}X'P_zY = (X'X)^{-1}X'Y = \widehat{\beta}_{OLS}$$

(ii) Note that the asymptotic variances of the two estimators are given by

$$\begin{aligned} \mathrm{AV}(\widehat{\beta}_1) &= \sigma^2 \left(Ex_iz_{1i}'(Ez_{1i}z_{1i}')^{-1}Ez_{1i}x_i'\right)^{-1} \\ \mathrm{AV}(\widehat{\beta}_2) &= \sigma^2 \left(Ex_iz_{2i}'(Ez_{2i}z_{2i}')^{-1}Ez_{2i}x_i'\right)^{-1} \end{aligned}$$

Hence, consistent estimators of these are

$$\begin{aligned} \widehat{\mathrm{AV}}_n(\widehat{\beta}_1) &= \sigma^2 \left(\frac{1}{n}\sum_{i=1}^n x_iz_{1i}' \left(\frac{1}{n}\sum_{i=1}^n z_{1i}z_{1i}'\right)^{-1} \frac{1}{n}\sum_{i=1}^n z_{1i}x_i'\right)^{-1} = n\sigma^2(X'P_{z1}X)^{-1} \\ \widehat{\mathrm{AV}}_n(\widehat{\beta}_2) &= \sigma^2 \left(\frac{1}{n}\sum_{i=1}^n x_iz_{2i}' \left(\frac{1}{n}\sum_{i=1}^n z_{2i}z_{2i}'\right)^{-1} \frac{1}{n}\sum_{i=1}^n z_{2i}x_i'\right)^{-1} = n\sigma^2(X'P_{z2}X)^{-1} \end{aligned}$$

Note that σ^2 is the true variance of ε_i, conditionally homoskedastic. For any n, $\widehat{\mathrm{AV}}_n(\widehat{\beta}_1) \geq \widehat{\mathrm{AV}}_n(\widehat{\beta}_2)$ a.s. in positive semidefinite sense. To see this, pick any n and any realization of random variables. Since

$$\begin{aligned} \widehat{\mathrm{AV}}_n(\widehat{\beta}_1) - \widehat{\mathrm{AV}}_n(\widehat{\beta}_2) \geq 0 &\Leftrightarrow \left[\widehat{\mathrm{AV}}_n(\widehat{\beta}_2)\right]^{-1} - \left[\widehat{\mathrm{AV}}_n(\widehat{\beta}_1)\right]^{-1} \geq 0 \\ &\Leftrightarrow X'P_{z2}X - X'P_{z1}X \geq 0 \end{aligned}$$

it suffices to prove that $X'P_{z2}X - X'P_{z1}X$ is a positive semidefinite matrix. Note that

$$\begin{aligned} X'P_{z2}X - X'P_{z1}X &= X'P_{z2}P_{z2}X - X'P_{z2}P_{z1}P_{z2}X \\ &= X'P_{z2}(I - P_{z1})P_{z2}X \\ &= X'P_{z2}(I - P_{z1})(I - P_{z1})P_{z2}X \end{aligned}$$

The first equality follows from the fact that P_{z2} is idempotent, and that $P_{z2}P_{z1} = P_{z1}P_{z2} = P_{z1}$ since Z_2 includes Z_1. The last equality holds since $(I - P_{z1})$ is idempotent. The expression has a quadratic form, so $X'P_{z2}X - X'P_{z1}X$ is positive semidefinite. Since this holds for every n and every realization of random variables, this holds in the limit as well.

$$\mathrm{plim}\widehat{\mathrm{AV}}_n(\widehat{\beta}_1) \geq \mathrm{plim}\widehat{\mathrm{AV}}_n(\widehat{\beta}_2)$$

or equivalently,

$$\text{AV}(\widehat{\beta}_1) \geq \text{AV}(\widehat{\beta}_2)$$

(iii) The statement is false. From (i), the (asymptotic) bias tends to go up as we use more instruments, even though they satisfy the moment conditions. But the variance gets smaller as more instruments are used provided that they are uncorrelated with the errors. We have verified this in (ii). So the choice of the optimal number of instruments is subject to the tradeoff between the bias and the variance.[1]

2. Finite Sample Properties of OLS and LAD Estimators

Results are provided in the following table. Consider first the basement specification: $n = 100$, $k = 2$ and $\beta = 0 \in \mathbb{R}^2$. As shown in the class, when $u_i \sim N(0, 1)$ (case 1), the OLS estimates are less dispersed, when $u_i \sim t_3$ (case 2), the LAD estimator works better, and when $u_i \sim t_1$ (case 3), the LAD estimator still works as well as in the other cases, but the OLS estimator suffers from huge outliers. This is basically because the OLS estimator has the variance proportional to $var(u_i)$, which has no impact on the variance of the LAD estimator theoretically. The ratio of the variances of the two estimators is consistent with the ARE (asymptotic relative efficiency), which is defined as the ratio of the asymptotic variances of the two estimators.

The mean biases of the two estimators are similar in case 1 and case 2, but the OLS estimator has a huge mean bias in case 3 while the LAD estimator does not. In fact, the OLS estimator is inconsistent when the error is Cauchy distributed, since the Cauchy distribution does not have the finite first moment, which means that the crucial condition of the OLS model, $Ex_iu_i = 0$, is not satisfied. But the median bias of the OLS estimator seems to be bounded, although it is still greater than that of the LAD estimator.

Consier now the other scenarios. β affects the result very little. When the same realizations of x_i and u_i are used, numerical values of the OLS estimator are not influenced by change of β at all. Those of the LAD estimator change a little, which is likely due to the mulitple minimizers of the criterion function. In contrast to the OLS model, the criterion function of the LAD model does not have a unique minimzer. k seems to have a little effect on the performance of the two estimators. There is a weak tendency that the variance of the LAD estimator gets less as k goes up. For too high k, the OLS estimator suffers from low degrees of freedom. So we can see in some cases that the LAD estimator has less variance than the OLS estimator. The bias seems to be unaffected. Larger n reduces the variance. Note that the asymptotic variance will be n times the observed variance. It seems to remain at the same level even if n gets larger. The ratio of the variances stays similar for different n, which justifies the theory. The bias becomes less for larger n, which is reasonable.

[1] Note that the bias does not systematically (monotonically) increase in the number of instruments, so we may obtain a less biased 2SLS estimator by including some instruments. But the bias can be calculated only when the true parameter is known, so practically it is not possible to see whether the bias really goes up or down by doing so.

TABLE1. Comparison of the OLS and the LAD estimators (the benchmark case)

```
*** n=100 / k=2 / beta=[ 0 0 ] ***

CASE 1: u~N(0,1)                OLS            LAD          ratio
-----------------------------------------------------------------
    Mean Bias    ||          0.0033         0.0030         1.1048
   Median Bias   ||          0.0053         0.0048         1.1100
    Variance     ||          0.0097         0.0152         0.6387
   M. S. Error   ||          0.0097         0.0152         0.6390
-----------------------------------------------------------------

CASE 2: u~t(3)                  OLS            LAD          ratio
-----------------------------------------------------------------
    Mean Bias    ||          0.0075         0.0079         0.9483
   Median Bias   ||         -0.0001         0.0057        -0.0194
    Variance     ||          0.0286         0.0194         1.4735
   M. S. Error   ||          0.0287         0.0195         1.4717
-----------------------------------------------------------------

CASE 3: u~Cauchy                OLS            LAD          ratio
-----------------------------------------------------------------
    Mean Bias    ||         40.8846        -0.0041     -9923.9689
   Median Bias   ||          0.0335        -0.0022       -15.0299
    Variance     ||    1639890.1841         0.0298  55010457.3049
   M. S. Error   ||    1639921.8409         0.0298  55035220.2315
-----------------------------------------------------------------
```

TABLE 2. The Effect of β

```
*** n=100 / k=2 / beta=[ 1 1 ] ***

CASE 1: u~N(0,1)                OLS            LAD          ratio
-----------------------------------------------------------------
    Mean Bias    ||          0.0033         0.0030         1.0915
   Median Bias   ||          0.0053         0.0046         1.1537
    Variance     ||          0.0097         0.0153         0.6378
   M. S. Error   ||          0.0097         0.0153         0.6381
-----------------------------------------------------------------

CASE 2: u~t(3)                  OLS            LAD          ratio
-----------------------------------------------------------------
    Mean Bias    ||          0.0075         0.0077         0.9743
   Median Bias   ||         -0.0001         0.0048        -0.0232
    Variance     ||          0.0286         0.0194         1.4758
   M. S. Error   ||          0.0287         0.0194         1.4742
-----------------------------------------------------------------

CASE 3: u~Cauchy                OLS            LAD          ratio
-----------------------------------------------------------------
    Mean Bias    ||         40.8846        -0.0041     -9894.9666
   Median Bias   ||          0.0335        -0.0019       -17.3639
    Variance     ||    1639890.1841         0.0299  54769039.6461
   M. S. Error   ||    1639921.8409         0.0299  54793648.4479
-----------------------------------------------------------------
```

```
*** n=100 / k=2 / beta=[ 1 2 ] ***

             OLS            LAD
-----------------------------
          0.0033         0.0030
          0.0053         0.0046
          0.0097         0.0153
          0.0097         0.0153
-----------------------------

             OLS            LAD
-----------------------------
          0.0075         0.0077
         -0.0001         0.0048
          0.0286         0.0194
          0.0287         0.0194
-----------------------------

             OLS            LAD
-----------------------------
         40.8846        -0.0041
          0.0335        -0.0019
    1639890.1841         0.0300
    1639921.8409         0.0299
-----------------------------
```

TABLE 3. The Effect of k

*** n=100 / k=5 / beta=[0 0 0 0 0] ***				*** n=100 / k=10 / beta=[0 ... 0] ***	
CASE 1: u~N(0,1)	**OLS**	**LAD**	**ratio**	**OLS**	**LAD**
Mean Bias	0.0023	-0.0006	-4.2359	-0.0009	0.0030
Median Bias	0.0033	0.0027	1.2166	-0.0062	0.0055
Variance	0.0104	0.0119	0.8757	0.0118	0.0102
M. S. Error	0.0104	0.0119	0.8762	0.0118	0.0102
CASE 2: u~t(3)	**OLS**	**LAD**	**ratio**	**OLS**	**LAD**
Mean Bias	0.0023	0.0026	0.8889	-0.0089	-0.0020
Median Bias	-0.0012	0.0053	-0.2193	-0.0091	-0.0003
Variance	0.0309	0.0139	2.2176	0.0323	0.0119
M. S. Error	0.0309	0.0139	2.2169	0.0324	0.0119
CASE 3: u~Cauchy	**OLS**	**LAD**	**ratio**	**OLS**	**LAD**
Mean Bias	-0.0759	-0.0046	16.4186	3.6157	0.0024
Median Bias	-0.0294	0.0034	-8.6046	-0.0230	0.0053
Variance	83.0287	0.0225	3693.8097	32959.4862	0.0193
M. S. Error	82.9515	0.0225	3690.5572	32939.5999	0.0193

TABLE 4. The Effect of n

*** n=1000 / k=2 / beta=[0 0] ***				*** n=10000 / k=2 / beta=[0 0] ***		
CASE 1: u~N(0,1)	**OLS**	**LAD**	**ratio**	**OLS**	**LAD**	
Mean Bias	-0.0003	0.0001	-5.1925	-0.0002	-0.0002	
Median Bias	0.0002	-0.0013	-0.1439	0.0001	-0.0000	
Variance	0.0010	0.0016	0.6226	0.0001	0.0002	(0.6274)
M. S. Error	0.0010	0.0016	0.6226	0.0001	0.0002	(0.6274)
CASE 2: u~t(3)	**OLS**	**LAD**	**ratio**	**OLS**	**LAD**	
Mean Bias	0.0002	0.0013	0.1625	-0.0004	0.0001	
Median Bias	-0.0019	0.0007	-2.5116	-0.0000	0.0005	
Variance	0.0029	0.0019	1.5239	0.0003	0.0002	(1.6553)
M. S. Error	0.0029	0.0019	1.5225	0.0003	0.0002	(1.6558)
CASE 3: u~Cauchy	**OLS**	**LAD**	**ratio**	**OLS**	**LAD**	
Mean Bias	-1.6025	-0.0001	13098.7920	0.2585	-0.0009	
Median Bias	0.0216	-0.0003	-71.0167	-0.0778	-0.0004	
Variance	6428.5969	0.0027	2406462.6118	590.8803	0.0002	
M. S. Error	6424.7362	0.0027	2407411.3199	590.3563	0.0002	

3. True or False

(i) FALSE.
It is mainly because the size of the Wald test does not remain at the level α that we do not use it when instruments are potentially weak. The power of the Wald test is high in many cases, compared to that of the LM test or the LR test.

(ii) FALSE.
It may be consistent in some special cases. Consider the model

$$y_i = x_i'\beta + z_i'\gamma + \varepsilon_i$$

where $Ex_i\varepsilon_i = 0$. When we omit z_i in the regression, we get

$$\widehat{\beta}_n = (X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'Z\gamma + (X'X)^{-1}X'\varepsilon$$

and thus

$$\widehat{\beta}_n - \beta = \underbrace{\Big(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\Big)^{-1}}_{\xrightarrow{p} Ex_i x_i'} \underbrace{\Big(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\Big)}_{\xrightarrow{p} Ex_i z_i'}\gamma + \underbrace{\Big(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\Big)^{-1}}_{\xrightarrow{p} Ex_i x_i'} \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i}_{\xrightarrow{p} Ex_i\varepsilon_i = 0}$$

Besides the trivial case that $\gamma = 0$,[2] $\widehat{\beta}_n$ is still consistent for β if $Ex_i z_i' = 0$. In other words, if the random variables x_i and z_i are orthogonal, then omitting z_i in the regression does not cause the problem.

(iii) FALSE.
The estimator obtained using suboptimal A_n in the GMM model is still consistent, as long as A_n converges in probability to some matrix. We do not require optimality of A_n when we prove the consistency of the estimator. Optimal weight matters only for the minimum of the asymptotic variance among all consistent estimators.

(iv) FALSE.
The LR statistic is not easy to implement. We need to solve both the unrestricted minimization and the restricted minimization problems. We also need to specify what c_n is, by deriving theoretically the ratio of Ω_0 to B_0. Although the formula of the LR statistic seems compact, the Wald statistic is more intuitive and thus may be regarded as compact in that sense (especially when the hypothesis is linear). The LM statistic has the most complicated one. The comparative advantage of the LR statistic (and the LM statistic) is the better size property.

(v) TRUE.

[2] This means that the true model is $y_i = x_i'\beta + \varepsilon_i$. This is not of our interest since there is no omitted regressor in this case.

Suppose $X_n = o_p(1)$. We want to show that there exists $\varepsilon > 0$ such that for any M

$$\limsup_{n\to\infty} \Pr\left(\frac{1}{|X_n|} > M\right) \geq \varepsilon$$

Take $\varepsilon = 0.5$ and any $M > 0$. Then, by definition of $o_p(1)$, there exists N such that for any $n > N$

$$\Pr\left(|X_n| < \frac{1}{M}\right) \geq \varepsilon$$

This is equivalent to

$$\Pr\left(\frac{1}{|X_n|} > M\right) \geq \varepsilon$$

Therefore $1/X_n$ is not $O_p(1)$.

Another proof is as follows. Suppose $1/X_n = O_p(1)$. We want to show that there exists $\varepsilon > 0$ such that

$$\lim_{n\to\infty} \Pr(|X_n| > \varepsilon) \neq 0$$

By definition of $O_p(1)$, there exists $M < \infty$ such that

$$\limsup_{n\to\infty} \Pr\left(\frac{1}{|X_n|} \geq M\right) < 0.5$$

Therefore

$$\liminf_{n\to\infty} \Pr\left(|X_n| > \frac{1}{M}\right) = 1 - \limsup_{n\to\infty} \Pr\left(|X_n| \leq \frac{1}{M}\right) > 0.5$$

Let $\varepsilon = 1/M$, then $\lim_{n\to\infty} \Pr(|X_n| > \varepsilon)$, if it exists, cannot be 0.

(vi) QUESTIONABLE.
First, it is not obvious whether an IQ score is indeed a relevant measure of ability affecting wage. Moreover, it is highly likely that there is an ability which is not captured by an IQ score but still affects wage. If it is correlated with education, including an IQ score as a regressor does not solve the inconsistency problem completely. In fact, it is not only the intellectual ability that affects the year of education. It would be better to use an IQ score as an instrument for education, since we would rather believe that an IQ score is highly correlated with education, but not with the other abilities in the error term. Of course, there might be objection to using it as an instrument as well.

4. Stationary Time Series

(a) Note that the non-causal AR(1) equation $X_t = \phi X_{t-1} + Z_t$ with $Z_t \sim WN(0, \sigma^2)$ and $|\phi| > 1$ has the following stationary solution.

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j} \tag{1}$$

We can take two different approaches to show that the above solution also satisfies the causal AR(1) equation $X_t = \phi^{-1} X_{t-1} + \widetilde{Z}_t$ as follows.

1. Use the above solution to obtain $\widetilde{Z}_t$ from the causal AR(1) equation and prove that $\widetilde{Z}_t$ is indeed a white noise process.
2. Obtain the solution to the causal AR(1) equation, and show that the two solutions are the same stationary time series. This requires showing that the two solutions have the same mean and autocovariance function.

Consider the 1st approach. Define $\widetilde{Z}_t := X_t - \phi^{-1}X_{t-1}$ where X_t is defined as in (1). This is indeed a white noise process. To see this, note first that

$$E\widetilde{Z}_t = EX_t - \phi^{-1}EX_{t-1} = 0$$

since $EX_t = 0$ from (1). To check the second moment, find the autocavariance function of X_t as follows. For an interger $h \geq 0$,

$$\begin{aligned}
\gamma_x(h) &= EX_tX_{t-h} \\
&= E\left[\sum_{j=1}^{\infty}\phi^{-j}Z_{t+j}\sum_{k=1}^{\infty}\phi^{-k}Z_{t-h+k}\right] \\
&= E\left[\sum_{j=1}^{\infty}\phi^{-j}Z_{t+j}\sum_{m=1-h}^{\infty}\phi^{-m-h}Z_{t+m}\right] && \leftarrow \text{change } m = k-h \\
&= \sum_{j=1}^{\infty}\phi^{-2j-h}EZ_{t+j}^2 && \leftarrow EZ_sZ_t = 0 \text{ for } s \neq t \\
&= \phi^{-h}\sum_{j=1}^{\infty}(\phi^{-2})^j\sigma^2 \\
&= \phi^{-h}\frac{\phi^{-2}\sigma^2}{1-\phi^{-2}} = \frac{\sigma^2}{\phi^h(\phi^2-1)} && (2)
\end{aligned}$$

Hence

$$\begin{aligned}
E\widetilde{Z}_t^2 &= EX_t^2 + \phi^{-2}EX_{t-1}^2 - 2\phi^{-1}EX_tX_{t-1} \\
&= \gamma_x(0) + \phi^{-2}\gamma_x(0) - 2\phi^{-1}\gamma_x(1) = \frac{\sigma^2}{\phi^2}
\end{aligned}$$

and also for $h > 0$,

$$\begin{aligned}
E\widetilde{Z}_t\widetilde{Z}_{t-h} &= EX_tX_{t-h} - \phi^{-1}EX_{t-1}X_{t-h} - \phi^{-1}EX_tX_{t-h-1} + \phi^{-2}EX_{t-1}X_{t-h-1} \\
&= \gamma_x(h) - \phi^{-1}\gamma_x(h-1) - \phi^{-1}\gamma(h+1) + \phi^{-2}\gamma_x(h) \\
&= \frac{\sigma^2}{\phi^h(\phi^2-1)} - \frac{\sigma^2}{\phi^h(\phi^2-1)} - \frac{\sigma^2}{\phi^{h+2}(\phi^2-1)} + \frac{\sigma^2}{\phi^{h+2}(\phi^2-1)} = 0
\end{aligned}$$

as desired. Defining $\widetilde{\sigma}^2 := \phi^{-2}\sigma^2$ completes the proof.

Now consider the 2nd approach. The causal AR(1) equation has the following solution.

$$X_t = \sum_{j=0}^{\infty} (\phi^{-1})^j \widetilde{Z}_{t-j} \tag{3}$$

Now it suffices to show that this also has mean 0 and autocovariance function as in (2). It is easy to check that

$$EX_t = \sum_{j=0}^{\infty} (\phi^{-1})^j E\widetilde{Z}_{t-j} = 0$$

The autocovariance function is hence defined as follows. For an integer $h \geq 0$,

$$\begin{aligned}
\gamma_x(h) &= EX_t X_{t+h} \\
&= E\left[\sum_{j=0}^{\infty} (\phi^{-1})^j \widetilde{Z}_{t-j} \sum_{k=0}^{\infty} (\phi^{-1})^k \widetilde{Z}_{t+h-k}\right] \\
&= E\left[\sum_{j=0}^{\infty} (\phi^{-1})^j \widetilde{Z}_{t-j} \sum_{m=-h}^{\infty} (\phi^{-1})^{h+m} \widetilde{Z}_{t-m}\right] && \leftarrow \text{change } m = k - h \\
&= \sum_{j=0}^{\infty} (\phi^{-1})^{h+2j} E\widetilde{Z}_{t-j}^2 && \leftarrow E\widetilde{Z}_s \widetilde{Z}_t = 0 \text{ for } s \neq t \\
&= \phi^{-h} \sum_{j=0}^{\infty} (\phi^{-2})^j \widetilde{\sigma}^2 \\
&= \phi^{-h} \frac{\widetilde{\sigma}^2}{1 - \phi^{-2}} = \frac{\widetilde{\sigma}^2}{\phi^{h-2}(\phi^2 - 1)} && (4)
\end{aligned}$$

Note that the two autocovariance functions obtained in (2) and (4) are the same when $\widetilde{\sigma}^2 = \phi^{-2}\sigma^2$, which completes the proof.

(b) Let $\phi_x(\cdot)$ and $\phi_y(\cdot)$ be polynomials of order p. Then two independent AR(p) processes X_t and Y_t can be defined as

$$\begin{aligned}
\phi_x(L) X_t &= \varepsilon_{xt} \\
\phi_y(L) Y_t &= \varepsilon_{yt}
\end{aligned}$$

where $\varepsilon_{xt} \sim WN(0, \sigma_x^2)$ and $\varepsilon_{yt} \sim WN(0, \sigma_y^2)$ are independent white noise processes. Define $Z_t := X_t + Y_t$, then,

$$\phi_x(L)\phi_y(L) Z_t = \phi_x(L)\phi_y(L) X_t + \phi_x(L)\phi_y(L) Y_t$$

So

$$\phi_x(L)\phi_y(L) Z_t = \phi_y(L)\varepsilon_{xt} + \phi_x(L)\varepsilon_{yt} \tag{5}$$

If we can write this as

$$\phi_z(L) Z_t = \theta(L)\varepsilon_{zt}$$

with $\phi_z(\cdot)$ a polynomial of order $2p$, $\theta(\cdot)$ a polynomial of order p, and ε_{zt} a white noise process, the proof completes. First, define $\phi_z(z) := \phi_x(z)\phi_y(z)$, then it is a polynomial of order $2p$, so the LHS of (5) is an AR($2p$) representation. Next, we want to find ε_{zt} and $\theta(\cdot)$ such that $\theta(L)\varepsilon_{zt} = \phi_y(L)\varepsilon_{xt} + \phi_x(L)\varepsilon_{yt}$. The equality means that both time series have the same mean and autocovariance function.[3] It is immediately verified that they have the same mean. The condition on autocovariance function defines the following system of $p+1$ equations.

$$\begin{aligned} var(\theta(L)\varepsilon_{zt}) &= var(\phi_y(L)\varepsilon_{xt} + \phi_x(L)\varepsilon_{yt}) \\ cov(\theta(L)\varepsilon_{zt}, \theta(L)\varepsilon_{zt-1}) &= cov(\phi_y(L)\varepsilon_{xt}, \phi_y(L)\varepsilon_{xt-1}) + cov(\phi_x(L)\varepsilon_{yt}, \phi_x(L)\varepsilon_{yt-1}) \\ &\vdots \\ cov(\theta(L)\varepsilon_{zt}, \theta(L)\varepsilon_{zt-p}) &= cov(\phi_y(L)\varepsilon_{xt}, \phi_y(L)\varepsilon_{xt-p}) + cov(\phi_x(L)\varepsilon_{yt}, \phi_x(L)\varepsilon_{yt-p}) \end{aligned}$$

The covariance between the terms with more than p lags is 0, since $\phi_y(L)\varepsilon_{xt}$ and $\phi_x(L)\varepsilon_{yt}$ are independent MA(p) processes. Writing the equations explicitly,

$$\begin{aligned} (1+\theta_1^2+\cdots+\theta_p^2)\sigma_z^2 &= (1+\phi_{y1}^2+\cdots\phi_{yp}^2)\sigma_x^2 + (1+\phi_{x1}^2+\cdots\phi_{xp}^2)\sigma_y^2 \\ (\theta_1+\theta_1\theta_2+\cdots+\theta_{p-1}\theta_p)\sigma_z^2 &= (\phi_{y1}+\phi_{y1}\phi_{y2}+\cdots+\phi_{yp-1}\phi_{yp})\sigma_x^2 \\ &\quad + (\phi_{x1}+\phi_{x1}\phi_{x2}+\cdots+\phi_{xp-1}\phi_{xp})\sigma_y^2 \\ &\vdots \\ \theta_p\sigma_z^2 &= \phi_{yp}\sigma_x^2 + \phi_{xp}\sigma_y^2 \end{aligned}$$

There are $p+1$ equations and $p+1$ unknowns $\theta_1, \cdots, \theta_p$ and σ_z^2, so we can find such $\theta(\cdot)$ and ε_{zt}. This implies that the RHS of (5) is an MA(p) representation. Therefore the equation (5) actually defines an ARMA($2p, p$) process.

The sum of two independent MA(p) processes is an MA(p) process. This is already proven in the above argument, since the choice of $\phi_x(\cdot)$, $\phi_y(\cdot)$, ε_{xt} and ε_{yt} are arbitrary.

[3] Another way to claim the equality is using a similar argument as in method 1 of (a). This requires definition of $\theta(\cdot)$, inversion of it and proof of ε_{zt} being a white noise process. Invertibility of an MA process is not covered, so we skip this method.