

## Matlab computer lab 2025 of spatial panels with common factors

Open Matlab version 2023a and store all files from the zip-file “Matlabcode2025spatialpanels\_CF” in the working directory. Then run the m-file “Course2025cigaretteCF”.

You will see that it generates the estimation results of different spatial econometric models for three spatial weight matrices: binary contiguity matrix, inverse distance matrix, and a parameterized inverse distance matrix (parameterized with factor 3). Next, it carries out a Bayesian comparison approach for static and dynamic spatial econometric models. Finally, the model is estimated as a dynamic spatial Durbin model with spatial and time-period fixed effects, using an inverse distance matrix as  $W$ , as well as with spatial fixed effects but time fixed effects replaced by cross-sectional averages of  $y(t)$  and  $y(t-1)$  and replaced by two principal components.

Just in case Matlab does not work, consult the notepad-file “Ouputfile\_Course2025\_results\_notepad” for the output.

### Questions

1. Consider the results of eight models generated for the binary contiguity matrix. Identify which type of spatial econometric model (eight models in total) has been estimated.
2. An important issue in spatial econometric models is the ratio between the spatial spillover effects and the direct effects. Calculate this ratio for the price variable in the SEM, SDEM and SDM models with  $W=WBC$ . Which patterns do you observe? What is your conclusion about these patterns from an empirical point of view?
3. Carry out manually Likelihood Ratio tests to see whether the results of the GNS model are significantly better than those of SDM and of SDEM ( $W=WBC$ ).
4. Write down the spatial interaction effects of the  $WY$  variable and of  $W*$ disturbance term of the GNS model ( $W=WBC$ ). Do these values make sense?

5. Which model according to the Bayesian comparison approach based on the binary contiguity matrix turns out to be most likely. Does this combination of model and  $W$  makes sense?
6. Which model and which  $W$  according to the Bayesian comparison approach turns out to be most likely. Does this combination of model and  $W$  makes sense? Answer this question both for static and dynamic versions of the cigarette demand model.
7. In a next step cigarette demand is explained by a dynamic spatial Durbin model with spatial and time-period fixed effects, using an inverse distance matrix as  $W$ . Consider the R-squared, log-likelihood function value, the CD test statistics, and the direct and indirect effects estimates to judge whether this model is an improvement, both from a statistical and an economic-theoretical viewpoint.
8. Next, cigarette demand is explained by a dynamic spatial Durbin model with two common factors in the form of cross-sectional averages, using an inverse distance matrix as  $W$ . Again, judge whether this model is an improvement, both from a statistical and an economic-theoretical viewpoint.
9. Finally, cigarette demand is explained by a dynamic spatial Durbin model with two common factors in the form of principal components, using an inverse distance matrix as  $W$ . Again, judge whether this model is an improvement, both from a statistical and an economic-theoretical viewpoint.

Answers:

1. 1=OLS, 2=SAR, 3=SEM, 4=SLX, 5=SDM, 6=SDEM, 7=SAC, 8=GNS
2. No direct and indirect effects estimates are reported for OLS, SEM, SLX and SDEM, since they follow from the coefficient estimates immediately. The direct effects are simply the coefficients estimates of the variables logp and logy. The indirect effects in OLS and SEM are zero by definition. The indirect effects in SLX and SDEM are the coefficient estimates of wlogp and wlogy. Ratio of indirect effect/direct effect for logp:

SEM:  $0/(-1.004579)=0$ .

SDEM:  $-0.176484/(-1.011159)=0.1745$ .

SDM:  $-0.2178/(-1.0140)=0.2148$ .

In SEM spillovers are zero by definition, which is unlikely from an empirical viewpoint.

In SDM and SDEM they can take any value. They are often close to each other in both models, which to some extent is worrying since spillovers in SDM are global and in SDEM are local.

3. GNS vs. SDM:  $-2*(1691.3787-1695.1232)=7.4890$ .

GNS vs. SDEM:  $-2*(1691.2000-1695.1232)=7.8464$ .

GNS contains one parameter more than SDM or than SDEM. This LR-ratio test statistic follows a chi-squared distribution with in this case 1 degree of freedom. 95% critical value is 3.84. So statistically, GNS outperforms SDM and SDEM.

4. Parameter estimate of the WY variable in GNS amounts to -0.481451 and of W\*disturbance term to 0.627830. Both are significant but opposite of sign. It seems that they blow up each other due to a kind of multicollinearity (overfitting), since if one of them is dropped smaller values are obtained: parameter estimate of the WY variable in SDM amounts to 0.267257 and of W\*disturbance term in SDEM to 0.268588. So although GNS outperforms SDM and SDEM statistically, GNS does not seem to make sense.
5. SDM with probability of 0.4591. SDM points to global spillovers. This goes together with a sparse W matrix, such as the binary contiguity matrix.
6. Static: SDEM with the parameterized inverse distance matrix (parameterized with factor 3) produces the highest probability of 1. SDEM points to local spillovers, which makes more sense from an economic-theoretical viewpoint. A local spillover model goes

together with a denser W matrix, since spillovers although small may occur at distant states.

Dynamic: SDEM with the inverse distance matrix produces the highest probability of 0.5247. Again the results point to SDEM, but a regular inverse distance matrix now appears to make more sense.

7. R-squared=0.9779, LogL=2624.7294, CD statistic raw data=101.519 points to common factors, CD statistic of residuals=-0.0064. Statistically, the model improved. Direct effects of price and income are significant and respectively negative and positive, both in the short and long-term, which are in line with standard economic-theory. Both short and long term indirect effect of price, although insignificant, are negative, which implies that no evidence is found in favor of bootlegging effect. This is disappointing.
8. Although LogL=3100.6 increased, R-squared is lower. Biggest problem is that this model does not converge, i.e., coefficients of  $y(t-1)$ ,  $Wy(t-1)$  and  $Wy(t)=0.8759+0.2803+0.0012>1$ . Due to this non-stability model has to be rejected.
9. R-squared=0.9837, LogL=3307.7, CD statistic raw data=101.519 points to common factors, CD statistic of residuals=-3.202. In terms of R-squared and LogL this model outperforms the dynamic spatial panel Durbin model with spatial and time-period fixed effects. The CD statistic however is still outside the interval  $[-1.96,+1.96]$ . Direct effects of price and income are again significant and respectively negative and positive, both in the short and long-term, which are in line with standard economic-theory. Importantly, we now also find evidence in favor of the bootlegging effect: the price spillover effect is positive and significant both in the short and in the long term. From an economic-theoretical viewpoint, this model is therefore to be preferred.