I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15815

Robust Dynamic Space-Time Panel Data Models Using ε-Contamination: An Application to Crop Yields and Climate Change

Badi H. Baltagi
Georges Bresson
Anoop Chaturvedi
Guy Lacroix

DECEMBER 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15815

Robust Dynamic Space-Time Panel Data Models Using ε-Contamination: An Application to Crop Yields and Climate Change

Badi H. Baltagi
Syracuse University and IZA

Anoop Chaturvedi
University of Allahabad

Georges Bresson
Universite Paris II

Guy Lacroix
Universite Laval and IZA

DECEMBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Robust Dynamic Space-Time Panel Data Models Using ε-Contamination: An Application to Crop Yields and Climate Change*

This paper extends the Baltagi et al. (2018, 2021) static and dynamic ε-contamination papers to dynamic space-time models. We investigate the robustness of Bayesian panel data models to possible misspecification of the prior distribution. The proposed robust Bayesian approach departs from the standard Bayesian framework in two ways. First, we consider the ε-contamination class of prior distributions for the model parameters as well as for the individual effects. Second, both the base elicited priors and the ε-contamination priors use Zellner (1986)'s g-priors for the variance-covariance matrices. We propose a general "toolbox" for a wide range of specifications which includes the dynamic space-time panel model with random effects, with cross-correlated effects à la Chamberlain, for the Hausman-Taylor world and for dynamic panel data models with homogeneous/heterogeneous slopes and cross-sectional dependence. Using an extensive Monte Carlo simulation study, we compare the finite sample properties of our proposed estimator to those of standard classical estimators. We illustrate our robust Bayesian estimator using the same data as in Keane and Neal (2020). We obtain short run as well as long run effects of climate change on corn producers in the United States.

JEL Classification: C11, C23, C26, Q15, Q54

Keywords: climate change, crop yields, dynamic model, ε-contamination, panel data, robust Bayesian estimator, space-time

Corresponding author:
Badi H. Baltagi
Department of Economics and Center for Policy Research
Syracuse University
426 Eggers Hall
Syracuse
NY 13244-1020
USA
E-mail: bbaltagi@syr.edu

* This paper is written in honor of Peter Schmidt for his many contributions to econometrics. In particular, his influential contributions to dynamic panel data models. We are grateful to Robin Sickles, Subal C. Kumbhakar and to two anonymous referees for their helpful comments and suggestions. The usual disclaimers apply.

1. Introduction

The space-time panel data models provide a general structure which accommodates feedback from lagged endogenous values, *i.e.*, state dependence, along with the spatial spillovers, spatial heterogeneity as well as interactive effects. Yu et al. (2008) have introduced a dynamic space-time panel specification with fixed effects, where both N (number of spatial sites or individuals) and T (number of time points) are large, and which allows one to treat spatial dependence in the dependent variable vector (see also Lee and Yu (2015)). They propose a concentrated quasi-maximum likelihood (QML) estimation and a bias correction for the estimators. They show that when T grows faster than $N^{1/3}$, the correction asymptotically eliminates the bias. Su and Yang (2015) propose a QML estimation of dynamic panel models with spatial errors for short panels (N large, T fixed), both for random effects and fixed effects worlds. They propose a residual-based bootstrap method for estimating the standard errors. This approach yields good results in finite samples only if the assumptions about the initial observations are satisfied.

As is well known, the ordinary least squares estimation of a spatial dynamic panel data model generally yields inconsistent parameter estimates due to the potential correlation between the spatially lagged dependent variables and the error term. Recently, Jin et al. (2020) proposed an efficient distribution-free least squares estimation method that utilizes the eigen decomposition of a weight matrix. They also present a penalized model selection procedure based on the proposed method. Their approach is very powerful compared to the well-known instrumental variable techniques and its applicability is demonstrated *via* a high-dimensional data example. Unfortunately, when N or T is small, their estimator is seriously biased even when using their proposed bias-corrected estimator.[1]

Parent and LeSage (2010) considered a dynamic space-time panel specification with random effects and proposed a Bayesian Markov Chain Monte Carlo (MCMC) method. They used a restriction on the parameter associated with spatial effects from the previous period, δ in equation (1) below, and showed that the restriction allows one to separate the space and time dimensions. This greatly simplifies the computation of the space-time covariance structure as well as the own- and cross-partial derivatives of the model (see also Parent and LeSage (2011)). Debarsy et al. (2012) considered ~~in~~ a dynamic space-time Durbin model with random effects. They remove the restriction on δ and, following Parent and LeSage (2010), use the same MCMC method where all parameters are *a priori* independent. LeSage et al. (2019) also proposed a dynamic space-time panel data model without individual-specific effects. Considering proper priors for the parameters and assuming that the joint distribution of these parameters is uniformly distributed, they adopt Metropolis Hastings steps and a reversible jump procedure for some number of initial MCMC draws to produce proposal values for the vector of parameters.

The present paper develops a general framework for robust Bayesian analysis of dynamic space-time panel data models using ε-contamination class of prior distributions. Bayesian inference procedures based on a single base prior distribution ignore the fact that a prior distribution in the neighborhood of this base prior may also represent the prior belief of the experimenter. Robust Bayes inference procedures based on a class of prior distributions usually perform better and are more robust in the sense that if the available base prior is irrelevant, the procedure

[1] We thank Yuehua Wu for the helpful discussions on this issue. Unfortunately, their bias correction method is ineffective when $T < 50$ irrespective of N.

articulately discards the base prior in favor of the sample information. The ε-contamination class of prior distributions, which is a mixture of a base prior and a contamination class of priors, is an attractive class of prior distributions. For a more authoritative discussion, one may refer to Berger (1985), Berger and Berliner (1986), Chaturvedi (1996), Baltagi et al. (2018) and the references cited therein. For selecting a specific prior distribution from the contamination class of priors, Berger and Berliner (1986) considered the type-II maximum likelihood (ML-II) procedure. ML-II was named and extensively studied in Good (1965), and it can be seen as a particular instance of empirical Bayes which, in general, "estimates" the hyperparameters from the data.[2] Section 2 presents the robust dynamic space-time panel data models. Section 3 derives the Type-II maximum likelihood posterior mean and the variance-covariance matrix of the coefficients utilizing a two-stage hierarchy approach. The finite sample performance of the proposed robust Bayes estimator is investigated in Section 4 using extensive Monte Carlo experiments. In Section 5, we use the same data as in Keane and Neal (2020) to illustrate our robust Bayesian estimator applied to a dynamic space-time mixed specification model of crop yields and climate change. One of the main benefits of such dynamic space-time mixed model is its ability to estimate short-run and long-run effects through the impact multiplier (weather) and the τ-period-ahead dynamic multiplier (climate) of a permanent change in the temperature or precipitations at time t. Finally, section 6 gives some concluding remarks.

2. A robust dynamic space-time panel data model

Let us start with the Gaussian dynamic space-time mixed model:

$$\begin{aligned} y_{ti} &= \phi y_{t-1,i} + \rho \sum_{j=1}^{N} w_{ij} y_{tj} + \delta \sum_{j=1}^{N} w_{ij} y_{t-1,j} + X'_{ti}\beta + D'_{ti} b_i + u_{ti} \ , \ i = 1, ..., N \ , \ t = 2, ..., T, \\ &= Z'_{ti}\theta + D'_{ti} b_i + u_{ti} \qquad (1) \\ &\quad \text{with } Z'_{ti} = \left[y_{t-1,i}, \sum_{j=1}^{N} w_{ij} y_{tj}, \sum_{j=1}^{N} w_{ij} y_{t-1,j}, X'_{ti} \right] \text{ and } \theta' = [\phi, \rho, \delta, \beta']', \end{aligned}$$

where the data is ordered in matrix form such that i is a faster index than t, X'_{ti} is a $(1 \times K_x)$ vector of explanatory variables including the intercept, and β is a $(K_x \times 1)$ vector of parameters. Let D'_{ti} denote a $(1 \times k_2)$ vector of covariates and b_i a $(k_2 \times 1)$ vector of parameters. The subscript i of b_i indicates that the model allows for heterogeneity on the D variables. u_{ti} is a remainder term assumed to be normally distributed, *i.e.* $u_{ti} \sim N\left(0, \tau^{-1}\right)$. The distribution of u_{ti} is parametrized in terms of its precision τ rather than its variance $\sigma_u^2 \, (= 1/\tau)$. The $W_N = (w_{ij})$ is a $(N \times N)$ spatial weights matrix whose diagonal elements are zero. Moreover, we also assume that W_N is row-normalized and that all eigenvalues are real and less than or equal to one. Connectivity between the N individuals is represented by the W_N spatial weights matrix. The distance between individuals i and j may be based on geography or some measure of economic distance, or defined as rook-style or queen-style contiguities, or as the k-nearest neighbors for instance.

[2] *"We consider the most commonly used method of selecting a hopefully robust prior in* Γ *(the* ε*-contamination class of prior distributions), namely choice of that prior* π *which maximizes the marginal likelihood* $m(y|\pi)$ *over* Γ*. This process is called Type II maximum likelihood by Good (1965)" (Berger and Berliner (1986) page 463).*

Pooling the N individuals for one time period, we can write

$$\begin{aligned} y_t &= \phi y_{t-1} + \rho W_N y_t + \delta W_N y_{t-1} + X_t\beta + D_t b + u_t \ , \ t = 2, ..., T, \\ &= Z_t\theta + D_t b + u_t \text{ with } Z_t = [y_{t-1}, W_N y_t, W_N y_{t-1}, X_t] \end{aligned} \tag{2}$$

where y_t is the N-dimensional vector of the dependent variable, y_{t-1} its lagged value, X_t the $(N \times K_x)$ matrix of covariates, D_t the $(N \times K_2)$ (with $K_2 = Nk_2$) matrix of other covariates,

$$D_t = diag\left(D'_{ti}\right)_{i=1,...,N} \text{ and } b = \left(b'_1, b'_2, \ldots, b'_N\right)', \tag{3}$$

β and b are $(K_x \times 1)$ and $(K_2 \times 1)$ vectors of coefficients associated with the covariates X_t and D_t. ϕ is the autoregressive time dependence parameter, ρ is the spatial dependence parameter and δ is the spatio-temporal diffusion parameter.[3] In order to ensure stable dynamic estimation, Yu et al. (2008), Parent and LeSage (2011) or LeSage et al. (2019) show that stationary conditions are satisfied if:[4]

$$\begin{cases} \phi + (\rho + \delta)\,\varpi_{\max} < 1 & \text{if} \quad \rho + \delta \geq 0 \\ \phi + (\rho + \delta)\,\varpi_{\min} < 1 & \text{if} \quad \rho + \delta < 0 \\ \phi - (\rho - \delta)\,\varpi_{\max} > -1 & \text{if} \quad \rho - \delta \geq 0 \\ \phi - (\rho - \delta)\,\varpi_{\min} > -1 & \text{if} \quad \rho - \delta < 0 \end{cases} \tag{4}$$

where $\varpi_{\min}$ and $\varpi_{\max}$ are the minimum and maximum eigenvalues of the spatial weights matrix W_N. Pooling the $T-1$ time periods, we get the dual form of the model:

$$y = \phi y_{-1} + \rho y^* + \delta y^*_{-1} + X\beta + Db + u = Z\theta + Db + u, \tag{5}$$

where y_{-1} is the $(T-1)N$-dimensional vector of the lagged dependent variable, y^* is the $(T-1)N$-dimensional vector of the spatially weighted dependent variable: $y^* = \left(y^*_{2,1}, ..., y^*_{2,N}, ..., y^*_{T,1}, ..., y^*_{T,N}\right)'$ with $y^*_{ti} = \sum_{j=1}^{N} w_{ij} y_{tj}$. y^*_{-1} is the $(T-1)N$-dimensional vector of the spatially weighted lagged dependent variable: $y^*_{-1} = \left(y^*_{1,1}, ..., y^*_{1,N}, ..., y^*_{T-1,1}, ..., y^*_{T-1,N}\right)'$ with $y^*_{t-1,i} = \sum_{j=1}^{N} w_{ij} y_{t-1,i}$ and θ is a K_1-vector of parameters with $K_1 = K_x + 3$.

In a Bayesian framework, it is customary to constrain the priors of the space-time parameters ϕ, ρ and δ over the stationary interval as in equation (4) and to use products of independent uniform distributions or mixtures of uniform distributions (see section A of the supplementary material for more discussion). In a non-spatial framework, much has been written about the desirability of imposing stationarity conditions. The choice of particular prior distributions that allow one to develop the posterior analysis of autoregressive models with (or without) the stationarity has also been much discussed in the literature (Phillips, 1991). However, most papers use uninformative (objective) priors and do not consider stationarity issues. As there is no clear consensus on these topics in the literature, we do not impose any particular constraints on the priors of the dependent parameters.

[3] Parent and LeSage (2010, 2011) use a restriction on $\delta(= -\rho \times \phi)$ allowing space and time to be separable.

[4] Yu et al. (2008) observed that y_t can have some nonstationary components if $\phi + \rho + \delta = 1$ but, as underlined by Parent and LeSage (2011), stationarity does not require that $|\phi| + |\rho| + |\delta| < 1$. LeSage et al. (2019) recall that the dependence parameters ϕ, ρ and δ associated with stable processes require $\phi + \rho + \delta < 1$ and, for cases where $\rho - \delta > 0$, it requires that $\phi - \rho + \delta > -1$. See also Parent and LeSage (2011).

Extending the Baltagi et al. (2018, 2021) non-spatial ε-contamination papers to the dynamic space-time model, we assume a Zellner g-prior, for the $\theta\big(=[\phi,\rho,\delta,\beta']'\big)$ vector encompassing all the coefficients of the covariates Z. In other words, we propose a very general two-stage hierarchy framework:

$$\begin{array}{rl} \text{First stage :} & y = Z\theta + Db + u,\ u \sim N(0,\Sigma), \Sigma = \tau^{-1} I_{(T-1)N} \\ \text{Second stage :} & \theta \sim N\left(\theta_0, \Lambda_\theta\right) \text{ and } b \sim N\left(b_0, \Lambda_b\right) \\ \text{with} & p(\tau) \propto \tau^{-1}\ ,\ \Lambda_\theta = (\tau g Z'Z)^{-1} \text{ and } \Lambda_b = (\tau h D'D)^{-1}\ . \end{array} \tag{6}$$

The second stage (also called *fixed effects model* in the Bayesian literature) updates the distribution of the parameters. Rather than specifying a Wishart distribution for the variance-covariance matrices as is customary, Zellner's g-prior ($\Lambda_\theta = (\tau g Z'Z)^{-1}$ for θ or $\Lambda_b = (\tau h D'D)^{-1}$ for b) has been widely adopted because of its computational efficiency in evaluating marginal likelihoods and because of its simple interpretation as arising from the design matrix of observables in the sample. Since the calculation of marginal likelihoods using a mixture of g-priors (resp. h-priors) involves only a one-dimensional integral, this approach provides an attractive computational solution that made the original g-priors popular while insuring robustness to misspecification of g (resp. h) (see Zellner (1986) and Fernández et al. (2001) to mention a few).[5] Since the calculation of marginal likelihoods using a mixture of g-priors involves only a one-dimensional integral, this approach provides an attractive computational solution that made the original g-priors popular while insuring robustness to misspecification of g (see Zellner (1986) and Fernández et al. (2001) to mention a few).

To guard against misspecifying the distributions of the priors, many suggest considering classes of priors (θ, b, τ) (see Berger (1985), Baltagi et al. (2018, 2021)). Here, we consider the ε-contamination class of prior distributions for (θ, b, τ):

$$\Gamma = \left\{\pi(\theta, b, \tau|g_0, h_0) = (1-\varepsilon)\,\pi_0(\theta, b, \tau|g_0, h_0) + \varepsilon q(\theta, b, \tau|g_0, h_0)\right\}. \tag{7}$$

$\pi_0(\cdot)$ is the base elicited prior, $q(\cdot)$ is the contamination belonging to some suitable class Q of prior distributions, $0 \leq \varepsilon \leq 1$ is given and reflects the amount of error in $\pi_0(\cdot)$. The precision τ is assumed to have a vague prior, $p(\tau) \propto \tau^{-1}$, $0 < \tau < \infty$, and $\pi_0(\theta, b, \tau|g_0, h_0)$ is the base prior assumed to be a specific g-prior with

$$\left\{\begin{array}{rcl} \theta & \sim & N\left(\theta_0 \iota_{K_1}, (\tau g_0 \Lambda_Z)^{-1}\right) \text{ with } \Lambda_Z = Z'Z \\ b & \sim & N\left(b_0 \iota_{NK_2}, (\tau h_0 \Lambda_D)^{-1}\right) \text{ with } \Lambda_D = D'D, \end{array}\right. \tag{8}$$

where ι_{K_1} is a $(K_1 \times 1)$ vector of ones. Furthermore, θ_0, b_0, g_0 and h_0 are known scalar hyper-parameters of the base prior $\pi_0(\theta, b, \tau|g_0, h_0)$. The probability density function (henceforth pdf) of the base prior $\pi_0(.)$ is given by:

$$\pi_0(\theta, b, \tau|g_0, h_0) = p(\theta|b, \tau, \theta_0, b_0, g_0, h_0) \times p(b|\tau, b_0, h_0) \times p(\tau). \tag{9}$$

[5] The literature generally recommends using the unit information prior (UIP) to set the g-priors (see section 4.1).

The possible class of contamination Q is defined as:

$$Q = \left\{ \begin{array}{c} q\left(\theta, b, \tau | g_0, h_0\right) = p\left(\theta | b, \tau, \theta_q, b_q, g_q, h_q\right) \times p\left(b | \tau, b_q, h_q\right) \times p\left(\tau\right) \\ \text{with } 0 < g_q \leq g_0,\, 0 < h_q \leq h_0 \end{array} \right\}, \tag{10}$$

with

$$\left\{ \begin{array}{rcl} \theta & \sim & N\left(\theta_q \iota_{K_1}, \left(\tau g_q \Lambda_Z\right)^{-1}\right) \\ b & \sim & N\left(b_q \iota_{NK_2}, \left(\tau h_q \Lambda_D\right)^{-1}\right), \end{array} \right. \tag{11}$$

where θ_q, b_q, g_q and h_q are unknown. The restrictions $g_q \leq g_0$ and $h_q \leq h_0$ imply that the base prior is the best possible so that the precision of the base prior is greater than any prior belonging to the contamination class. The ε-contamination class of prior distributions for (θ, b, τ) is then conditional on known g_0 and h_0.

Following Baltagi et al. (2018, 2021), we use a two-step strategy because it simplifies the derivation of the predictive densities (or marginal likelihoods):[6]

1. Let $y^* = (y - Db)$. Derive the conditional ML-II posterior distribution of θ given the specific effects b.
2. Let $\widetilde{y} = (y - Z\theta)$. Derive the conditional ML-II posterior distribution of b given the coefficients θ.

We condition the likelihood on the first period observation of y_1 and consider the latter as exogenous and known. As stressed above, and in line with most of the literature, we do not impose stationarity constraints. Likewise, we adhere to the philosophy of the ε-contamination class approach and use data-driven priors.

3. The robust dynamic space-time model in the two-stage hierarchy

The marginal likelihoods (or predictive densities) corresponding to the base priors are:

$$m\left(y^* | \pi_0, b, g_0\right) = \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\theta, \tau | g_0\right) \times p\left(y^* | Z, b, \tau\right)\ d\theta\ d\tau,$$

where K_1 is the dimension of θ. Further

$$m\left(\widetilde{y} | \pi_0, \theta, h_0\right) = \int_0^\infty \int_{\mathbb{R}^{NK_2}} \pi_0\left(b, \tau | h_0\right) \times p\left(\widetilde{y} | D, \theta, \tau\right)\ db\ d\tau,$$

where K_2 is the dimension of b and

$$\begin{array}{rcl} \pi_0\left(\theta, \tau | g_0\right) & = & \left(\dfrac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \tau^{-1} \left|\Lambda_Z\right|^{1/2} \exp\left(-\dfrac{\tau g_0}{2}\left(\theta - \theta_0 \iota_{K_1}\right)' \Lambda_Z (\theta - \theta_0 \iota_{K_1})\right)\right), \\ \pi_0\left(b, \tau | h_0\right) & = & \left(\dfrac{\tau h_0}{2\pi}\right)^{\frac{NK_2}{2}} \tau^{-1} \left|\Lambda_D\right|^{1/2} \exp\left(-\dfrac{\tau h_0}{2}(b - b_0 \iota_{NK_2})' \Lambda_D (b - b_0 \iota_{NK_2})\right). \end{array}$$

[6] A one-step estimation of the ML-II posterior distribution is possible but hardly feasible. This is because the probability density functions of y and that of the base prior $\pi_0\left(\theta, b, \tau | g_0, h_0\right)$ need to be combined to get the predictive density. The resulting expression is highly complex and its integration with respect to (θ, b, τ) is quite involved.

Solving these equations is considerably easier than solving the equivalent expression corresponding to a one-step approach.

For the first step of the robust Bayesian estimator ($y^* = y - Db$), combining the pdf of y^* and the pdf of the base prior allows one to get the predictive density $m(y^*|\pi_0, b, g_0)$ corresponding to the base prior.[7] Likewise, we can obtain the predictive density $m(y^*|q, b, g_0)$ corresponding to the contaminated prior for the distribution $q(\theta, \tau|g_0, h_0) \in Q$ from the class Q of possible contamination distributions. As the ε-contamination of the prior distributions for (θ, τ) is defined by $\pi(\theta, \tau|g_0) = (1-\varepsilon)\pi_0(\theta, \tau|g_0) + \varepsilon q(\theta, \tau|g_0)$, the corresponding predictive density is given by:

$$m(y^*|\pi, b, g_0) = (1-\varepsilon) m(y^*|\pi_0, b, g_0) + \varepsilon m(y^*|q, b, g_0).$$

Let $\pi_0^*(\theta, \tau|g_0)$ denote the posterior density of (θ, τ) based upon the prior $\pi_0(\theta, \tau|g_0)$. Let $q^*(\theta, \tau|g_0)$ denote the posterior density of (θ, τ) based upon the prior $q(\theta, \tau|g_0)$. Then the ML-II posterior density of (θ, τ) is given by

$$\begin{aligned}
\widehat{\pi}^*(\theta, \tau|g_0) &= \frac{p(y^*|X, b, \tau)\,\widehat{\pi}(\theta, \tau \mid g_0)}{\int_0^\infty \int_{\mathbb{R}^{K_1}} p(y^*|X, b, \tau)\,\widehat{\pi}(\theta, \tau|g_0)\; d\theta\; d\tau} \\
&= \widehat{\lambda}_\theta \left(\frac{p(y^*|X, b, \tau)\,\pi_0(\theta, \tau|g_0)}{m(y^* \mid \pi_0, b, g_0)} \right) + \left(1 - \widehat{\lambda}_\theta\right) \left(\frac{p(y^*|X, b, \tau)\,\widehat{q}(\theta, \tau|g_0)}{m(y^*|\widehat{q}, b, g_0)} \right),
\end{aligned}$$

with

$$\widehat{\lambda}_{\theta, g_0} = \left[1 + \frac{\varepsilon m(y^*|\widehat{q}, b, g_0)}{(1-\varepsilon) m(y^*|\pi_0, b, g_0)} \right].$$

and $m(y^*|\widehat{q}, b, g_0) = \sup_{q \in Q} m(y^*|q, b, g_0)$. Integration of $\widehat{\pi}^*(\theta, \tau|g_0)$ with respect to τ leads to the marginal ML-II posterior density of θ :

$$\begin{aligned}
\widehat{\pi}^*(\theta|g_0) &= \int_0^\infty \widehat{\pi}^*(\theta, \tau|g_0)\, d\tau = \widehat{\lambda}_{\theta, g_0} \int_0^\infty \pi_0^*(\theta, \tau|g_0)\, d\tau + \left(1 - \widehat{\lambda}_{\theta, g_0}\right) \int_0^\infty q^*(\theta, \tau|g_0)\, d\tau \\
&= \widehat{\lambda}_{\theta, g_0} \pi_0^*(\theta|g_0) + \left(1 - \widehat{\lambda}_{\theta, g_0}\right) \widehat{q}^*(\theta|g_0), \qquad (12)
\end{aligned}$$

where $\pi_0^*(\theta|g_0)$ is the pdf of a multivariate t-distribution where the mean vector $\theta_*(b|g_0)$ is the Bayes estimate of θ for the prior distribution $\pi_0(\theta, \tau)$. $\widehat{q}^*(\theta|g_0)$ is the pdf of a multivariate t-distribution where the mean vector $\widehat{\theta}_{EB}(b|g_0)$ is the empirical Bayes estimator of θ for the contaminated prior distribution $q(\theta, \tau)$ (see section B of the supplementary material). The mean of the ML-II posterior density of θ is then:

$$\begin{aligned}
\widehat{\theta}_{ML-II} &= \widehat{\lambda}_{\theta, g_0} E[\pi_0^*(\theta|g_0)] + \left(1 - \widehat{\lambda}_{\theta, g_0}\right) E[\widehat{q}^*(\theta|g_0)] \qquad (13) \\
&= \widehat{\lambda}_{\theta, g_0} \theta_*(b|g_0) + \left(1 - \widehat{\lambda}_{\theta, g_0}\right) \widehat{\theta}_{EB}(b|g_0).
\end{aligned}$$

The ML-II posterior density of θ, given b and g_0 is a shrinkage estimator. It is a weighted average of the Bayes estimator $\theta_*(b|g_0)$ under the base prior g_0 and the data-dependent empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. If the base prior is consistent with the data, the weight $\widehat{\lambda}_{\theta, g_0} \to 1$ and the ML-II posterior density of θ gives more weight to the posterior $\pi_0^*(\theta|g_0)$ derived from the elicited

[7] More information is given in section B of the supplementary material and in Baltagi et al. (2018, 2021).

prior. In this case $\widehat{\theta}_{ML-II}$ is close to the Bayes estimator $\theta_*(b|g_0)$. Conversely, if the base prior is not consistent with the data, the weight $\widehat{\lambda}_{\theta,g_0} \to 0$ and the ML-II posterior density of θ is then close to the posterior $\widehat{q}^*(\theta|g_0)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. The ability of the ε-contamination model to extract more information from the data is what makes it superior to the classical Bayes estimator based on a single base prior.

The second step of the robust Bayesian estimator focuses on $\widetilde{y} = y - Z\theta$. Moving along the lines of the first step, the ML-II posterior density of b is given by:

$$\widehat{\pi}^*(b|h_0) = \widehat{\lambda}_{b,h_0}\pi_0^*(b|h_0) + \left(1 - \widehat{\lambda}_{b,h_0}\right)\widehat{q}^*(b|h_0), \tag{14}$$

where $\widehat{\lambda}_{b,h_0}$ is an estimated weight, $\pi_0^*(b|h_0)$ is the pdf of a multivariate t-distribution where the mean vector $b_*(\theta|h_0)$ is the Bayes estimate of b for the prior distribution $\pi_0(b,\tau|h_0)$, $q^*(b|h_0)$ is the pdf of a multivariate t-distribution where the mean vector $\widehat{b}_{EB}(\theta|h_0)$ is the empirical Bayes estimator of b for the contaminated prior distribution $q(b,\tau|h_0)$ (see section B of the supplementary material). The mean of the ML-II posterior density of b is hence given by:

$$\widehat{b}_{ML-II} = \widehat{\lambda}_{b,h_0}b_*(\theta|h_0) + \left(1 - \widehat{\lambda}_{b,h_0}\right)\widehat{b}_{EB}(\theta|h_0). \tag{15}$$

Many have raised concern about the unbiasedness of the posterior variance-covariance matrices of $\widehat{\theta}_{ML-II}$ and $\widehat{b}_{ML-II}$. Following Berger (1985), Baltagi et al. (2018) derived the analytical ML-II posterior variance-covariance matrices of $\widehat{\theta}_{ML-II}$ and $\widehat{b}_{ML-II}$. Unfortunately, both are biased towards zero as $\widehat{\lambda}_{\theta,g_0}$ and $\widehat{\lambda}_{b,h_0} \to 0$ and converge to the empirical variance which is known to underestimate the true variance (see *e.g.* Berger and Berliner (1986); Gilks et al. (1997); Robert (2007)). Consequently, to approximate the true ML-II variances, Baltagi et al. (2018, 2021) proposed two different strategies, each with different desirable properties: 1) MCMC with multivariate t-distributions or 2) block resampling bootstrap. In addition, they proposed a mixture of multivariate skewed (or non-skewed) t-distributions to decrease the computational time (see section B of the supplementary material). In what follows, we will use block resampling bootstrap and mixtures of multivariate t-distributions.

4. A Monte Carlo simulation study

4.1. The DGP of the Monte Carlo simulation study

We consider a number of distinct statistical worlds. These include the random effects (RE) world, the Chamberlain (1982)-type fixed effects (FE) world and the Hausman and Taylor (1981) (HT) world. We extend the DGPs used in Baltagi et al. (2018, 2021) to the dynamic space-time case. For the dynamic space-time panel data model with common trends or with common correlated effects, we draw inspiration from the DGP of Chudik and Pesaran (2015a,b) and Baltagi et al. (2021).

Consider the dynamic space-time panel data model:

$$\begin{aligned} y_{ti} &= \phi y_{t-1,i} + \rho y_{ti}^* + \delta y_{t-1,i}^* + x_{1,ti}\beta_1 + x_{2,ti}\beta_2 + V_{1,i}\eta_1 + V_{2,i}\eta_2 + \mu_i + u_{ti}, \\ &\quad \text{for } i = 1, ..., N \text{ , } t = 2, ..., T, \text{ with} \end{aligned} \tag{16}$$

$$\begin{aligned} x_{1,ti} &= 0.7x_{1,t-1,i} + \zeta_i + \varkappa_{ti} \\ u_{ti} &\sim N\left(0, \tau^{-1}\right),\ (\zeta_i, \varkappa_{ti}) \sim U(-6,6). \end{aligned}$$

where $y^*_{ti} = \sum_{j=1}^N w_{ij} y_{tj}$ and $y^*_{t-1,i} = \sum_{j=1}^N w_{ij} y_{t-1,j}$.

1. For a dynamic space-time random effects (RE) world, we assume that:

$$\begin{aligned} \eta_1 &= \eta_2 = 0 \\ x_{2,ti} &= 0.7x_{2,t-1,i} + \kappa_i + \vartheta_{ti}\ ,\ (\kappa_i, \vartheta_{ti}) \sim U(-6,6) \\ \mu_i &\sim N\left(0, \sigma^2_\mu\right)\ ,\ \sigma^2_\mu = 4\tau^{-1}. \end{aligned}$$

   Furthermore, $x_{1,ti}$ and $x_{2,ti}$ are assumed to be exogenous in that they are not correlated with μ_i and u_{ti}.

2. For a dynamic space-time Chamberlain-type fixed effects (FE) world, we assume that:

$$\begin{aligned} \eta_1 &= \eta_2 = 0; \\ x_{2,ti} &= \delta_{2,i} + \omega_{2,ti}\ ,\ \delta_{2,i} \sim N(m_{\delta_2}, \sigma^2_{\delta_2}),\ \omega_{2,ti} \sim N(m_{\omega_2}, \sigma^2_{\omega_2}); \\ m_{\delta_2} &= m_{\omega_2} = 1,\ \sigma^2_{\delta_2} = 8,\ \sigma^2_{\omega_2} = 2; \\ \mu_i &= x_{2,1i}\pi_1 + x_{2,2i}\pi_2 + ... + x_{2,Ti}\pi_T + \nu_i,\ \nu_i \sim N(0, \sigma^2_\nu); \\ \sigma^2_\nu &= 1,\ \pi_t = (0.8)^{T-t} \text{ for } t = 1, ..., T. \end{aligned}$$

   $x_{1,ti}$ is assumed to be exogenous but $x_{2,ti}$ is correlated with μ_i and we assume an exponential growth for the correlation coefficient π_t.

3. For a dynamic space-time Hausman-Taylor (HT) world, we assume that:

$$\begin{aligned} \eta_1 &= \eta_2 = 1; \\ x_{2,ti} &= 0.7x_{2,t-1,i} + \mu_i + \vartheta_{ti}\ ,\ \vartheta_{ti} \sim U(-6,6); \\ V_{1,i} &= 1,\ \forall i; \\ V_{2,i} &= \mu_i + \zeta_i + \theta_i + \xi_i\ ,\ (\theta_i, \xi_i) \sim U(-6,6); \\ \mu_i &\sim N\left(0, \sigma^2_\mu\right) \text{ and } \sigma^2_\mu = 4\tau^{-1}. \end{aligned}$$

   $x_{1,ti}$ and $V_{1,i}$ are assumed to be exogenous while $x_{2,ti}$ and $V_{2,i}$ are endogenous because they are correlated with μ_i but not with the u_{ti}.

4. For a dynamic space-time homogeneous panel data world with common trends, (see Chudik and Pesaran (2015a,b)), we assume that

$$\begin{aligned} y_{ti} &= \phi y_{t-1,i} + \rho y^*_{t,i} + \delta y^*_{t-1,i} + x_{ti}\beta_1 + x_{t-1,i}\beta_2 + f'_t\gamma_i + u_{ti}, \qquad (17) \\ &\quad \text{for } i = 1, ..., N\ ,\ t = 2, ..., T, \end{aligned}$$

   with

$$\begin{aligned} x_{ti} &= f'_t\gamma_{x_i} + \omega_{x_{ti}} \\ \omega_{x_{ti}} &= \varrho_{x_i}\omega_{x_{t-1,i}} + \zeta_{x_{ti}} \\ \gamma_{il} &= \gamma_l + \eta_{i,\gamma_l}, \text{ for } l = 1, ..., m \\ \gamma_{x_{il}} &= \gamma_{x_l} + \eta_{i,\gamma_{x_l}}, \text{ for } l = 1, ..., m \end{aligned}$$

where

$$\zeta_{x_{ti}} \sim U(-3,3)\ , \qquad \eta_{i,\gamma_l} \sim N(0,\sigma^2_{\gamma_l})\ , \qquad \eta_{i,\gamma_{x_l}} \sim N(0,\sigma^2_{\gamma_{x_l}})$$
$$\sigma^2_{\gamma_l} = \sigma^2_{\gamma_{x_l}} = 0.2^2\ , \qquad \gamma_l = \sqrt{l \times c_\gamma}\ , \qquad \gamma_{x_l} = \sqrt{l \times c_{x,l}}$$
$$c_\gamma = (1/m) - \sigma^2_{\gamma_l}\ , \quad c_{x,l} = \tfrac{2}{m(m+1)} - \tfrac{2\sigma^2_{\gamma_{x_l}}}{(m+1)}\ , \quad \text{and } u_{ti} \sim N\left(0,\tau^{-1}\right).$$

f_t and γ_i are $(m \times 1)$ vectors. We consider $m = 2$ deterministic known common trends: one linear trend $f_{t,1} = t/T$ and one polynomial trend: $f_{t,2} = t/T + 1.4(t/T)^2 - 3(t/T)^3$ for $t = 2, ..., T$.

5. For a dynamic space-time homogeneous panel data world with correlated common effects (see Chudik and Pesaran (2015a,b), Yang (2021)), we assume that the m common trends f_t in the model (17), are replaced with unobserved common factors:

$$f_{tl} = \rho_{fl} f_{t-1,l} + \xi_{ftl}\ ,\ \xi_{ftl} \sim U(-0.1, 0.1)\ ,\ l = 1, ..., m$$

   We suppose that the common factors are independent stationary $AR(1)$ processes with $\rho_{fl} = 0.5$ for $l = 1, ..., m$.
6. For a dynamic space-time heterogeneous panel data world with correlated common effects (see Chudik and Pesaran (2015a,b)), we assume that, in the model (17), ϕ (resp. ρ, δ and β_1) is replaced by individual coefficients $\phi_i \sim U(0.6, 0.9)$ (resp. $\rho_i \sim U(0.65, 0.95)$, $\delta_i = -\phi_i\rho_i$ and $\beta_{1i} \sim U(0.5, 1)$) for $i = 1, ..., N$ and keep the m unobserved common factors as defined as previously.

For each set-up, we vary the size of the sample and the duration of the panel. We choose several (N, T) pairs with $N = 63$, 120 and $T = 10$, 20 for cases 1 to 3 and $N = (63, 120)$ and $T = (30, 50)$ for cases 4 to 6. Following Bivand et al. (2008), we use the census tract data set for Central New York State counties featured in Waller and Gotway (2004). More precisely, we work on two subsets of the map consisting of the $N = 63$ census tracts within Syracuse City and the $N = 120$ census tracts within Syracuse City and its neighborhood. We use several weighting matrices $W_N(= \{w_{ij}\})$ which essentially differ in their degree of sparseness. First, we create inverse distance weighting matrices with $w_{ij} = 1/dist(i, j)$ where $dist(i, j)$ is the distance (in km) between two census tracts i and j. The matrix W_N is full save for its diagonal elements which are set to zero. Second, we create weighting matrices from the census tract rook-style and queen-style contiguities, by analogy with movements on a chessboard. Lastly, we create k-nearest neighbors weighting matrices with the $k = 4$ or 10 individuals (see Figures 1 to 3 in section D in the supplementary material). All the weighting matrices are row normalized.

The autoregressive and spatial coefficients take several values $(0.75, 0.3)$ for ϕ and $(0.8, 0.4)$ for ρ while the spatio-temporal diffusion parameter is fixed ($\delta = -\phi\rho$) in most cases and $\beta_1 = \beta_2 = 1$. We set the initial values of y_{ti}, $x_{1,1,ti}$, $x_{1,2,ti}$ and $x_{2,ti}$, x_{ti}, ... to zero. Next, we generate all the $x_{1,ti}$, $x_{2,ti}$, x_{ti}, y_{ti}, u_{ti}, ζ_{ti}, ς_{ti}, $\omega_{2,ti}$, ... over $T + T_0$ time periods and we drop the first $T_0(= 50)$ observations to reduce the dependence on the initial values.

The robust Bayesian estimators for the two-stage hierarchy are estimated with $\varepsilon = 0.5$, though we also investigate their robustness to various values of ε.[8] We must set the hy-

[8] $\varepsilon = 0.5$ is an arbitrary value. We implicitly assume that the amount of error in the base elicited prior is

perparameters values $\theta_0, b_0, g_0, h_0, \tau$ for the initial distributions of $\theta \sim N\left(\theta_0 \iota_{K_1}, (\tau g_0 \Lambda_Z)^{-1}\right)$ and $b \sim N\left(b_0 \iota_{NK_2}, (\tau h_0 \Lambda_D)^{-1}\right)$ where $\theta = [\phi, \rho, \delta, \beta_1, \beta_2, \eta_1, \eta_2]'$ for the first three cases and $\theta = [\phi, \rho, \delta, \beta_1, \beta_2]'$ for the last three cases. While we can choose arbitrary values for θ_0, b_0 and τ, the literature generally recommends using the unit information prior (UIP) to set the g-priors.[9] In the normal regression case, and following Kass and Wasserman (1995), the UIP corresponds to $g_0 = h_0 = 1/((T-1)N)$, leading to Bayes factors that behave like the Bayesian Information Criterion (BIC).

For the 2S robust estimators, we use $BR = 20$ samples in the block resampling bootstrap. For each experiment, we run $R = 1,000$ replications and we compute the means, the standard errors and the root mean squared errors (RMSEs) of the coefficients, the variances of the specific effects and the residual variances. To save on space, we only include tables and comments for the random effects world, the Chamberlain-type fixed effects world and the homogeneous panel data world with common trends. Results for other statistical worlds (Hausman-Taylor, homogeneous (resp. heterogeneous) panel data world with correlated common effects) are reported in the supplementary material.

4.2. Simulation results

4.2.1. The dynamic space-time random effects world

Rewrite the general dynamic model (6) as follows:

$$\begin{aligned} y =& Z\theta + Db + u = Z\theta + Z_\mu \mu + u \\ & \text{with } Z'_{ti} = \left[y_{t-1,i}, y^*_{ti}, y^*_{t-1,i}, x_{1,ti}, x_{2,ti}\right] \ , \theta' = [\phi, \rho, \delta, \beta_1, \beta_2] \,, \end{aligned}$$

where $u \sim N(0, \Sigma)$, $\Sigma = \tau^{-1} I_{(T-1)N}$, $Z_\mu = \iota_{(T-1)} \otimes I_N$ is $((T-1)N \times N)$, $\otimes$ is the Kronecker product, $\iota_{(T-1)}$ is a $((T-1) \times 1)$ vector of ones and $\mu (\equiv b)$ is an $(N \times 1)$ vector of idiosyncratic parameters. When $D \equiv Z_\mu$, the random effects, $\mu \sim N\left(0, \sigma^2_\mu I_N\right)$, are associated with the error term $\nu = Z_\mu \mu + u$ with $\text{Var}(\nu) = \sigma^2_\mu \left(J_{(T-1)} \otimes I_N\right) + \sigma^2_u I_{(T-1)N}$, where $J_{(T-1)} = \iota_{(T-1)} \iota'_{(T-1)}$.

This model can also be estimated using MCMC Gibbs sampling and quasi-maximum likelihood (QML) (see Yu et al. (2008), Kripfganz (2016), Bun et al. (2017), Hsiao and Zhou (2018), Moral-Benito et al. (2019)). In what follows, we compare our Bayesian two-stage two-step estimator (B2S2S) with the latter two estimators.[10,11]

Table 1 reports the results of fitting the Bayesian two-stage two-step model (B2S2S) along with those from the QMLE and the MCMC Gibbs sampling, each in a separate panel respectively for $(N = 63, T = 10)$ and $(N = 120, T = 20)$ using a row normalized inverse distance weighting matrix, W_N. The true parameter values appear in the first row of the Table. The last column reports the computation time in seconds.[12] Note that the computation time increases significantly as we move from a small sample to a larger one. The B2S2S estimator with mixtures of

50%. In other words, $\varepsilon = 0.5$ means that we elicit the π_0 prior but feel we could be as much as 50% off (in terms of implied probability sets).

[9] We chose: $\theta_0 = 0, b_0 = 0$ and $\tau = 1$.

[10] See section C in the supplementary material. For the MCMC Gibbs sampling, we explicitly introduce uniform distributions for ϕ, ρ and δ. We use $1,000$ draws and a warmup of 500 burn-in draws.

[11] We use our own R codes for the Bayesian two-stage two-step model (B2S2S) and the MCMC Gibbs sampling and the "xtdpdqml" Stata command for the QML estimator. We use the same DGP set under R and Stata environments to compare the three methods.

[12] The simulations were conducted using R version 3.3.2 on a MacBook Pro, 2.8 GHz core i7 with 16Go 1600 MGz DDR3 ram.

t-distributions for the standard errors (hereafter se_mixt in the Tables) is the fastest, followed by the B2S2S estimator with block resampling bootstrap for the standard errors (hereafter se_boot in the Tables), whereas the MCMC Gibbs sampling needs considerably more computation time to get very similar estimates.[13]

The first noteworthy feature of the Table is that all the estimators yield parameter estimates, standard errors and RMSEs that are very close.[14] The B2S2S estimator yields a slightly underestimated σ_μ^2 whereas the MCMC Gibbs sampling yields a very precise estimate. On the other hand, the latter is obtained at a huge computational cost. The numerical standard errors ("nse") and the convergence diagnostic ("cd") confirm the good mixing of the MCMC draws.[15] We first estimate the Bayesian two-stage two-step model (B2S2S) with block resampling bootstrap.[16] It is worth mentioning that only the estimates of the variance of the specific effects are biased when using the B2S2S and QMLE estimators. The biases are nevertheless relatively small (resp. -3.25% and -2.75% for B2S2S and QMLE) and decrease as N and T increase (resp. -1.25% and -0.25% for B2S2S and QMLE). The estimated values of the other parameters are virtually unbiased (1% or less). Table 1 confirms that the base prior is not consistent with the data since $\hat{\lambda}_{\theta,g_0}$ is close to zero. The ML-II posterior density of θ is close to the posterior $\widehat{q}^*(\theta|g_0)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. Conversely, $\widehat{\lambda}_\mu$ is close to 0.5 so the Bayes estimator $b_*(\theta|h_0)$ under the base prior h_0 and the empirical Bayes estimator $\widehat{b}_{EB}(\theta|h_0)$ each contributes similarly to the random effects $b_i(\equiv \mu_i)$. Below the table we stress that the stationarity conditions of the B2S2S estimator are satisfied. The QLME gives similar results but is computationally considerably more demanding. It is important to note that the standard deviations of ϕ, β_1 and β_2 when using the B2S2S estimator with mixtures of t-distributions (B2S2S_mixt) are slightly underestimated relative to those of B2S2S_boot, QMLE or the full Bayesian estimator. There is thus a trade-off between slightly biased standard deviations and exceedingly large computation time.

We next simulate the model when the spatial dependence parameter (ρ) is decreased from 0.8 to 0.4. To save space, the results are reported in Table G.1 of the supplementary material. As above, we consider the row normalized inverse distance weighting matrix W_N. In a nut shell, the B2S2S and QMLE estimators do just as well as the MCMC when $N = 63$ and $T = 10$ but in considerably less computation time (we do not run MCMC for $N = 120$ and $T = 20$ due to the excessive computation time). Once again, our B2S2S estimator satisfies the implicit stationarity conditions of the dynamic space-time structure. We also simulate the model by setting ϕ to 0.3 instead 0.75 while maintaining ρ at 0.8 (see Table G.2 in the supplementary material). We draw the same conclusions as for Table G.1, namely that the B2S2S and QMLE estimators perform

[13] For the sake of brevity, we will henceforth write B2S2S_mixt and B2S2S_boot when referring to the B2S2S estimators with mixtures of t-distributions and with block resampling bootstrap, respectively.

[14] Strictly speaking, we should mention "posterior means" and "posterior standard errors" whenever we refer to Bayesian estimates and "coefficients" and "standard errors" when discussing frequentist ones. For the sake of brevity, we will use "coefficients" and "standard errors" in both cases.

[15] The "nse", often referred to as the Monte-Carlo error, is equal to the difference between the mean of the sampled values and the true posterior mean. As a rule of thumb, as many simulations as necessary should be conducted so as to ensure that the Monte Carlo error of each parameter of interest is less than approximately 10% of the sample standard error. As shown in the Table, the estimated nse easily satisfy this criterion. The "cd" compares means calculated from the first 10% and last 40% draws of the Markov chain. Under the null hypothesis of no difference between these means, $cd \sim N(0,1)$ and indicates that a sufficiently large number of draws have been taken. See Koop (2003); Koop et al. (2007).

[16] Recall that we use only $BR = 20$ individual block bootstrap samples. Fortunately, the results are very robust to the value of BR. For instance, increasing BR from 20 to 200 in the random effects world increases the computation time tenfold but yields practically the same results.

as well as the MCMC when $N = 63$ and $T = 10$ and that their estimates are very close to one another for $N = 120$ and $T = 20$, although the B2S2S estimator is considerably faster. Finally, we report the results when setting $\phi = 0.3$ and $\rho = 0.4$ in Table G.3 of the supplementary material. The same conclusions hold as those for Tables G.1 and G.2.

Next, we investigate the properties of our estimators when the autoregressive time dependence parameter is close to the unit root, *i.e.* $\phi = 0.98$ for $N = 63$ and $T = 10$. The spatial dependence parameter takes two values: $\rho = (0.8; 0.4)$ (See Table G.4 of the supplementary material). Interestingly, in such an environment the stationarity conditions are still satisfied as confirmed by the 95% HPDI. It does not therefore seem necessary to impose a stationarity constraint on the prior distribution of ϕ (nor on ρ and consequently on δ). Three features of the simulation results are worth mentioning. First, the B2S2S and MCMC estimators yield a bias of similar magnitude but in opposite direction for σ_μ^2 ($\pm 5.4\%$). On the other hand, when the spatial dependence parameter is reduced to $\rho = 0.4$, the bias of the MCMC estimator is lower (-1.5%). Conversely, the bias of the QMLE is very large (-34.6%). Second, the Stata procedure "xtdpdqml" which corresponds to the QML estimator yields an unrealistic estimate of the variance σ_u^2 of the remainder disturbance. Third, the other parameters of the model (ϕ, ρ, δ and β) are not biased in any significant way, regardless of the estimation method.

We next investigate the sensitivity of our results to two different types of weighting matrices. All the simulations are conducted by setting $\phi = 0.75$ and $\rho = 0.8$ for $N = 63$ and $T = 10$. First, we use the census tracts of the City of Syracuse to compute rook-style and queen-style contiguity weighting matrices. The non sparsity rates of both matrices are smaller than that of the inverse distance weighting matrix (see Figures 1 and 2 in section D in the supplementary material).[17] Once again, the B2S2S and QMLE estimators perform as well as the MCMC but are both considerably faster (See Table G.5). Second, we compute the 4-nearest and 10-nearest neighbors weighting matrices using the same census tracks. The non-sparsity rates of these weighting matrices are also smaller than that of the inverse distance weighting matrices (see Figures 1 and 3 in section D in the supplementary material).[18] We still conclude that the B2S2S and QMLE estimators do as well as the MCMC but both exhibit more reasonable computation times (See table G.6).

As a last exercise, we study the behavior of the estimators in the context of an explosive process. We thus set $\phi = 1.05$ as in Tao and Yu (2020).[19]. Since $\rho = 0.8$ and $\delta = -\phi\rho$, we are clearly outside the stationarity conditions.[20] As reported in Table G.7 of the supplementary material, the B2S2S and MCMC Gibbs sampling estimators give good results although the variance of the specific effects, σ_μ^2, of the B2S2S is once again slightly downward biased. The

[17] For the $N = 63$ census tract rook-style and queen-style contiguities within Syracuse city, the non sparsity rates are respectively 8.72% and 7.76% while that of the inverse distance weighting matrix is 98.41%.

[18] The with 4-nearest and 10-nearest neighbors weighting matrices have non-sparsity rates of 6.35% and 15.87%, respectively.

[19] In a time series: $x_t = \phi x_{t-1} + u_t$, $t = 1, ..., T$, x_t is said to be local-to-unit-root from the explosive side (LTUE) if $\phi = 1 + 1/T$. x_t is said to be mildly explosive (ME) if $\phi = 1 + (T^\alpha)/T$, with $\alpha = 0.1$ or 0.3 and x_t is said to be explosive (EX) if $\phi > 1$. When T is large, $\phi_{\text{LTUE}} < \phi_{\text{ME}} < \phi_{\text{EX}}$ which is not necessarily the case when T is small (see for instance Phillips, 1987; Phillips and Magdalinos, 2007; Tao and Yu, 2020)

[20] As $\phi = 1.05$, $\rho = 0.8$, $\delta = -0.84$, $\varpi_{\min} = -0.0963$ and $\varpi_{\max} = 1$ where $\varpi_{\min}$ and $\varpi_{\max}$ are the minimum and maximum eigenvalues of the spatial weights matrix W_N, we cannot respect one of the two stationarity conditions (4) in footnote 4:

$$\begin{cases} \phi + (\rho + \delta)\,\varpi_{\min} < 1 & \text{if} \quad \rho + \delta < 0 \quad \rightarrow \quad 1.0538 \not< 1, \\ \phi - (\rho - \delta)\,\varpi_{\max} > -1 & \text{if} \quad \rho - \delta \geq 0 \quad \rightarrow \quad -0.59 > -1. \end{cases}$$

narrow 95% HPDI of ϕ ($[1.0499; 1.0502]$) confirms the presence of an explosive root, rejecting the hypothesis of a unit root or a stationary process. While the QMLE also yields similar results for ϕ, ρ, δ, β_1 and β_2, the estimates of σ_u^2 and σ_μ^2 are not only strongly biased but highly unlikely.

In a RE world, one can legitimately argue that the B2S2S yields as good results as the MCMC Gibbs sampling, irrespective of the autoregressive time dependence parameter, ϕ, the spatial dependence parameter, ρ, and the spatio-temporal diffusion parameter, δ, and whether or not the stationarity conditions are satisfied. Conversely, the QMLE yields similar results to those of the B2S2S and MCMC Gibbs sampling if we are not too close to (or do not exceed) the stationarity conditions. In the majority of cases, the B2S2S and QMLE are similar to MCMC Gibbs sampling but are both undoubtedly preferable from a computational point of view.[21] Given the above results, and for the sake of brevity, the other statistical worlds will be investigated through the B2S2S and QMLE estimators only using a row-normalized inverse distance weighting matrix.

4.2.2. The dynamic space-time Chamberlain-type fixed effects world

For the Chamberlain (1982)-type specification, the individual effects ($D_t b \equiv \mu$) are given by $\mu = \underline{X}\Pi + \nu$, where $\underline{X}$ is a $(N \times (T-1)K_1)$ matrix with $\underline{X}_i = (X'_{i2}, ..., X'_{iT})$ and $\Pi = (\pi'_2, ..., \pi'_T)'$ is a $((T-1)K_1 \times 1)$ vector. Here π_t is a $(K_1 \times 1)$ vector of parameters to be estimated. We compare the QML estimator to our B2S2S estimator. These are based on the transformed model: $y_{ti} = \phi y_{t-1,i} + \rho y^*_{ti} + \delta y^*_{t-1,i} + x_{1,ti}\beta_1 + x_{2,ti}\beta_2 + \sum_{t=2}^{T} x_{2,ti}\pi_t + \nu_i + u_{ti}$ or $y = Z^*\theta^* + Db + u$ where $Z^* = \left[y_{-1}, y^*, y^*_{-1}, x_1, x_2, \underline{x_2}\right]$, $\theta^{*'} = (\phi, \rho, \delta, \beta_1, \beta_2, \Pi)'$, $D = \iota_T \otimes I_N$ and $b = \nu$.

Table 2 shows that once again the results of the B2S2S are very close to — or even better than — those of the QML estimator.[22] Our B2S2S estimator fits the variance parameter of the specific effects, σ_μ^2, better than the QML estimator does. Note that the computation times of the QMLE are 46 (resp. 3) times greater than those of the B2S2S with the mixture approach (resp. with bootstrap). Tables G.8 and G.9 in the supplementary material report the estimates of the π_t coefficients. Both estimators yield estimates that are close to the true values.

4.2.3. The dynamic space-time homogeneous panel data world with common trends

The dynamic homogeneous panel data world with common trends is defined as:

$$y_{ti} = \phi y_{t-1,i} + \rho y^*_{ti} + \delta y^*_{t-1,i} + x_{ti}\beta_1 + x_{t-1,i}\beta_2 + f'_t\gamma_i + u_{ti}$$

Since the m common trends f_t are known, we can rewrite the model as follows:

$$\begin{aligned} y =& Z\theta + Db + u = Z\theta + F\Gamma + u \\ & \text{with } Z'_{ti} = \left[y_{t-1,i}, y^*_{ti}, y^*_{t-1,i}, X'_{ti}\right] \ , \ \theta' = \left[\phi, \rho, \delta, \beta'\right]' \text{ and } X'_{ti} = \left[x_{ti}, x_{t-1,i}\right], \end{aligned}$$

[21] We only used 1,000 draws and 500 burn-in draws for each replication, which is small for MCMC. Despite this, 1,000 replications with $N = 63$, $T = 10$ (resp. $N = 120$, $T = 20$) require more than one hour of CPU time (resp. almost 5 hours). Had we used 10,000 draws and 1,000 burn-in draws, it would have taken 8 (resp. 34) hours for $N = 63$, $T = 10$ (resp. $N = 120$, $T = 20$). The computation times of B2S2S and QMLE are considerably shorter. For instance, in Table 1 the respective computation times are 3min and 7min for $N = 63$, $T = 10$ and 12 min and 20 min for $N = 120$, $T = 20$. When using mixtures of t-distributions, the B2S2S requires as little as 15 sec for $N = 63$, $T = 10$ and 52 sec for $N = 120$, $T = 20$.

[22] We do not provide simulations for other combinations of ϕ, ρ and δ for the sake of brevity.

where $u \sim N(0, \Sigma)$, $\Sigma = \tau^{-1} I_N$. The $((T-1)N \times Nm)$ matrix F of the m common trends is given by

$$F = \left[I_N \otimes f_t' \right]_{t=2,\cdots,T} = \begin{pmatrix} I_N \otimes f_2' \\ \cdots \\ I_N \otimes f_T' \end{pmatrix} \text{ with } f_t' = (f_{t1}, f_{t2}, \cdots, f_{tm})$$

and Γ is the $(Nm \times 1)$ individual varying coefficients vector:

$$\Gamma = vec \begin{pmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{N1} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{N2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{1m} & \gamma_{2m} & \cdots & \gamma_{Nm} \end{pmatrix}$$

The primal form of this model cannot be estimated as is using the dynamic common correlated effects pooled estimator (CCEP) (see Pesaran (2006) and Chudik and Pesaran (2015a,b)). The introduction of spatial terms may bias the CCEP estimator. Bailey et al. (2016) have proposed a two-stage approach to estimate dynamic space-time models with strong and weak cross-sectional dependence but do not consider explanatory variables (*e.g.*, $y_{ti} = \phi y_{t-1,i} + \rho y_{ti}^* + \delta y_{t-1,i}^* + f_t' \gamma_i + u_{ti}$). More recently, Yang (2021) proposed a two-stage least squares (2SLS) and a GMM estimators for a spatial autoregressive model with common factors (*e.g.*, $y_{ti} = \rho y_{ti}^* + x_{ti}\beta + f_t' \gamma_i + u_{ti}$). Yang shows that 2SLS exhibits very small biases and declining RMSEs as N and/or T increase. The IV matrix of instruments is defined as $Q_t = \left(x_t, W_N x_t, W_N^2 x_t\right)$. Interestingly, the GMM estimator provides similar results but does not clearly dominate the 2SLS estimator.[23]

We compare our B2S2S estimator with the 2SLS estimator extended to the dynamic space-time case, but unlike Yang (2021) we do not use only $q = 2$ in our Monte Carlo simulation study (*e.g.* $Q_t = \left(X_t, W_N X_t, W_N^2 X_t\right)$, a $(N \times (q+1)K_1)$ matrix) since it leads to biased estimates and large standard errors.[24] We must use $q = 7$ (*e.g.* $Q_t = \left(X_t, W_N X_t, W_N^2 X_t, \cdots, W_N^7 X_t\right)$) to get good results. The larger the dimension $(N \times (q+1)K_1)$ of the IV matrix Q_t, the better the estimates, especially with respect to the standard errors. We chose samples in which the time span is large $T = 30$ or $T = 50$ with $N = 63$ or $N = 120$ census tracts (in the spirit of Chudik and Pesaran (2015a) and Yang (2021) in their simulations).

Table 3 shows that the results of the B2S2S estimator are close to those of the 2SLS estimator and both yield very small bias. The computation time is greater with our estimator when using the bootstrap procedure. On the other hand, when using the mixture approach the computation time is drastically reduced and our estimator is computationally more efficient as N and T increase.[25] Most importantly, our parameter estimates exhibit much smaller standard errors. This is a major shortcoming of instrumental variable methods: The loss of efficiency is the price to pay when using these methods (not to mention the delicate choice of the instrument set).

[23] With Monte Carlo simulations for a SAR model with i.i.d errors, Yang (2021) shows that the biases (resp. RMSEs) ($\times 100$) of $\rho (= 0.4)$ for 2SLS are smaller (resp. close) to those of GMM: 0.05 (resp. 1.58) for 2SLS and -0.64 (resp. 1.52) for GMM when $N = 50$, $T = 30$ and 0.01 (resp. 0.81) for 2SLS and -0.31 (resp. 0.75) for GMM when $N = 100$, $T = 50$. Similar results are obtained for the coefficient β.

[24] See section E in the supplementary material for more details on the 2SLS estimator of Yang (2021) extended to the dynamic space-time case. We use our own R codes for our Bayesian estimator and the 2SLS estimator.

[25] For $N = 63$, $T = 30$ (resp. $T = 50$), the gain factor is 1.4 (resp. 3.2) and for $N = 120$, $T = 30$ (resp. $T = 50$), the gain factor is 3.3 (resp. 7.8).

4.2.4. The dynamic space-time models for the other statistical worlds

For the Hausman-Taylor world, our Bayesian two-stage two-step (B2S2S) estimation method is compared with the two-stage quasi-maximum likelihood (TSQML) sequential approach proposed by Kripfganz and Schwarz (2019) and adapted to the dynamic space-time framework (see section G.3 and Tables G.4 and G.5 of the supplementary material). The estimates are very close to each other. Yet, the B2S2S has a RMSE of the coefficient of the time-invariant variable of about 50% to that of the TSQML. Interestingly, the standard error of that coefficient is smaller when using the Bayesian estimator as compared to the two-stage QMLE. We also reached the same conclusion in non-spatial static and dynamic models (see Baltagi et al. (2018, 2021)). Finally, the computation times of the two-stage QML sequential approach are huge compared to those of the B2S2S with mixtures of t-distributions or with bootstrap.

For the homogeneous (resp. heterogeneous) panel data world with correlated common effects, we compare our B2S2S estimator with the 2SLS estimator of Yang (2021) extended to the dynamic space-time homogeneous (resp. heterogeneous) case (see sections G.4 and G.5 in the supplementary material). For the homogeneous case, the results of B2S2S are very close to those of the 2SLS of the extended Yang's estimator and lead to better efficiency properties, less computation time, and absence of bias. Lastly, when we introduce a dynamic space-time heterogeneous panel data world with correlated common effects, the results of the B2S2S estimator are also close to those of the 2SLS estimator but the RMSEs of the B2S2S are generally smaller than those of 2SLS.

5. Application to crop yields and climate change

Since the seminal work of Wallace (1920), agricultural economists have shown great interest in estimating crop yield production functions. Most papers have focused on corn as it is the largest crop in the U.S. in terms of tonnage. Annual yields have usually been regressed against observed temperatures and precipitations during the growing season.[26] As pointed out by Burke and Emerick (2016), empirical studies have originally either exploited cross-sectional variations to compare outcomes between warm and cool regions (*e.g.*, Mendelsohn et al. (1994), Schlenker et al. (2005)), or have used time series to contrast outcomes under warm and cool conditions within a given area (*e.g.*, Deschênes and Greenstone (2007, 2011), Schlenker and Roberts (2009), Dell et al. (2012)). More recently, analysts have modeled crop yields within a panel data framework. In addition, some have estimated the effects of temperature on crop yields using the "degree day" approach in order to control for spatial (*e.g.*, soil quality) and common time effects. This specification acknowledges that too high temperatures may harm crop yields while moderate temperatures are likely beneficial (see *e.g.*, Schlenker and Roberts (2009), Lobell et al. (2013), Butler and Huybers (2013), Burke and Emerick (2016)).

Given that climate change evolves on a time scale of several decades, the main empirical challenge is to anticipate the ability of producers to adapt to these long-term trends.[27] Depending

[26] The growing season is generally defined as ranging from April 1st to September 30th in the literature. More specifically, it starts at sowing and lasts approximately 150 days.

[27] As pointed out by Keane and Neal (2020), this may involve the use of more heat-tolerant hybrids, improved water retention in fields, irrigation, adjustment of sowing rates, etc. This adaptation includes all sources of covariation between heat and heat sensitivity of agricultural yields. It implies the active adaptation of farmers to temperature for growing techniques, as well as any other factors (not controlled by farmers) that make yields less sensitive to heat in warmer conditions.

on the speed of adjustment, the deleterious effects of climate change may be minimal or sizeable. While the literature provides mixed results on behavioral adjustments (see for instance Lobell and Burke (2008), Schlenker and Roberts (2009), Butler and Huybers (2013), Porter et al. (2014), Burke and Emerick (2016)), these are necessarily intrinsic within the spatial and temporal components of the historical data which maps weather to crop yields.

A standard specification of the "degree-day" approach may be written as

$$\log y_{ti} = \beta_1 gdd_{ti} + \beta_2 kdd_{ti} + \beta_3 prec_{ti} + \beta_4 prec_{ti}^2 + c_i + \lambda_t + u_{ti}, \tag{18}$$

where y_{ti} is the yield at year t for region (or county) i. The growing degree days, $gdd_{j,ti}$ (resp. the "killing degree days", $kdd_{j,ti}$), is the total time over the growing season during which the crops are exposed to temperatures up to a maximum threshold (resp. above the threshold).[28] Total yield is customarily written as a quadratic function of cumulative precipitation during the growing season, $prec_{ti}$. The spatial (county) and time effects are represented by c_i and λ_t, respectively, and aim to capture intercept heterogeneity (such as soil quality) and changes in total factor productivity that are assumed common across space. The key parameter of interest is $\beta_2 < 0$, which captures the extent to which high temperatures reduce crop yields. To take into account potential adaptation to high temperatures, the specification (18) may be extended as follows:

$$\log y_{ti} = \beta_1 gdd_{ti} + \beta_{2,0} kdd_{ti} + \beta_{2,1} \left(\log\left(kdd_{ti}\right) kdd_{ti} - kdd_{ti}\right) + \beta_3 prec_{ti} + \beta_4 prec_{ti}^2 + c_i + \lambda_t + u_{ti}, \tag{19}$$

leading to a marginal effect of yields with respect to kdd given by $\beta_{2,0} + \beta_{2,1} \log\left(kdd_{it}\right)$. This specification incorporates the strong relationship between the sensitivity of the yields to the climatology of the kdds (Butler and Huybers (2013), Keane and Neal (2020)). *A priori*, we expect a positive effect of gdd ($\beta_1 > 0$), a concave effect of precipitations ($\beta_3 > 0$, $\beta_4 < 0$) and a positive coefficient $\beta_{2,1}$ leading to smaller kdd effects in warmer regions since $\beta_{2,0} < 0$.

Keane and Neal (2020) have also considered adaptation across both regions and time. Since variations in heat sensitivity can occur across space and over time, they estimate a model with both spatial and temporal heterogeneity in the slope coefficients:

$$\log y_{ti} = \beta_{1,ti} gdd_{ti} + \beta_{2,ti} kdd_{ti} + \beta_{3,ti} prec_{ti} + \beta_{4,ti} prec_{ti}^2 + c_i + \lambda_t + u_{ti}. \tag{20}$$

They allow the heterogeneous slopes to be correlated with the regressors as they focus on additive heterogeneity across the county/time dimensions:

$$\beta_{k,ti} = \beta_k + \beta_{k,i} + \beta_{k,t} \ , \ k = 1, ..., 4 \tag{21}$$

They propose a "mean observation OLS" (MO-OLS) method for models that contain both county and time fixed effects in the slope coefficients. This novel static panel data method allows to flexibly estimate the extent of historical adaptation to high temperatures. Their specification implies that each county's relative sensitivity to weather is fixed over time.

Our application uses the same data as in Keane and Neal (2020). Our model acknowledges that crop yields are likely spatially correlated and that time effects may be persistent at the

[28] The threshold for corn is 29°C.

county level. These features argue in favor of a dynamic space-time model defined as

$$\begin{aligned}\log y_{ti} &= \phi \log y_{t-1,i} + \rho \sum_{j=1}^{N} w_{ij} \log y_{tj} + \delta \sum_{j=1}^{N} w_{ij} \log y_{t-1,j} \\ &+\beta_1 gdd_{ti} + \beta_2 \left(\log\left(gdd_{ti}\right) gdd_{ti} - gdd_{ti}\right) + \beta_3 kdd_{ti} + \beta_4 \left(\log\left(kdd_{ti}\right) kdd_{ti} - kdd_{ti}\right) \\ &+\beta_5 prec_{ti} + \beta_6 prec_{ti}^2 + V_i'\eta + f_t'\gamma_i + u_{ti} \ , \ i = 1, ..., N \ , \ t = 2, ..., T, \end{aligned} \tag{22}$$

where the exogenous variables are as specified above. The row-normalized spatial weights, w_{ij}, correspond to the inverse of the squared distances ($w_{ij} = 1/dist^2(i,j)$, in km) between counties i and j. Likewise, f_t is a $(m \times 1)$ vector of common trends defined as the time means of *gdd*, *kdd* and *prec*. These trends capture the U.S.-wide trend changes in temperature and precipitation observed over a long time period. Finally, V_i is $(K_v \times 1)$ vector of time-invariant dummy variables which correspond to the 1980-2016 U.S Köppen-Geiger climate classification (see section H in the supplementary material).

Our specification accounts for potential adaptation to high temperatures *via* the non-linear relationship between the climatology of *kdd*s and the sensitivity of the yield to *kdd* as in Keane and Neal (2020). In addition, it allows potential adaptation to *gdd* through the non-linear relationship between the climatology of *gdd*s and the sensitivity of the yield to *gdd*.[29] According to our specification, as the number of growing degree days increases, the need to adapt lessens if $\beta_1 > 0$ and $\beta_2 < 0$. On the other hand, if global warming implies more killing degree days increases, the need to adapt increases significantly if $\beta_3 < 0$ and $\beta_4 > 0$.

Specifications (20-21) and (22) are two different approaches to the same problem. The Keane and Neal (2020) specification is static and non-spatial but with heterogeneous slope coefficients, the space/time heterogeneity being additive. Moreover, the estimated τ-period-ahead forecasts of the dependent variable also depend on the future values $\beta_{k,t+\tau}$. We must therefore make assumptions about the dynamic time path of the $\beta_{k,t+\tau}$ slope coefficients. In contrast, the specification we propose is dynamic, spatial and with constant slope coefficients. In addition, the specification can include time-invariant covariates as well as unobserved or known common factors. This specification also allows one to discriminate between short-run and long-run effects and to take into account the spatial correlation of marginal effects *via* the spatial matrix W_N.

5.1. Data

The county-level crop yields, the temperature and the precipitation data are taken from the supplementary material of Keane and Neal (2020) and cover the period 1950-2015.[30] Annual growing (resp. killing) degree days gdd_{ti} (resp. kdd_{ti}) are converted into total hours over the growing season (see section H in the supplementary material for the description of data). Likewise, precipitation corresponds to total inches of rain over the growing season. A number of counties had missing values at different years. These were interpolated using the inverse distance weighted method. Doing so yields a balanced panel of $N = 2,678$ corn-growing counties over

[29] Indeed, the MO-OLS estimation on the static model

$$\log y_{ti} = \beta_{1,ti} gdd_{ti} + \beta_{2,ti} kdd_{ti} + \beta_{3,ti} prec_{ti} + \beta_{4,ti} prec_{ti}^2 + c_{ti} + u_{ti} \ , \ i = 1, ..., N \ , \ t = 1, ..., T,$$

implies a non-linear relation between $\hat{\beta}_{1,ti}$ and $\log gdd_{ti}$ and between $\hat{\beta}_{2,ti}$ and $\log kdd_{ti}$ (see Table H.4 and Figures 10 and 11 in the supplementary material).

[30] Their yield data came from the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service. Temperatures and precipitations data were drawn from Schlenker and Roberts (2009).

$T = 66$ years, *i.e.* as many as 176,748 observations. The spatial weight matrix was computed using the counties spatial polygon coordinates from an ESRI Shapefile downloaded from the US Census (see section H of the supplementary material).

The spatial patterns of corn yields, growing and killing degree days and precipitations are displayed in Figure 1.[31] The maps exhibit considerable heterogeneity in crop yields ranging from 17 to 159 bushels per acre. They also underline the high productivity of the corn belt and that of some southwestern and western states (west of the 100th meridian). Growing and killing degree days show a marked separation between the southern and northern counties around the 35th parallel. On the other hand, maximum precipitations occur east of the 100th meridian from south to north.[32]

5.2. Estimation Results

Table 4 reports the robust parameter estimates of the ε-contamination model in equation (22) for years $1951 - 2015$. Except for some Köppen-Geiger climate classification dummies, all coefficients are significantly different from zero. The estimated values of the autoregressive time dependence parameter (ϕ) (resp. the spatial dependence parameter (ρ) and the spatio-temporal diffusion parameter (δ)) are 0.606 (resp. 0.912 and -0.537). The impact of the spatial dependence is stronger than that of the time dependence and the estimated spatio-temporal diffusion parameter is very close to the product of $-\rho \times \phi = -0.553$. The parameter estimates thus satisfy the stationarity conditions.

The adaptation to the effect of the growing degree days on crop yields is statistically significant. As the temperatures of the growing season approach the upper bound of 29°C from below, the positive marginal effect gets smaller. Likewise, the adaptation to the effect of the killing degree days is also statistically significant. Its negative marginal effects also gets smaller as the temperature rises above 29°C. Further, the relation between yields and precipitation is concave, as expected. According to our estimates, only three classes of the Köppen-Geiger climate classification impact crop yields: Cfa , Cfb and Dwa. All three have a negative coefficient and imply lower corn yields of between 7% to 10% relative to other classes. Finally, note that the model exhibits a very good fit ($R^2 = 0.9985$ and $\sigma_u^2 = 0.0282$).[33] The right-hand side of Table 4 reports the 5/95 percentile range of the γ_i parameters associated with common factors f_t which includes the time means of *gdd*, *kdd* and *prec*. These capture the country-wide trends in temperature and precipitation observed over our 65-year sample window. The table shows that the counties are impacted differently by these common trends as there is considerable heterogeneity in the parameters estimates.

As noted earlier, one of the advantages of the dynamic space-time mixed model is its ability to estimate short-run (weather) and long-run (climate) effects through impact multipliers, as well as the τ-period-ahead impact of a (permanent) change in temperature or precipitation at time t. Specifically, it is readily seen from equation (22) that $\partial \log y_{ti} / \partial X_{k,ti}$ represents the contemporaneous direct effect on county i's yield growth rate arising from a change in the kth explanatory variable in county i (see Debarsy et al. (2012), Elhorst (2014)). Furthermore, the

[31] Enlargements of these maps are reported in Figures 6 to 8 of the supplementary material.

[32] See the supplementary material for additional maps and descriptive statistics, as well as data on the distribution of the Köppen-Geiger climate classification across counties.

[33] This estimation is significantly better than that obtained by MO-OLS using the static non-spatial model which yields an $R^2 = 0.793$ and a residual variance $\sigma_u^2 = 0.071$. See Table H.4 in the supplementary material.

cross-partial derivative $\partial \log y_{tj}/\partial X_{k,ti}$ measures the contemporaneous spatial spillover effect on county j, $j \neq i$. Finally, $\partial \log y_{t+\tau,i}/\partial X_{k,tj}$ gives the own ($i = j$) and cross ($i \neq j$) marginal effects on the yield growth rate in county i at time $t+\tau$ of an increase in the kth variable at time t in a specific county. Written in matrix form, $\partial \log y_{t+\tau}/\partial X'_{k,t}$ is a $(N \times N)$ matrix of dynamic multipliers. Following LeSage and Pace (2009), the cumulative direct effect (*i.e.*, cumulative own-county impacts) is computed as the average of the diagonal elements, while the cumulative indirect effect (*i.e.*, diffusion over space and time) is computed as the average of the row sums of the off-diagonal elements. The total cumulative effect corresponds to the sum of the cumulative direct and indirect effects.[34,35]

Table 5 reports the direct, indirect and total impact multipliers as well as the 30-year-ahead multipliers for growing and killing degree days and for precipitations. For the growing degree days, the mean short-run (weather) direct, indirect and total effects on yield growth are 0.008%, 0.05% and 0.06%, respectively. The mean 30-year impacts are estimated at 0.01%, 0.22% and 0.23%, respectively. As shown in the table, the indirect effects (*i.e.*, diffusion over space and time) clearly dominate. This follows from the fact that the value of the spatial dependence parameter, ρ, is larger than that of the autoregressive time dependence parameter, ϕ. Importantly, the table shows that the short-run direct, indirect and total effects as well as the long-run effect vary considerably across counties. Thus, an additional growing degree day leads to an increase in overall corn yields of between 0.19% and 0.29% in the long-run and between 0.05% and 0.08% in the short-run.

The next panel of the table focuses on the killing degree days. Unfortunately, the short-run direct, indirect and total effects on corn yields are larger in absolute value than those of growing degree days. In the long run, an additional *kdd* today is expected to decrease corn yields by as much as -3.31%. Once again, the spatio-temporal diffusion effects dominate the time dependence effect as evidence by a comparison of the direct, indirect and total effects. Thus, an additional degree-day above 29°C leads to a decrease in overall corn yields between -6.63% and -0.70% in the long-run (the climate effect) while the instantaneous effect (the weather effect) is between -1.95% and -0.16%.

The last panel of the table focuses on precipitations. All short-run effects are positive. The mean total impact corresponds to an increase of 0.021% in corn yield. In the long-run, the mean total impact is estimated to be 0.077%. As with *gdd* and *kdd*, the spatio-temporal effects dominate the temporal dependence effect. According to the parameter estimates, an additional inch of precipitation would lead to a mean increase in corn yield of 0.08% in the long-run and to an instantaneous mean increase of 0.02%.

To add to the discussion of Table 5, Figure 2 maps the geographic patterns of the long-run total effects associated with the growing and killing degree days and with precipitations. These figures are very instructive as they unearth interesting spatial patterns.[36] Thus a unit increase in any of the covariates at time t (*i.e.*, in 2015) leads to specific waves of spatial long-term effects. Thus, 30 years hence (*i.e.*, in 2045), Figure 2a shows that the marginal impact of an addition growing degree day will be spread northwesterly with increasing intensity. States that will benefit most include Washington, Montana, Wyoming, Utah, North and South Dakota, Minnesota and

[34] We note that it is not possible to separate out the time from space and space-time diffusion effects in this model except if we constrain δ to be equal to $\delta = -\phi\rho$.

[35] The derivation of the dynamic multipliers is given in section H.2 in the supplementary material.

[36] Enlargements of these maps are reported in Figures 12 to 14 of the supplementary material.

Wisconsin. Corn yields are expected to increase between $[0.26\%, 0.29\%)$ per year. States located further south will not gain as much while the southernmost states located east of Texas will benefit very little.

Surprisingly, the long-term effect of an additional killing degree day spreads into parallel waves with increasing intensity from southwest to northeast states, as depicted in Figure 2b. Producers located in North Dakota, Minnesota, Wisconsin, Michigan, Ohio, New York and the central Appalachians states will be hurt the most. Yields are expected to decrease between $[-6.63\%, -4.82\%)$ per year. On the other hand, the least impacted states will be those from Florida to Texas and Oklahoma.

Lastly, Figure 2c depicts the long-run marginal impacts of an additional unit of precipitation. The vertical line that stretches more or less from North Dakota to West Texas delineate states that will benefit most from those who will not benefit much, if at all. To the west, the long-run total marginal effects are estimated to range between $[0.15\%, 0.42\%)$ per year. To the east, the gains in productivity are modest and vary between $[0.07\%, 0.015\%)$ per year.

A comparison between Figures 1 and 2 helps understand the adaptation mechanisms the are likely to occur in the face of long term climate changes. Focusing first on the growing degree days, it is readily apparent that states that have numerous growing hours will benefit little from an addition *gdd* and vice versa (Figures 1b and 2a). On the other hand, northwestern states who benefit most from an additional *gdd* are also the most vulnerable to an additional killing degree day. Yet, these states face much fewer *kdd* during the growing season than the southern states who also appear to be less vulnerable to an additional *kdd*. This suggests that the crop yields in the northwestern states are much more sensitive to climate changes than the other corn producing states.

6. Conclusion

The dynamic space-time panel data models considered in the paper allow one to account for feedback from lagged endogenous values, state dependence, spatial spillovers, spatial heterogeneity and the interactive effects. The models are based upon an ε-contamination class of priors and are cast within a two-stage hierarchical approach. This setup can potentially extract more information from the data than the classical Bayes estimator with a single base prior. In addition, we show that our approach encompasses a variety of classical or frequentist specifications. The Type-II maximum likelihood procedure leads to posterior distributions of the slope coefficients and the individual effects that are convex combinations of the conditional posterior densities derived from the elicited prior and the ε-contaminated prior. The estimator assigns more weight to the conditional posterior density derived from the former if the base prior is consistent with the data and to the latter otherwise. The finite sample performance of the two-stage hierarchical models is investigated using extensive Monte Carlo experiments. With such a unified toolbox, our estimators are shown to be at least as good as the alternative classical estimators for the statistical worlds we consider.

We use our estimator to investigate the ability of corn producers in the United States to adapt to climate change using the same data as in Keane and Neal (2020). Our robust Bayesian two-stage two-step approach provides a very good fit to the data. As stressed in the paper, one of the advantages of this dynamic space-time mixed model is its ability to decompose the short-run (weather) and long-run (climate) effects into their direct and indirect components through

impact multipliers and τ-period-ahead impacts of a (permanent) change in the temperature or precipitation at time t. Our results show that the spatial dependence largely dominates that of the time dependence, and that the estimated spatio-temporal diffusion parameter is very close to their product. An additional growing degree day has a statistically significant positive but decreasing marginal impact on crop yields. The converse holds for an additional killing degree day. The impact of increased precipitations on crop yield is found to be concave. Finally, the estimates suggest that corn production in the northwestern states is more sensitive to climate changes than elsewhere.

Table 1: Dynamic Space-Time Random Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	4			
$N = 63$	B2S2S coef	0.7510	0.8087	-0.6089	1.0029	1.0028	0.9958	3.8737	$< 10^{-4}$	0.4381	
$T = 10$	se_boot	0.0036	0.0163	0.0168	0.0147	0.0103	0.0580	0.7130			199.38
	se_mixt	0.0019	0.0149	0.0151	0.0081	0.0072	0.0579	0.7137			14.997
	rmse	0.0040	0.0191	0.0194	0.0155	0.0108	0.0581	0.7238			
	QMLE coef	0.7502	0.8090	-0.6084	0.9993	0.9997	0.9928	3.8977			400.25
	se	0.0038	0.0174	0.0178	0.0159	0.0111	0.0618	0.7437			
	rmse	0.0039	0.0205	0.0204	0.0156	0.0109	0.0622	0.7503			
	MCMC coef	0.7507	0.8083	-0.6084	0.9990	0.9991	1.0004	4.0029			4048.75
	se	0.0039	0.0187	0.0192	0.0150	0.0106	0.0579	0.7334			
	rmse	0.0040	0.0205	0.0210	0.0151	0.0106	0.0579	0.7331			
	nse	0.0001	0.0005	0.0005	0.0007	0.0005	0.0002	0.0075			
	cd	0.2350	0.2200	0.1930	0.5040	0.4770	0.4290	0.3220			
$N = 120$	B2S2S coef	0.7509	0.8068	-0.6071	1.0022	1.0020	0.9981	3.9511	$< 10^{-4}$	0.4692	
$T = 20$	se_boot	0.0019	0.0091	0.0094	0.0071	0.0050	0.0298	0.5155			713.19
	se_mixt	0.0009	0.0087	0.0088	0.0041	0.0038	0.0298	0.5154			52.31
	rmse	0.0022	0.0119	0.0122	0.0077	0.0055	0.0299	0.5176			
	QMLE coef	0.7500	0.8072	-0.6066	0.9999	0.9996	0.9967	3.9920			1167.23
	se	0.0018	0.0094	0.0096	0.0074	0.0051	0.0307	0.5308			
	rmse	0.0018	0.0120	0.0117	0.0073	0.0051	0.0309	0.5306			
	MCMC coef	0.7500	0.8069	-0.6064	1.0000	0.9997	0.9987	4.0218			15971.54
	se	0.0019	0.0104	0.0106	0.0073	0.0051	0.0298	0.5295			
	rmse	0.0019	0.0125	0.0123	0.0073	0.0051	0.0298	0.5297			
	nse	0.0001	0.0003	0.0003	0.0003	0.0002	0.0001	0.0031			
	cd	0.3020	0.2170	0.1980	0.5350	0.5440	0.4540	0.3410			

B2S2S : Bayesian two-stage two-step estimation.
se_boot: standard errors computed with individual block resampling bootstrap.
se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
QMLE: quasi-maximum likelihood estimation.
MCMC: MCMC Gibbs sampling with $1,000$ draws and 500 burn-in draws.
Stationarity conditions for B2S2S for $N = 63$, $T = 10$ and for $N = 120$, $T = 20$: $\phi + (\rho + \delta)\,\varpi_{\max} = 0.95(< 1)$ as $\rho + \delta = 0.2(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.65(> -1)$ as $\rho - \delta = 1.4(\geq 0)$.

Table 2: Dynamic Space-Time Chamberlain-type Fixed Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	224.3295			
$N = 63$	B2S2S coef	0.7485	0.8072	-0.6055	1.0030	0.9992	0.9913	224.3587	$< 10^{-4}$	0.4842	
$T = 10$	se_boot	0.0038	0.0140	0.0145	0.0104	0.0144	0.0580	42.9346			339.59
	se_mixt	0.0046	0.0272	0.0278	0.0178	0.0280	0.0580	42.7039			22.57
	rmse	0.0041	0.0160	0.0157	0.0110	0.0158	0.0587	42.9133			
	QMLE coef	0.7503	0.8096	-0.6094	0.9986	0.9998	1.0017	220.4818			981.16
	se	0.0037	0.0149	0.0154	0.0113	0.0158	0.2568	42.3261			
	rmse	0.0046	0.0365	0.0385	0.014	0.0164	0.2566	42.4706			
	true	0.75	0.8	-0.6	1	1	1	222.9121			
$N = 120$	B2S2S coef	0.7491	0.8052	-0.6044	1.0017	0.9997	0.9986	223.0591	$< 10^{-4}$	0.4950	
$T = 20$	se_boot	0.0018	0.0083	0.0085	0.0049	0.0072	0.0280	28.6932			2068.82
	se_mixt	0.0024	0.0166	0.0168	0.0092	0.0146	0.0280	28.6973			100.66
	rmse	0.0021	0.0099	0.0097	0.0051	0.0073	0.0280	28.6793			
	QMLE coef	0.7501	0.8055	-0.6055	0.9997	0.9999	0.9991	220.8988			4221.21
	se	0.0018	0.0087	0.0089	0.0052	0.0076	0.0292	28.4953			
	rmse	0.0019	0.0102	0.0104	0.0049	0.0074	0.0292	28.5521			

B2S2S : Bayesian two-stage two-step estimation.
se_boot: standard errors computed with individual block resampling bootstrap.
se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
QMLE: quasi-maximum likelihood estimation.
The parameters π_t are omitted from the table, see Tables F.8 and F.9 in the supplementary material.

Table 3: Dynamic Space-Time Homogeneous Panel Data Model with Common Trends and row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1			
$N = 63$	B2S2S coef	0.7490	0.8095	-0.6030	1.0001	1.0011	1.0038	$< 10^{-4}$	0.5078	
$T = 30$	se_boot	0.0025	0.0092	0.0093	0.0091	0.0104	0.0322			6199.62
	se_mixt	0.0022	0.0073	0.0077	0.0107	0.0124	0.0322			427.95
	rmse	0.0027	0.0132	0.0098	0.0091	0.0105	0.0324			
	2SLS coef	0.7511	0.7996	-0.6006	1.0004	0.9978	0.9909			615.00
	se	0.0177	0.0106	0.0217	0.0078	0.0349	0.1085			
	rmse	0.0181	0.0117	0.0216	0.0077	0.0354	0.1088			
$N = 63$	B2S2S coef	0.7495	0.8075	-0.6021	0.9997	1.0006	1.0027	$< 10^{-4}$	0.4860	
$T = 50$	se_boot	0.0018	0.0068	0.0068	0.0072	0.0081	0.0248			6858.87
	se_mixt	0.0016	0.0061	0.0064	0.0080	0.0093	0.0248			446.54
	rmse	0.0018	0.0101	0.0072	0.0072	0.0081	0.0250			
	2SLS coef	0.7499	0.7994	-0.5988	0.9998	1.0004	1.0161			1453.37
	se	0.0156	0.0079	0.0179	0.0054	0.0351	0.0984			
	rmse	0.0163	0.0083	0.0180	0.0056	0.0366	0.0996			
$N = 120$	B2S2S coef	0.7490	0.8130	-0.6049	1.0000	1.0016	1.0030	$< 10^{-4}$	0.4941	
$T = 30$	se_boot	0.0018	0.0092	0.0093	0.0068	0.0076	0.0242			8097.24
	se_mixt	0.0016	0.0061	0.0064	0.0078	0.0091	0.0242			530.32
	rmse	0.0021	0.0159	0.0105	0.0068	0.0078	0.0244			
	2SLS coef	0.7497	0.8003	-0.6005	1.0001	1.0006	0.9875			1778.98
	se	0.0171	0.0100	0.0202	0.0059	0.0336	0.1017			
	rmse	0.0181	0.0106	0.0209	0.0060	0.0353	0.1025			
$N = 120$	B2S2S coef	0.7495	0.8097	-0.6026	1.0002	1.0007	1.0032	$< 10^{-4}$	0.4869	
$T = 50$	se_boot	0.0013	0.0070	0.0070	0.0051	0.0058	0.0193			8608.59
	se_mixt	0.0012	0.0055	0.0057	0.0058	0.0068	0.0193			590.15
	rmse	0.0014	0.0120	0.0075	0.0051	0.0059	0.0195			
	2SLS coef	0.7503	0.8001	-0.6005	1.0001	0.9991	1.0087			4591.31
	se	0.0148	0.0076	0.0168	0.0039	0.0336	0.0819			
	rmse	0.0152	0.0086	0.0171	0.0039	0.0340	0.0824			

B2S2S : Bayesian two-stage two-step estimation.
se_boot: standard errors computed with individual block resampling bootstrap.
se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
2SLS: two-stage least squares estimator of Yang (2021) extended to the case of a dynamic space-time model.

Table 4: Robust estimation using ε-contamination of the impacts of temperatures and precipitations on U.S. corn yields for the $N = 2,678$ counties and the $T = 65$ years (1951-2015), ($NT = 174,070$ observations).

B2S2S	coef	se_mixt		γ_{i1} ($\overline{gdd}$)	γ_{i2} ($\overline{kdd}$)	γ_{i3} ($\overline{prec}$)
ϕ	0.605913	0.001483	5%	-0.000244	-0.002851	-0.000958
ρ	0.912530	0.001730	10%	-0.000189	-0.001913	-0.000658
δ	-0.537365	0.002188	25%	-0.000117	-0.000797	-0.000322
Growing Degree Days			mean	-0.000057	0.000159	-0.000004
gdd	0.000399	0.000090	75%	0.000006	0.001114	0.000309
$\log(gdd)gdd - gdd$	-0.000042	0.000011	90%	0.000073	0.002361	0.000679
Killing Degree Days			95%	0.000126	0.003161	0.000965
kdd	-0.002084	0.000078				
$\log(kdd)kdd - kdd$	0.000347	0.000015				
Precipitation						
$prect$	0.000112	0.000010				
$prec^2$	-0.000083	0.000008				
Köppen-Geiger climate classification						
KG_Aw	-0.037331	0.040117				
KG_BSh	-0.023433	0.038771				
KG_BSk	-0.020882	0.037975				
KG_BWh	0.060343	0.040854				
KG_BWk	-0.015472	0.040647				
KG_Cfa	-0.065704	0.038188				
KG_Cfb	-0.069362	0.038221				
KG_Csa	0.006116	0.038678				
KG_Csb	-0.032743	0.037914				
KG_Dfa	-0.053815	0.038257				
KG_Dfb	-0.049014	0.037479				
KG_Dfc	0.017874	0.037555				
KG_Dsb	-0.035670	0.037555				
KG_Dsc	0.019681	0.043040				
KG_Dwa	-0.106968	0.040369				
KG_Dwb	-0.046474	0.039083				
σ_u^2	0.028204					
λ_θ	$< 10^{-6}$					
λ_μ	0.328997					
R^2	0.998539					

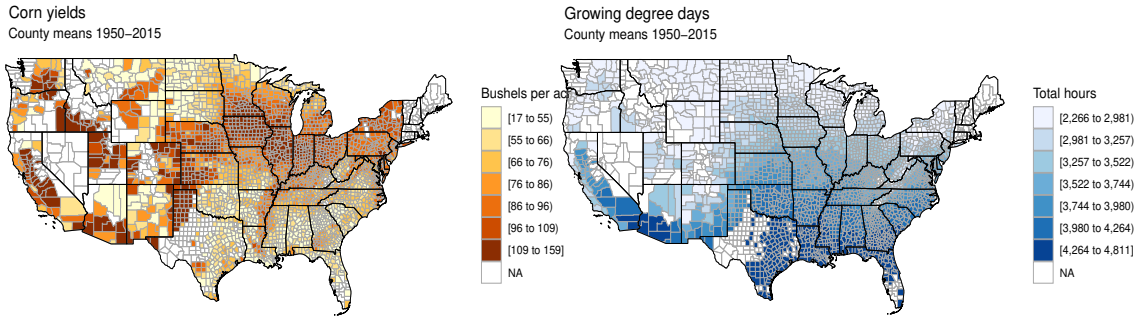
B2S2S: Bayesian two-stage two-step estimation.
se_mixt: standard errors of the parameters θ computed with mixture of
t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
$f_t' = \left(\overline{gdd}, \overline{kdd}, \overline{prec}\right)$.
Stationarity conditions for B2S2S :
$\phi + (\rho + \delta)\,\varpi_{\max} = 0.981(< 1)$ as $\rho + \delta = 0.375(\geq 0)$ and
$\phi - (\rho - \delta)\,\varpi_{\max} = -0.844(> -1)$ as $\rho - \delta = 1.449(\geq 0)$.

KG: Köppen-Geiger climate classification dummies.

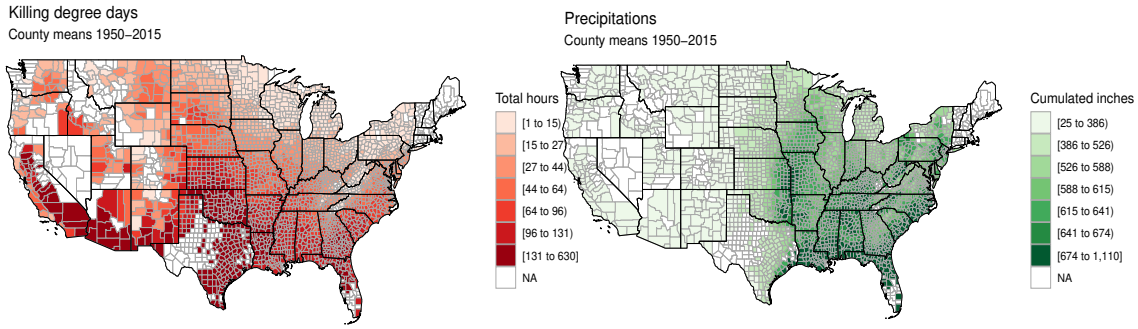
Aw	Tropical wet and dry climate	BSh	Warm semi-arid climate
BSk	Cold semi-arid climate	BWh	Warm desert climate
BWk	Cold desert climate	Cfa	Warm oceanic climate/Humid subtropical climate
Cfb	Temperate oceanic climate	Csa	Warm Mediterranean climate
Csb	Temperate Mediterranean climate	Dfa	Warm/Humid continental climate
Dfb	Temperate/Humid continental climate	Dfc	Cool continental climate/Subarctic climate
Dsb	Warm/Humid continental climate	Dsc	Temperate/Humid continental climate
Dwa	Cool continental climate/Subarctic climate	Dwb	Temperate/Mediterranean continental climate

Table 5: Short-run (weather) and long-run (climate) direct, indirect and total effects of growing and killing degree days and precipitations on growth rates of corn yields for the $N = 2,678$ counties (in percent).

Growing Degree Days	τ		min	10%	25%	mean	75%	90%	max
short-run	$\tau = 0$	direct	0.0051	0.0072	0.0077	0.0089	0.0095	0.0108	0.0212
(weather)		indirect	0.0419	0.0492	0.0522	0.0567	0.0610	0.0649	0.0717
		total	0.0524	0.0567	0.0602	0.0656	0.0706	0.0752	0.0821
long-run	$\tau = 30$	direct	0.0080	0.0128	0.0140	0.0168	0.0183	0.0217	0.0570
(climate)		indirect	0.1690	0.1918	0.2039	0.2202	0.2361	0.2503	0.2726
		total	0.1914	0.2059	0.2188	0.2370	0.2544	0.2704	0.2931
Killing Degree Days	τ		min	10%	25%	mean	75%	90%	max
short-run	$\tau = 0$	direct	-0.4706	-0.2102	-0.1635	-0.1250	-0.0733	-0.0563	0.0266
(weather)		indirect	-1.7299	-1.2820	-1.0459	-0.7936	-0.4886	-0.3999	-0.1457
		total	-1.9490	-1.4889	-1.2124	-0.9186	-0.5613	-0.4609	-0.1585
long-run	$\tau = 30$	direct	-1.2686	-0.4055	-0.3097	-0.2377	-0.1368	-0.1024	0.0509
(climate)		indirect	-6.2462	-4.8440	-4.0157	-3.0725	-1.9397	-1.5779	-0.6740
		total	-6.6329	-5.2439	-4.3283	-3.3102	-2.0752	-1.6989	-0.7014
Precipitations	τ		min	10%	25%	mean	75%	90%	max
short-run	$\tau = 0$	direct	-0.0097	-0.0004	0.0006	0.0031	0.0043	0.0090	0.0303
(weather)		indirect	-0.0362	-0.0004	0.0046	0.0186	0.0265	0.0541	0.1027
		total	-0.0446	-0.0005	0.0052	0.0216	0.0308	0.0637	0.1183
long-run	$\tau = 30$	direct	-0.0226	-0.0008	0.0012	0.0061	0.0082	0.0169	0.0873
(climate)		indirect	-0.1202	0.0004	0.0186	0.0710	0.0987	0.2032	0.3942
		total	-0.1337	0.0002	0.0199	0.0770	0.1065	0.2223	0.4224

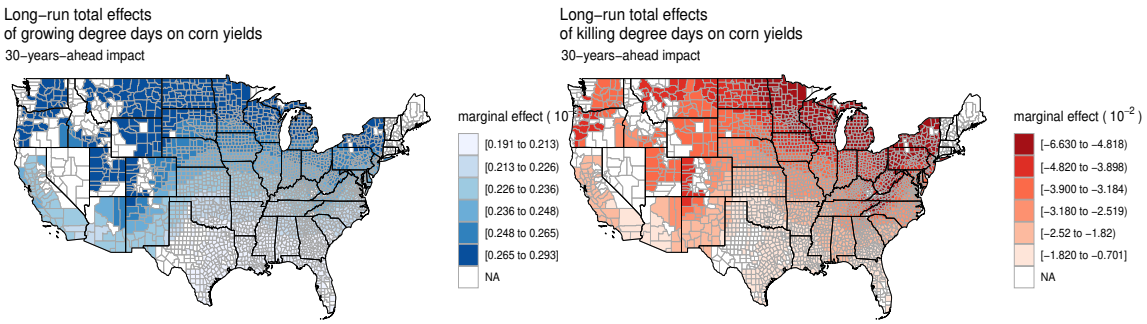


(a) (b)

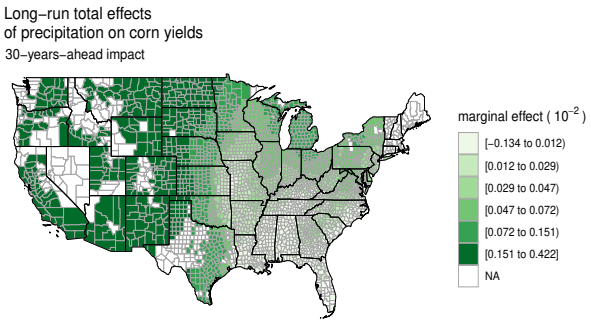


(c) (d)

Figure 1: US county means over 1950-2015. (a) Corn yields. (b) Growing degree days. (c) Killing degree days. (d) Precipitations.



(a) (b)



(c)

Figure 2: Long-run total effects on corn yields. (a) Growing degree days. (b) Killing degree days. (c) Precipitations.

References

Bailey, N., Holly, S., Pesaran, M.H., 2016. A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. Journal of Applied Econometrics 31, 249–280.

Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G., 2018. Robust linear static panel data models using ε-contamination. Journal of Econometrics 202, 108–123.

Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G., 2021. Robust dynamic panel data models using ε-contamination, in: Chudik, A., Hsiao, C., Timmermann, A. (Eds.), Advances in Econometrics, Essays in honor of M. Hashem Pesaran. Emerald Publishing (*forthcoming*).

Berger, J., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Berger, J., Berliner, M., 1986. Robust Bayes and empirical Bayes analysis with ε-contaminated priors. Annals of Statistics 14, 461–486.

Bivand, R.S., Gomez-Rubio, V., Pebesma, E.J., 2008. Applied Spatial Data Analysis with R. Springer.

Bun, M.J.G., Carree, M.A., Juodis, A., 2017. On maximum likelihood estimation of dynamic panel data models. Oxford Bulletin of Economics and Statistics 79, 463–494.

Burke, M., Emerick, K., 2016. Adaptation to climate change: Evidence from US agriculture. American Economic Journal: Economic Policy 8, 106–40.

Butler, E.E., Huybers, P., 2013. Adaptation of US maize to temperature variations. Nature Climate Change 3, 68–72.

Chamberlain, G., 1982. Multivariate regression models for panel data. Journal of Econometrics 18, 5–46.

Chaturvedi, A., 1996. Robust Bayesian analysis of the linear regression. Journal of Statistical Planning and Inference 50, 175–186.

Chudik, A., Pesaran, M.H., 2015a. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393–420.

Chudik, A., Pesaran, M.H., 2015b. Large panel data models with cross-sectional dependence: A survey, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 3–45.

Debarsy, N., Ertur, C., LeSage, J.P., 2012. Interpreting dynamic space–time panel data models. Statistical Methodology 9, 158–171.

Dell, M., Jones, B.F., Olken, B.A., 2012. Temperature shocks and economic growth: Evidence from the last half century. American Economic Journal: Macroeconomics 4, 66–95.

Deschênes, O., Greenstone, M., 2007. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. The American Economic Review 97, 354–385.

Deschênes, O., Greenstone, M., 2011. Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. American Economic Journal: Applied Economics 3, 152–85.

Elhorst, J.P., 2014. Spatial Econometrics: From Cross-Sectional Data to Spatial Panels. Springer.

Fernández, C., Ley, E., Steel, M.F.J., 2001. Benchmark priors for Bayesian model averaging. Journal of Econometrics 100, 381–427.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1997. Markov Chain Monte Carlo in Practice. 2nd ed., Chapman & Hall, London, UK.

Good, I.J., 1965. The Estimation of Probabilities. MIT Press, Cambridge, MA.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.

Hsiao, C., Zhou, Q., 2018. Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models. Journal of Econometrics 207, 114–128.

Jin, B., Wu, Y., Rao, C.R., Hou, L., 2020. Estimation and model selection in general spatial dynamic panel data models. Proceedings of the National Academy of Sciences 117, 5235–5241.

Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. Journal of the American Statistical Association 90, 928–934.

Keane, M., Neal, T., 2020. Climate change and US agriculture: Accounting for multidimensional slope heterogeneity in panel data. Quantitative Economics 11, 1391–1429.

Koop, G., 2003. Bayesian Econometrics. Wiley, New York.

Koop, G., Poirier, D.J., Tobias, J.L., 2007. Bayesian econometric methods. Cambridge University Press.

Kripfganz, S., 2016. Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. The Stata Journal 16, 1013–1038.

Kripfganz, S., Schwarz, C., 2019. Estimation of linear dynamic panel data models with time-invariant regressors. Journal of Applied Econometrics 34, 526–546.

Lee, L.F., Yu, J., 2015. Spatial panel data models, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 363–401.

LeSage, J., Pace, R., 2009. An Introduction to Spatial Econometrics. CRC Press, Taylor-Francis.

LeSage, J.P., Chih, Y.Y., Vance, C., 2019. Markov chain Monte carlo estimation of spatial dynamic panel models for large samples. Computational Statistics & Data Analysis 138, 107–125.

Lobell, D.B., Burke, M.B., 2008. Why are agricultural impacts of climate change so uncertain? the importance of temperature relative to precipitation. Environmental Research Letters 3, 034007.

Lobell, D.B., Hammer, G.L., McLean, G., Messina, C., Roberts, M.J., Schlenker, W., 2013. The critical role of extreme heat for maize production in the United States. Nature Climate Change 3, 497–501.

Mendelsohn, R., Nordhaus, W.D., Shaw, D., 1994. The impact of global warming on agriculture: a Ricardian analysis. The American Economic Review , 753–771.

Moral-Benito, E., Allison, P., Williams, R., 2019. Dynamic panel data modelling using maximum likelihood: an alternative to Arellano-Bond. Applied Economics 51, 2221–2232.

Parent, O., LeSage, J.P., 2010. A spatial dynamic panel model with random effects applied to commuting times. Transportation Research Part B: Methodological 44, 633–645.

Parent, O., LeSage, J.P., 2011. A space-time filter for panel data models containing random effects. Computational Statistics & Data Analysis 55, 475–490.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.

Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.

Phillips, P.C.B., 1991. To criticize the critics: An objective Bayesian analysis of stochastic trends. Journal of Applied Econometrics 6, 333–364.

Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. Journal of Econometrics 136, 115–130.

Porter, J.R., Xie, L., Challinor, A.J., Cochrane, K., Howden, S.M., Iqbal, M.M., Travasso, M.I., 2014. Food security and food production systems, in: IPCC, T.I.P.o.C.C. (Ed.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 485–533.

Robert, C.P., 2007. The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation. 2nd ed., Springer, New York, USA.

Schlenker, W., Hanemann, W.M., Fisher, A.C., 2005. Will US agriculture really benefit from global warming? accounting for irrigation in the hedonic approach. The American Economic Review 95, 395–406.

Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proceedings of the National Academy of Sciences 106, 15594–15598.

Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. Journal of Econometrics 185, 230–258.

Tao, Y., Yu, J., 2020. Model selection for explosive models, in: Essays in Honor of Cheng Hsiao, Advances in Econometrics. Emerald Publishing Limited. volume 41, pp. 73–104.

Wallace, H.A., 1920. Mathematical inquiry into the effect of weather on corn yield in the eight corn belt states. Monthly Weather Review 48, 439–446.

Waller, L.A., Gotway, C.A., 2004. Applied Spatial Statistics for Public Health Data. John Wiley & Sons.

Yang, C.F., 2021. Common factors and spatial dependence: an application to US house prices. Econometric Reviews 40, 14–50.

Yu, J., De Jong, R., Lee, L.F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146, 118–134.

Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distribution, in: Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, Studies in Bayesian Econometrics. North-Holland, Amsterdam. volume 6, pp. 389–399.

Contents

A Some comments related to the debate on stationarity conditions and priors in spatial Bayesian models **2**

B The robust dynamic space-time model in the two-stage hierarchy **3**
B.1 The first step of the robust Bayesian estimator 4
B.2 The second step of the robust Bayesian estimator 9
B.3 Estimating the ML-II posterior variance-covariance matrix 11
B.4 A simple and efficient way to drastically reduce the computation time of our Bayesian two-stage two-step estimator. 12

C Full Bayesian estimator for the random effects world **15**
C.1 Gibbs sampling for the RE world 15
C.2 Derivation of the posterior densities of the Gibbs sampling for the RE world 17

D The spatial weighting matrices **21**

E A two-stage least squares (2SLS) estimator for the dynamic space-time homogeneous panel data world with correlated common factors. **25**

F A two-stage least squares (2SLS) estimator for the dynamic space-time heterogeneous panel data world with correlated common factors. **26**

G Some Monte Carlo simulation results. **29**
G.1 Some results for the dynamic space-time random effects world 29
G.2 Some results for the dynamic space-time Chamberlain-type fixed effects world 37
G.3 The dynamic space-time Hausman-Taylor world: results of the Monte Carlo simulation study 40
G.4 The dynamic space-time homogeneous panel data world with correlated common effects: results of the Monte Carlo simulation study 43
G.5 The dynamic space-time heterogeneous panel data world with correlated common effects: results of the Monte Carlo simulation study 45

H Application on crop yields and climate change **48**
H.1 The dataset 48
H.2 Direct effects, spatial spillover effects and total effects 60

A. Some comments related to the debate on stationarity conditions and priors in spatial Bayesian models

For Bayesian estimation, the spatial literature tells us that the priors for the space-time parameters ϕ, ρ and δ should be defined over the stationary interval (eq.(4) in the main text). As a uniform joint prior distribution over this interval does not produce vague marginal priors, and following Parent and LeSage (2010), a prior can be constructed that takes the form of a product of probability density functions: $p(\phi, \rho, \delta) = p(\rho)p(\delta|\rho)p(\phi|\rho, \delta)$. If the parameter space for ρ is assumed to be a compact subset of $(-1, 1)$, then the conditional prior $p(\phi|\rho, \delta) \sim U(-1+|\rho-\delta|, 1-|\rho+\delta|)$ and the conditional prior $p(\delta|\rho) \sim U(-1+|\rho|, 1-|\rho|)$. The last prior is therefore $p(\rho) \sim U(-1, 1)$. Then, the joint prior is a uniform distribution and equal to $1/2$ over the parameter space defined by the stationary interval (eq.(4) in the main text). But Parent and LeSage (2010) adopt independent uniform priors for ϕ, ρ and δ over the interval $(-1, 1)$ and, in a very standard way, the use Normal distribution for the prior of β and the inverse-Gamma distributions for the priors of the specific effects variance and the remainder variance. They follow the block sampling method proposed by Chib and Carlin (1999) who suggest first sampling β marginalized b and then sampling b conditioned on β. Posterior distributions are standard and can be found in Koop (2003), LeSage and Pace (2009) or Chan et al. (2019). Parent and LeSage (2010), Parent and LeSage (2011) use a restriction on $\delta(= -\rho \times \phi)$ allowing space and time to be separable.

In a non-spatial framework, the pros and cons of imposing a stationarity hypothesis in a Bayesian setup have focused on the implementation of different prior distributions to develop the posterior analysis of autoregressive models with (or without) the stationarity assumption (see for instance Phillips (1991)).[1] Ghosh and Heo (2003) introduced a comparative study to some selected uninformative (objective) priors for the $AR(1)$ model. Ibazizen and Fellag (2003), assumed a uninformative prior for the autoregressive parameter without considering the stationarity assumption for the $AR(1)$ model. However, most literature considers a uninformative (objective) prior for the Bayesian analysis of $AR(1)$ model without considering the stationarity assumption. See for example, DeJong and Whiteman (1991), Schotman and Van Dijk (1991), Sims and Uhlig (1991). For dynamic random coefficients panel data models, Hsiao and Pesaran (2008) do not impose any constraint on the coefficient of the lag dependent variable, ϕ_i. But, following Liu and Tiao (1980), they suggest that one way to impose the stability condition on individual units would be to assume that ϕ_i follows a rescaled Beta distribution on $(0, 1)$. In the time series framework, and for an $AR(1)$ models, Karakani et al. (2016) have performed a posterior sensitivity analysis based on Gibbs sampling with four different priors: natural conjugate prior, Jeffreys' prior, truncated normal prior and g-prior. Their respective performances are compared in terms of highest posterior density region criterion. They show that truncated normal distribution outperforms slightly the g-prior and more strongly the other priors especially when the time dimension is small. On the other hand, for a larger time span, there is no significant difference between truncated normal distribution and g-prior. Nevertheless introducing a truncated normal distribution for ϕ poses very complex integration problems due to the presence of the normal cdf function as integrand in the marginal likelihoods with ε-contamination class of prior distributions.[2]

[1] This debate may apply to the stationarity constraints on the spatial coefficients.

[2] In a dynamic panel data framework with ε-contamination class of prior distributions, Baltagi et al. (2021) have shown that, to avoid these problems of integration, one could assume that ϕ is $U(-1, 1)$. In this case, the mean is (0) and the variance is (1/3), so we do not need to introduce an ε-contamination class of prior distributions for ϕ.

These various debates on the introduction of more or less strong constraints on the priors make us think of some alerts emitted by eminent statisticians. We think of Gelman et al. (2013), Simpson et al. (2017) or Gelman and Yao (2020). Indeed, constraining a prior to follow a $U(-1,1)$ or an even more constraining distribution like that of $p(\phi|\rho,\delta) \sim U(-1+|\rho-\delta|, 1-|\rho+\delta|)$ in LeSage et al. (2019) leads to relatively heavier procedures. As underlined by Lemoine (2019), uniform priors offer no regularization whatsoever even though it is one of the main advantages of going Bayesian. Even worse, if one chooses *e.g.* a $U(0,10)$ prior then one is placing 0 mass outside of this interval and the Bernstein-von Mises theorem does not hold. In other words, even if one had infinite data, one can still be terribly wrong with such priors. As Gelman and Yao (2020) point out, Bayesian statisticians are in resounding agreement that uniform priors are silly and Andrew Gelman, in the Stan documentation.[3], discourages uniform priors.

B. The robust dynamic space-time model in the two-stage hierarchy

The marginal likelihoods (or predictive densities) corresponding to the base priors are:

$$m\left(y^*|\pi_0, b, g_0\right) = \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\theta, \tau|g_0\right) \times p\left(y^*|Z, b, \tau\right) \; d\theta \; d\tau$$

where K_1 is the dimension of θ. Further

$$m\left(\widetilde{y}|\pi_0, \theta, h_0\right) = \int_0^\infty \int_{\mathbb{R}^{NK_2}} \pi_0\left(b, \tau|h_0\right) \times p\left(\widetilde{y}|D, \theta, \tau\right) \; db \; d\tau,$$

where K_2 is the dimension of b and

$$\begin{aligned}
\pi_0\left(\theta, \tau|g_0\right) &= \left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \tau^{-1} \left|\Lambda_Z\right|^{1/2} \exp\left(-\frac{\tau g_0}{2}\left(\theta - \theta_0 \iota_{K_1}\right)' \Lambda_Z (\theta - \theta_0 \iota_{K_1})\right)\right), \\
\pi_0\left(b, \tau|h_0\right) &= \left(\frac{\tau h_0}{2\pi}\right)^{\frac{NK_2}{2}} \tau^{-1} \left|\Lambda_D\right|^{1/2} \exp\left(-\frac{\tau h_0}{2}(b - b_0 \iota_{NK_2})' \Lambda_D (b - b_0 \iota_{NK_2})\right).
\end{aligned}$$

Solving these equations is considerably easier than solving the equivalent expression in the one-step approach.

Unfortunately, the results obtained using Monte Carlo simulations result in biased estimates of ϕ, β and the residual variances.

[3]Prior choice recommendations by Andrew Gelman in https://github.com/stan-dev/stan/wiki

B.1. The first step of the robust Bayesian estimator

Let $y^* = y - Db$. Combining the pdf of y^* and the pdf of the base prior, we get the predictive density corresponding to the base prior[4]:

$$\begin{aligned} m\left(y^*|\pi_0, b, g_0\right) &= \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\theta, \tau|g_0\right) \times p\left(y^*|Z, b, \tau\right)\, d\theta\, d\tau \qquad \text{(B.1)} \\ &= \widetilde{H}\left(\frac{g_0}{g_0+1}\right)^{K_1/2}\left(1+\left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\theta_0}}{1-R^2_{\theta_0}}\right)\right)^{-\frac{(T-1)N}{2}} \end{aligned}$$

with $\widetilde{H} = \frac{\Gamma\left(\frac{(T-1)N}{2}\right)}{\pi^{\left(\frac{(T-1)N}{2}\right)} v(b)^{\left(\frac{(T-1)N}{2}\right)}}$, $R^2_{\theta_0} = \frac{(\widehat{\theta}(b)-\theta_0\iota_{K_1})'\Lambda_Z(\widehat{\theta}(b)-\theta_0\iota_{K_1})}{(\widehat{\theta}(b)-\theta_0\iota_{K_1})'\Lambda_Z(\widehat{\theta}(b)-\theta_0\iota_{K_1})+v(b)}$, $\widehat{\theta}\left(b\right) = \Lambda_Z^{-1}Z'y^*$ and $v\left(b\right) = (y^* - Z\widehat{\theta}\left(b\right))'(y^* - Z\widehat{\theta}\left(b\right))$, and where $\Gamma\left(\cdot\right)$ is the Gamma function.

Likewise, we can obtain the predictive density corresponding to the contaminated prior for the distribution $q\left(\theta, \tau|g_0, h_0\right) \in Q$ from the class Q of possible contamination distributions:

$$m\left(y^*|q, b, g_0\right) = \widetilde{H}\left(\frac{g_q}{g_q+1}\right)^{\frac{K_1}{2}}\left(1+\left(\frac{g_q}{g_q+1}\right)\left(\frac{R^2_{\theta_q}}{1-R^2_{\theta_q}}\right)\right)^{-\frac{(T-1)N}{2}}, \qquad \text{(B.2)}$$

where

$$R^2_{\theta_q} = \frac{(\widehat{\theta}\left(b\right) - \theta_q\iota_{K_1})'\Lambda_Z(\widehat{\theta}\left(b\right) - \theta_q\iota_{K_1})}{(\widehat{\theta}\left(b\right) - \theta_q\iota_{K_1})'\Lambda_Z(\widehat{\theta}\left(b\right) - \theta_q\iota_{K_1}) + v\left(b\right)}.$$

As the ε-contamination of the prior distributions for (θ, τ) is defined by $\pi\left(\theta, \tau|g_0\right) = (1-\varepsilon)\,\pi_0\left(\theta, \tau|g_0\right) + \varepsilon q\left(\theta, \tau|g_0\right)$, the corresponding predictive density is given by:

$$m\left(y^*|\pi, b, g_0\right) = (1-\varepsilon)\, m\left(y^*|\pi_0, b, g_0\right) + \varepsilon m\left(y^*|q, b, g_0\right)$$

and

$$\sup_{\pi\in\Gamma} m\left(y^*|\pi, b, g_0\right) = (1-\varepsilon)\, m\left(y^*|\pi_0, b, g_0\right) + \varepsilon \sup_{q\in Q} m\left(y^*|q, b, g_0\right).$$

The maximization of $m\left(y^*|\pi, b, g_0\right)$ requires the maximization of $m\left(y^*|q, b, g_0\right)$ with respect to θ_q and g_q. The first-order conditions lead to

$$\widehat{\theta}_q = \left(\iota'_{K_1}\Lambda_Z\iota_{K_1}\right)^{-1}\iota'_{K_1}\Lambda_Z\widehat{\theta}\left(b\right) \qquad \text{(B.3)}$$

and

$$\begin{aligned} \widehat{g}_q &= \min\left(g_0, g^*\right), \qquad \text{(B.4)} \\ \text{with } g^* &= \max\left[\left(\frac{((T-1)N-K_1)}{K_1}\frac{(\widehat{\theta}\left(b\right) - \widehat{\theta}_q\iota_{K_1})'\Lambda_Z(\widehat{\theta}\left(b\right) - \widehat{\theta}_q\iota_{K_1})}{v\left(b\right)} - 1\right)^{-1}, 0\right] \\ &= \max\left[\left(\frac{((T-1)N-K_1)}{K_1}\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right) - 1\right)^{-1}, 0\right]. \end{aligned}$$

[4]Derivation of all the following expressions can be found in the supplementary appendix of Baltagi et al. (2018).

Denote $\sup_{q\in Q} m\left(y^{*}|q,b,g_{0}\right)=m\left(y^{*}|\widehat{q},b,g_{0}\right)$. Then

$$m\left(y^{*}|\widehat{q},b,g_{0}\right)=\widetilde{H}\left(\frac{\widehat{g}_{q}}{\widehat{g}_{q}+1}\right)^{\frac{K_{1}}{2}}\left(1+\left(\frac{\widehat{g}_{q}}{\widehat{g}_{q}+1}\right)\left(\frac{R_{\widehat{\theta}_{q}}^{2}}{1-R_{\widehat{\theta}_{q}}^{2}}\right)\right)^{-\frac{(T-1)N}{2}}.$$

If $\pi_{0}^{*}\left(\theta,\tau\mid g_{0}\right)$ denotes the posterior density of (θ,τ) for the prior $\pi_{0}\left(\theta,\tau\right)$ and if $q^{*}\left(\theta,\tau\mid g_{0}\right)$ denotes the posterior density of (θ,τ) for the prior $q\left(\theta,\tau\right)$, then the ML-II posterior density of (θ,τ) is given by

$$\begin{aligned}\widehat{\pi}^{*}\left(\theta,\tau\mid g_{0}\right) &= \frac{p\left(y^{*}\mid X,b,\tau\right)\widehat{\pi}\left(\theta,\tau\mid g_{0}\right)}{\int_{0}^{\infty}\int_{\mathbb{R}^{K_{1}}}p\left(y^{*}\mid X,b,\tau\right)\widehat{\pi}\left(\theta,\tau\mid g_{0}\right)\,d\theta\,d\tau}\\ &= \frac{p\left(y^{*}\mid X,b,\tau\right)\left\{\left(1-\varepsilon\right)\pi_{0}\left(\theta,\tau\mid g_{0}\right)+\varepsilon\widehat{q}\left(\theta,\tau\mid g_{0}\right)\right\}}{\int_{0}^{\infty}\int_{\mathbb{R}^{K_{1}}}p\left(y^{*}\mid X,b,\tau\right)\left\{\left(1-\varepsilon\right)\pi_{0}\left(\theta,\tau\mid g_{0}\right)+\varepsilon\widehat{q}\left(\theta,\tau\mid g_{0}\right)\right\}\,d\theta\,d\tau}\\ &= \frac{\left(1-\varepsilon\right)p\left(y^{*}\mid X,b,\tau\right)\pi_{0}\left(\theta,\tau\mid g_{0}\right)+\varepsilon p\left(y^{*}\mid X,b,\tau\right)\widehat{q}\left(\theta,\tau\mid g_{0}\right)}{\left(\begin{array}{c}\left(1-\varepsilon\right)\int_{0}^{\infty}\int_{\mathbb{R}^{K_{1}}}p\left(y^{*}\mid X,b,\tau\right)\pi_{0}\left(\theta,\tau\mid g_{0}\right)d\theta d\tau\\ +\varepsilon\int_{0}^{\infty}\int_{\mathbb{R}^{K_{1}}}p\left(y^{*}\mid X,b,\tau\right)\widehat{q}\left(\theta,\tau\mid g_{0}\right)d\theta d\tau\end{array}\right)}.\end{aligned}$$

Since

$$\begin{aligned}\widehat{\pi}^{*}\left(\theta,\tau\mid g_{0}\right) &= \frac{\left(1-\varepsilon\right)p\left(y^{*}\mid X,b,\tau\right)\pi_{0}\left(\theta,\tau\mid g_{0}\right)+\varepsilon p\left(y^{*}\mid X,b,\tau\right)\widehat{q}\left(\theta,\tau\mid g_{0}\right)}{\left(1-\varepsilon\right)m\left(y^{*}\mid\pi_{0},b,g_{0}\right)+\varepsilon m\left(y^{*}\mid\widehat{q},b,g_{0}\right)}\\ &= \widehat{\lambda}_{\theta}\left(\frac{p\left(y^{*}\mid X,b,\tau\right)\pi_{0}\left(\theta,\tau\mid g_{0}\right)}{m\left(y^{*}\mid\pi_{0},b,g_{0}\right)}\right)+\left(1-\widehat{\lambda}_{\theta}\right)\left(\frac{p\left(y^{*}\mid X,b,\tau\right)\widehat{q}\left(\theta,\tau\mid g_{0}\right)}{m\left(y^{*}\mid\widehat{q},b,g_{0}\right)}\right),\end{aligned}$$

then

$$\widehat{\pi}^{*}\left(\theta,\tau\mid g_{0}\right)=\widehat{\lambda}_{\theta,g_{0}}\pi_{0}^{*}\left(\theta,\tau\mid g_{0}\right)+\left(1-\widehat{\lambda}_{\theta,g_{0}}\right)q^{*}\left(\theta,\tau\mid g_{0}\right)$$

with

$$\widehat{\lambda}_{\theta,g_{0}}=\frac{\left(1-\varepsilon\right)m\left(y^{*}\mid\pi_{0},b,g_{0}\right)}{\left(1-\varepsilon\right)m\left(y^{*}\mid\pi_{0},b,g_{0}\right)+\varepsilon m\left(y^{*}\mid\widehat{q},b,g_{0}\right)}.$$

$$\begin{aligned}\widehat{\lambda}_{\theta,g_{0}} &= \left[1+\frac{\varepsilon m\left(y^{*}\mid\widehat{q},b,g_{0}\right)}{\left(1-\varepsilon\right)m\left(y^{*}\mid\pi_{0},b,g_{0}\right)}\right]\\ &= \left[1+\frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{g}}{\widehat{g}+1}}{\frac{g_{0}}{g_{0}+1}}\right)^{K_{1}/2}\left(\frac{1+\left(\frac{g_{0}}{g_{0}+1}\right)\frac{(\widehat{\theta}(b)-\theta_{0}\iota_{K_{1}})'\Lambda_{X}(\widehat{\theta}(b)-\theta_{0}\iota_{K_{1}})}{v(b)}}{1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{(\widehat{\theta}(b)-\widehat{\theta}_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\theta}(b)-\widehat{\theta}_{q}\iota_{K_{1}})}{v(b)}}\right)^{\frac{N(T-1)}{2}}\right]^{-1}\\ &= \left[1+\frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{g}}{\widehat{g}+1}}{\frac{g_{0}}{g_{0}+1}}\right)^{K_{1}/2}\left(\frac{1+\left(\frac{g_{0}}{g_{0}+1}\right)\left(\frac{R_{\theta_{0}}^{2}}{1-R_{\theta_{0}}^{2}}\right)}{1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\left(\frac{R_{\theta_{q}}^{2}}{1-R_{\theta_{q}}^{2}}\right)}\right)^{\frac{N(T-1)}{2}}\right]^{-1}\end{aligned}$$

Integration of $\widehat{\pi}^*(\theta, \tau \mid g_0)$ with respect to τ leads to the marginal ML-II posterior density of θ :

$$\widehat{\pi}^*(\theta \mid g_0) = \int_0^\infty \widehat{\pi}^*(\theta, \tau \mid g_0)\, d\tau = \widehat{\lambda}_{\theta,g_0} \int_0^\infty \pi_0^*(\theta, \tau \mid g_0)\, d\tau + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) \int_0^\infty q^*(\theta, \tau \mid g_0)\, d\tau.$$

We must first define $\pi_0^*(\theta, \tau \mid g_0)$ and $q^*(\theta, \tau \mid g_0)$. As

$$\pi_0^*(\theta, \tau \mid g_0) = \frac{p(y^* \mid X, b, \tau)\, \pi_0(\theta, \tau \mid g_0)}{m(y^* \mid \pi_0, b, g_0)} = \frac{p(y^* \mid X, b, \tau)\, \pi_0(\theta, \tau \mid g_0)}{\int_0^\infty \int_{\mathbb{R}^{K_1}} p(y^* \mid X, b, \tau)\, \pi_0(\theta, \tau \mid g_0)\; d\theta d\tau},$$

where

$$\begin{aligned} m(y^* \mid \pi_0, b) &= \frac{\Gamma\left(\frac{N(T-1)}{2}\right)}{\pi^{\left(\frac{N(T-1)}{2}\right)} v(b)^{\left(\frac{N(T-1)}{2}\right)}} \left(\frac{g_0}{g_0+1}\right)^{K_1/2} \\ &\quad \times \left(1 + \left(\frac{g_0}{g_0+1}\right) \frac{(\widehat{\theta}(b) - \theta_0 \iota_{K_1})' \Lambda_X (\widehat{\theta}(b) - \theta_0 \iota_{K_1})}{v(b)}\right)^{-\frac{N(T-1)}{2}}, \end{aligned}$$

and where

$$\begin{aligned} p(y^* \mid X, b, \tau)\, \pi_0(\theta, \tau \mid g_0) &= \left(\begin{array}{c} \left(\frac{\tau}{2\pi}\right)^{\frac{N(T-1)}{2}} \left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \tau^{-1} |\Lambda_X|^{1/2} \\ \times \exp\left(-\frac{\tau g_0}{2}(\theta - \theta_0 \iota_{K_1})' \Lambda_X (\theta - \theta_0 \iota_{K_1})\right) \\ \times \exp\left(-\frac{\tau}{2}\left\{v(b) + (\theta - \widehat{\theta}(b))' \Lambda_X (\theta - \widehat{\theta}(b)\right\}\right) \end{array} \right) \\ &= \tau^{\left(\frac{N(T-1)+K_1}{2} - 1\right)} |\Lambda_X|^{1/2} \left(\frac{1}{2\pi}\right)^{\frac{N(T-1)+K_1}{2}} g_0^{\frac{K_1}{2}} \times \exp\left(-\frac{\tau}{2}\varphi_{\pi_0,\theta}\right), \end{aligned}$$

with

$$\begin{aligned} \varphi_{\pi_0,\theta} &= v(\theta) + (g_0+1)(\theta - \theta_*(b))' \Lambda_X (\theta - \theta_*(b)) \\ &\quad + \left(\frac{g_0}{g_0+1}\right) \left(\widehat{\theta}(b) - \theta_0 \iota_{K_1}\right)' \Lambda_X \left(\widehat{\theta}(b) - \theta_0 \iota_{K_1}\right), \end{aligned}$$

then

$$\pi_0^*(\theta, \tau \mid g_0) = L_0(b) \times \tau^{\left(\frac{N(T-1)+K_1}{2} - 1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\pi_0,\theta}\right),$$

where

$$\begin{aligned} L_0(b) &= \frac{2^{-\left(\frac{N(T-1)+K_1}{2}\right)}}{\Gamma\left(\frac{N(T-1)}{2}\right) . \pi^{K_1/2}} . (g_0+1)^{\frac{K_1}{2}} . v(b)^{\frac{N(T-1)}{2}} . |\Lambda_X|^{1/2} \\ &\quad \times \left[\left(1 + \left(\frac{g_0}{g_0+1}\right) \frac{\left(\widehat{\theta}(b) - \theta_0 \iota_{K_1}\right)' \Lambda_X \left(\widehat{\theta}(b) - \theta_0 \iota_{K_1}\right)}{v(b)}\right)^{\left(\frac{N(T-1)}{2}\right)}\right]. \end{aligned}$$

Similarly, the expression of $q^*(\theta, \tau \mid g_0)$ is defined as:

$$\begin{aligned} q^*(\theta, \tau \mid g_0) &= \frac{p(y^* \mid X, b, \tau)\, \widehat{q}(\theta, \tau \mid g_0)}{m(y^* \mid \widehat{q}, b, g_0)} = \frac{p(y^* \mid X, b, \tau)\, \widehat{q}(\theta, \tau \mid g_0)}{\int_0^\infty \int_{\mathbb{R}^{K_1}} p(y^* \mid X, b, \tau)\, \widehat{q}(\theta, \tau \mid g_0)\ d\theta\ d\tau} \\ &= L_{\widehat{q}}(b) \times \tau^{\left(\frac{N(T-1)+K_1}{2}-1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\widehat{q},\theta}\right), \end{aligned}$$

with

$$\begin{aligned} \varphi_{\widehat{q},\theta} &= v(\theta) + (\widehat{g}+1)\left(\theta - \widehat{\theta}_{EB}(b \mid g_0)\right)' \Lambda_X \left(\theta - \widehat{\theta}_{EB}(b \mid g_0)\right) \\ &\quad + \left(\frac{\widehat{g}}{\widehat{g}+1}\right)\left(\widehat{\theta}(b) - \widehat{\theta}_q \iota_{K_1}\right)' \Lambda_X \left(\widehat{\theta}(b) - \widehat{\theta}_q \iota_{K_1}\right) \end{aligned}$$

and

$$\begin{aligned} L_{\widehat{q}}(b) &= \frac{2^{-(K_1)}}{\Gamma\left(\frac{N(T-1)}{2}\right)\pi^{K_1/2}} (\widehat{g}+1)^{\frac{K_1}{2}}\, v(b)^{\left(\frac{N(T-1)}{2}\right)} |\Lambda_X|^{1/2} \\ &\quad \times \left[\left(1 + \left(\frac{\widehat{g}}{\widehat{g}+1}\right) \frac{\left(\widehat{\theta}(b) - \widehat{\theta}_q \iota_{K_1}\right)' \Lambda_X \left(\widehat{\theta}(b) - \widehat{\theta}_q \iota_{K_1}\right)}{v(\theta)}\right)^{\left(\frac{N(T-1)}{2}\right)}\right], \end{aligned}$$

and where $\widehat{\theta}_{EB}(b \mid g_0)$ is the empirical Bayes estimator of θ for the contaminated prior distribution $q(\theta, \tau)$ (see the derivation below):

$$\widehat{\theta}_{EB}(b \mid g_0) = \frac{\widehat{\theta}(b) + \widehat{g}_q \widehat{\theta}_q \iota_{K_1}}{\widehat{g}_q + 1}.$$

Integration of $\widehat{\pi}^*(\theta, \tau \mid g_0)$ with respect to τ leads to the marginal ML-II posterior density of θ :

$$\begin{aligned} \widehat{\pi}^*(\theta \mid g_0) &= \int_0^\infty \widehat{\pi}^*(\theta, \tau \mid g_0)\, d\tau \\ &= \widehat{\lambda}_{\theta,g_0} \int_0^\infty \pi_0^*(\theta, \tau \mid g_0)\, d\tau + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) \int_0^\infty q^*(\theta, \tau \mid g_0)\, d\tau \\ &= \widehat{\lambda}_{\theta,g_0} \pi_0^*(\theta \mid g_0) + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) \widehat{q}^*(\theta \mid g_0) \end{aligned} \tag{B.5}$$

So,

$$\begin{aligned} \pi_0^*(\theta \mid g_0) &= \int_0^\infty \pi_0^*(\theta, \tau \mid g_0)\, d\tau \\ &= L_0(b) \int_0^\infty \tau^{\left(\frac{N(T-1)+K_1}{2}-1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\pi_0,\theta}\right) d\tau \\ &= L_0(b) \times 2^{\left(\frac{N(T-1)+K_1}{2}\right)} \varphi_{\pi_0,\theta}^{\left(-\frac{N(T-1)+K_1}{2}\right)} \Gamma\left(\frac{N(T-1)+K_1}{2}\right). \end{aligned}$$

Then $\pi_0^*(\theta \mid g_0)$ is given by

$$\begin{aligned}\pi_0^*(\theta \mid g_0) &= \frac{\Gamma\left(\frac{N(T-1)+K_1}{2}\right)}{\Gamma\left(\frac{N(T-1)}{2}\right)\pi^{\frac{K_1}{2}}}|\Lambda_X|^{1/2}(g_0+1)^{\frac{K}{2}}\,v(b)^{\left(\frac{N(T-1)}{2}\right)} \times \varphi_{\pi_0,\theta}^{\left(-\frac{N(T-1)+K_1}{2}\right)} \\ &\quad \times \left(1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(\mu)-\theta_0\iota_{K_1}\right)}{v(b)}\right)^{\left(\frac{N(T-1)}{2}\right)}.\end{aligned}$$

We therefore get

$$\pi_0^*(\theta \mid g_0) = \widetilde{H}_{\pi_0}\frac{(g_0+1)^{K_1/2}}{\left((g_0+1)\frac{(\theta-\theta_*(b))'\Lambda_X(\theta-\theta_*(b))}{v(b)}+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)}{v(b)}+1\right)^{\frac{N(T-1)+K_1}{2}}},$$

with

$$\begin{aligned}\widetilde{H}_{\pi_0} &= \frac{\Gamma\left(\frac{N(T-1)+K_1}{2}\right)|\Lambda_X|^{1/2}}{\pi^{K/2}\Gamma\left(\frac{N(T-1)}{2}\right)v(\theta)^{K_1/2}} \\ &\quad \times \left(1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)}{v(b)}\right)^{\frac{N(T-1)}{2}}.\end{aligned}$$

If we suppose that $M_{0,\theta} = \frac{(g_0+1)}{v(b)}\Lambda_X$, then $|M_{0,\theta}|^{1/2} = \left(\frac{g_0+1}{v(b)}\right)^{K_1/2}|\Lambda_X|^{1/2}$ and

$$\begin{aligned}\pi_0^*(\theta \mid g_0) &= \frac{\Gamma\left(\frac{N(T-1)+K_1}{2}\right)|M_{0,\theta}|^{1/2}}{\pi^{K_1/2}\Gamma\left(\frac{N(T-1)}{2}\right)}(\xi_{0,\theta})^{N(T-1)/2}\left[(\theta-\theta_*(b))'M_{0,\theta}(\theta-\theta_*(b))+\xi_{0,\theta}\right]^{-\frac{N(T-1)+K_1}{2}}, \\ \text{with } \xi_{0,\theta} &= 1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(b)-\theta_0\iota_{K_1}\right)}{v(b)}. \qquad \text{(B.6)}\end{aligned}$$

and where $\theta_*(b)$ is the Bayes estimate of θ for the prior distribution $\pi_0(\theta,\tau)$:

$$\theta_*(b) = \frac{\widehat{\theta}(b)+g_0\theta_0\iota_{K_1}}{g_0+1}. \qquad \text{(B.7)}$$

So $\pi_0^*(\theta \mid g_0)$ is the pdf of a multivariate t-distribution with mean vector $\theta_*(b)$, variance-covariance matrix $\left(\frac{\xi_{0,\theta}M_{0,\theta}^{-1}}{NT-2}\right)$ and degrees of freedom $(N(T-1))$ (see Bauwens *et al.* (2005)). $q^*(\theta \mid g_0)$ is defined equivalently by:

$$\widehat{q}^*(\theta \mid g_0) = \int_0^\infty \widehat{q}^*(\theta,\tau \mid g_0)\,d\tau = L_{\widehat{q}}(b)\int_0^\infty \tau^{\left(\frac{N(T-1)+K_1}{2}-1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\widehat{q},\theta}\right)d\tau.$$

Then $q^*(\theta)$ is given by

$$q^*(\theta \mid g_0) = \widetilde{H}_q \frac{(\widehat{g}+1)^{K_1/2}}{\left\{(\widehat{g}+1)\frac{\left(\theta-\widehat{\theta}_{EB}(b)\right)'\Lambda_X\left(\theta-\widehat{\theta}_{EB}(b)\right)}{v(b)} + \left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\theta}(b)-\widehat{\theta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(\mu)-\widehat{\theta}_q\iota_{K_1}\right)}{v(b)} + 1\right\}^{\frac{N(T-1)+K_1}{2}}},$$

with

$$\begin{aligned}\widetilde{H}_q &= \frac{\Gamma\left(\frac{N(T-1)+K_1}{2}\right)|\Lambda_X|^{1/2}}{\pi^{K_1/2}\Gamma\left(\frac{N(T-1)}{2}\right)v(b)^{K_1/2}} \\ &\quad\times\left(1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\theta}(b)-\widehat{\theta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(b)-\widehat{\theta}_q\iota_{K_1}\right)}{v(b)}\right)^{\frac{N(T-1)}{2}}.\end{aligned}$$

Notice that $q^*(\theta \mid g_0)$ is the pdf of a multivariate t-distribution with mean vector $\widehat{\theta}_{EB}(b)$, variance-covariance matrix $\left(\frac{\xi_{q,\theta}M_{q,\theta}^{-1}}{N(T-1)-2}\right)$ and degrees of freedom $(N(T-1))$ with

$$\xi_{q,\theta} = 1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\theta}(b)-\widehat{\theta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\theta}(b)-\widehat{\theta}_q\iota_{K_1}\right)}{v(b)} \text{ and } M_{q,\theta} = \left(\frac{(\widehat{g}+1)}{v(\theta)}\right)\Lambda_X. \quad \text{(B.8)}$$

The mean of the ML-II posterior density of θ is then:

$$\begin{aligned}\widehat{\theta}_{ML-II} &= E\left[\widehat{\pi}^*(\theta|g_0)\right] \quad \text{(B.9)}\\ &= \widehat{\lambda}_{\theta,g_0}E\left[\pi_0^*(\theta|g_0)\right] + \left(1-\widehat{\lambda}_{\theta,g_0}\right)E\left[\widehat{q}^*(\theta|g_0)\right] \\ &= \widehat{\lambda}_{\theta,g_0}\theta_*(b|g_0) + \left(1-\widehat{\lambda}_{\theta,g_0}\right)\widehat{\theta}_{EB}(b|g_0).\end{aligned}$$

The ML-II posterior density of θ, given b and g_0 is a shrinkage estimator. It is a weighted average of the Bayes estimator $\theta_*(b|g_0)$ under base prior g_0 and the data-dependent empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. If the base prior is consistent with the data, the weight $\widehat{\lambda}_{\theta,g_0} \to 1$ and the ML-II posterior density of θ gives more weight to the posterior $\pi_0^*(\theta|g_0)$ derived from the elicited prior. In this case $\widehat{\theta}_{ML-II}$ is close to the Bayes estimator $\theta_*(b|g_0)$. Conversely, if the base prior is not consistent with the data, the weight $\widehat{\lambda}_{\theta,g_0} \to 0$ and the ML-II posterior density of θ is then close to the posterior $\widehat{q}^*(\theta|g_0)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. The ability of the ε-contamination model to extract more information from the data is what makes it superior to the classical Bayes estimator based on a single base prior.[5]

B.2. The second step of the robust Bayesian estimator

Let $\widetilde{y} = y - Z\theta$. Moving along the lines of the first step, the ML-II posterior density of b is given by:

$$\widehat{\pi}^*(b|h_0) = \widehat{\lambda}_{b,h_0}\pi_0^*(b|h_0) + \left(1-\widehat{\lambda}_{b,h_0}\right)\widehat{q}^*(b|h_0) \quad \text{(B.10)}$$

[5]Following Berger (1985), Baltagi et al. (2018) have derived the analytical ML-II posterior variance-covariance matrix of θ (see the supplementary appendix of Baltagi et al. (2018)).

with

$$\widehat{\lambda}_{b,h_0} = \left[1 + \frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{h}}{\widehat{h}+1}}{\frac{h_0}{h_0+1}}\right)^{NK_2/2}\left(\frac{1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{b_0}}{1-R^2_{b_0}}\right)}{1+\left(\frac{\widehat{h}}{\widehat{h}+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)}\right)^{\frac{(T-1)N}{2}}\right]^{-1},$$

where

$$R^2_{b_0} = \frac{(\widehat{b}(\theta) - b_0\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - b_0\iota_{NK_2})}{(\widehat{b}(\theta) - b_0\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - b_0\iota_{NK_2}) + v(\theta)},$$

$$R^2_{\widehat{b}_q} = \frac{(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})}{(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2}) + v(\theta)},$$

with $\widehat{b}(\theta) = \Lambda_D^{-1}D'\widetilde{y}$ and $v(\theta) = (\widetilde{y} - D\widehat{b}(\theta))'(\widetilde{y} - D\widehat{b}(\theta))$,

$$\widehat{b}_q = \left(\iota'_{NK_2}\Lambda_D\iota_{NK_2}\right)^{-1}\iota'_{NK_2}\Lambda_D\widehat{b}(\theta) \tag{B.11}$$

and

$$\begin{aligned}\widehat{h}_q &= \min(h_0, h^*)\\ \text{with } h^* &= \max\left[\left(\frac{((T-1)N - NK_2)}{NK_2}\frac{(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})}{v(\theta)} - 1\right)^{-1}, 0\right]\\ &= \max\left[\left(\frac{((T-1)N - NK_2)}{NK_2}\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right) - 1\right)^{-1}, 0\right].\end{aligned}$$

$\pi_0^*(b|h_0)$ is the pdf of a multivariate t-distribution with mean vector $b_*(\theta|h_0)$, variance-covariance matrix $\left(\frac{\xi_{0,b}M_{0,b}^{-1}}{(T-1)N-2}\right)$ and degrees of freedom $((T-1)N)$ with

$$M_{0,b} = \frac{(h_0+1)}{v(\theta)}\Lambda_D \text{ and } \xi_{0,b} = 1 + \left(\frac{h_0}{h_0+1}\right)\frac{(\widehat{b}(\theta) - b_0\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - b_0\iota_{NK_2})}{v(\theta)}. \tag{B.12}$$

$b_*(\theta|h_0)$ is the Bayes estimate of b for the prior distribution $\pi_0(b, \tau|h_0)$:

$$b_*(\theta|h_0) = \frac{\widehat{b}(\theta) + h_0 b_0\iota_{NK_2}}{h_0+1}. \tag{B.13}$$

$q^*(b|h_0)$ is the pdf of a multivariate t-distribution with mean vector $\widehat{b}_{EB}(\theta|h_0)$, variance-covariance matrix $\left(\frac{\xi_{1,b}M_{1,b}^{-1}}{(T-1)N-2}\right)$ and degrees of freedom $((T-1)N)$ with

$$\xi_{1,b} = 1 + \left(\frac{\widehat{h}_q}{\widehat{h}_q+1}\right)\frac{(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})'\Lambda_D(\widehat{b}(\theta) - \widehat{b}_q\iota_{NK_2})}{v(\theta)} \text{ and } M_{1,b} = \left(\frac{\widehat{h}+1}{v(\theta)}\right)\Lambda_D$$

and where $\widehat{b}_{EB}(\theta|h_0)$ is the empirical Bayes estimator of b for the contaminated prior distribution $q(b,\tau|h_0)$:

$$\widehat{b}_{EB}(\theta|h_0) = \frac{\widehat{\theta}(b) + \widehat{h}_q \widehat{b}_q \iota_{NK_2}}{\widehat{h}_q + 1}. \tag{B.14}$$

The mean of the ML-II posterior density of b is hence given by:

$$\widehat{b}_{ML-II} = \widehat{\lambda}_{b,h_0} b_*(\theta|h_0) + \left(1 - \widehat{\lambda}_{b,h_0}\right) \widehat{b}_{EB}(\theta|h_0). \tag{B.15}$$

The ML-II posterior variance-covariance matrix of b can be derived in a similar fashion[6] to that of $\hat{\theta}_{ML-II}$.

B.3. Estimating the ML-II posterior variance-covariance matrix

Many have raised concerns about the unbiasedness of the posterior variance-covariance matrices of $\widehat{\theta}_{ML-II}$ and $\widehat{b}_{ML-II}$. Indeed, they will both be biased towards zero as $\widehat{\lambda}_{\theta,g_0}$ and $\widehat{\lambda}_{b,h_0} \to 0$ and converge to the empirical variance which is known to underestimate the true variance (see *e.g.* Berger and Berliner (1986); Gilks et al. (1997); Robert (2007)). Consequently, the assessment of the performance of either $\widehat{\theta}_{ML-II}$ or $\widehat{b}_{ML-II}$ using standard quadratic loss functions cannot be conducted using the analytical expressions. What is needed is an unbiased estimator of the true ML-II variances. Baltagi et al. (2018) have proposed two different strategies to approximate these, each with different desirable properties: MCMC with multivariate t-distributions or block resampling bootstrap. They have shown that fortunately, one needs as few as 20 bootstrap samples to achieve acceptable results[7]. Here, as in Baltagi et al. (2018, 2021), we will use the same individual block resampling bootstrap method. Following Bellman et al. (1989); Andersson and Karlsson (2001); Kapetanios (2008), and for an $(N \times (T-1))$ matrix Y, individual block resampling consists in drawing an $(N \times (T-1))$ matrix Y^{BR} whose rows are obtained by resampling those of Y with replacement. Conditionally on Y, the rows of Y^{BR} are independent and identically distributed. The following algorithm is used to approximate the variance matrices:

1. Loop over BR samples
2. In the first step, compute the mean of the ML-II posterior density of θ using our initial shrinkage procedure

$$\begin{aligned} \widehat{\theta}_{ML-II,br} &= E\left[\widehat{\pi}^*(\theta|g_0)\right] \\ &= \widehat{\lambda}_{\theta,g_0} \theta_*(b|g_0) + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) \widehat{\theta}_{EB}(b|g_0). \end{aligned}$$

3. In the second step, compute the mean of the ML-II posterior density of b:

$$\widehat{b}_{ML-II,br} = \widehat{\lambda}_{b,h_0} b_*(\theta|h_0) + \left(1 - \widehat{\lambda}_{b,h_0}\right) \widehat{b}_{EB}(\theta|h_0)$$

[6] See the supplementary appendix of Baltagi et al. (2018).

[7] For convenience, the number of bootstrap samples BR is relatively small compared to the sample size N. Increasing the number of bootstrap samples does not change the results but increases the computation time considerably.

4. Once the BR bootstraps are completed, use the $(K_1 \times BR)$ matrix of coefficients $\theta^{(BR)}$ and the $(N \times BR)$ matrix of coefficients $b^{(BR)}$ to compute:

$$\widehat{\theta}_{ML-II} = E\left[\theta^{(BR)}\right] , \qquad \widehat{\sigma}_{\theta_{ML-II}} = \sqrt{diag\left(Var\left[\theta^{(BR)}\right]\right)}$$
$$\widehat{b}_{ML-II} = E\left[b^{(BR)}\right] , \qquad \widehat{\sigma}_{b_{ML-II}} = \sqrt{diag\left(Var\left[b^{(BR)}\right]\right)}$$

We can also use a mixture of multivariate skewed (or non-skewed) t-distributions (see Baltagi et al. (2021)). The ML-II posterior density of θ in (B.17) is a two-component finite mixture of multivariate t-distributions whose location parameters and scale matrices are given in (B.7) and in (B.8). Following McLachlan and Lee (2013) and Baltagi et al. (2021), one can generate mixture of multivariate skewed (or non-skewed) t-distributions via an EM Algorithm approach. Thus, generating 1000 (or more) random samples of K_1 (or NK_2)-dimensional multivariate t observations with location parameters, scale matrices given in (B.7) and in (B.8) and degrees of freedom $N(T-1)$, allows to sample a mixture of the two components to get 1000 (or more) random vectors of θ_{ML-II}. The latter can then be used to compute the variances of the K_1 (or NK_2) parameters.[8] Using this device reduces the computation time by at least 90% in all cases, although small discrepancies with the bootstrapped standard errors may occur in specific cases.

B.4. A simple and efficient way to drastically reduce the computation time of our Bayesian two-stage two-step estimator.

The first stage of the Gaussian dynamic linear mixed model (eq.(6) in the main text) is given by

$$y = Z\theta + Db + u \ , \ u \sim N(0, \tau^{-1} I_{N(T-1)}) \tag{B.16}$$

where y is $(N(T-1) \times 1)$. Z is $(N(T-1) \times K_1)$, D is $(N(T-1) \times K_2)$ and u is $(N(T-1) \times 1)$. The mean of the ML-II posterior density of θ is:

$$\begin{aligned} \widehat{\theta}_{ML-II} &= E\left[\widehat{\pi}^*\left(\theta|g_0\right)\right] = \widehat{\lambda}_{\theta,g_0} E\left[\pi_0^*\left(\theta|g_0\right)\right] + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) E\left[\widehat{q}^*\left(\theta|g_0\right)\right] \\ &= \widehat{\lambda}_{\theta,g_0}\theta_*(b|g_0) + \left(1 - \widehat{\lambda}_{\theta,g_0}\right)\widehat{\theta}_{EB}\left(b|g_0\right). \end{aligned} \tag{B.17}$$

Baltagi et al. (2018) have shown that the ML-II posterior variance-covariance matrix of θ is given by

$$\begin{aligned} Var\left(\widehat{\theta}_{ML-II}\right) &= \widehat{\lambda}_{\theta,g_0} Var\left[\pi_0^*\left(\theta \mid g_0\right)\right] + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) Var\left[\widehat{q}^*\left(\theta \mid g_0\right)\right] \\ &+ \widehat{\lambda}_{\theta,g_0}\left(1 - \widehat{\lambda}_{\theta,g_0}\right)\left(\theta_*(b \mid g_0) - \widehat{\theta}_{EB}\left(b \mid g_0\right)\right)\left(\theta_*(b \mid g_0) - \widehat{\theta}_{EB}\left(b \mid g_0\right)\right)' \end{aligned} \tag{B.18}$$

Since this expression underestimates the true variance, Baltagi et al. (2018, 2021) have proposed two different strategies to approximate it, each with different desirable properties: MCMC with multivariate t-distributions or individual block resampling bootstrap. They have shown that fortunately, one needs as few as 20 bootstrap samples to achieve acceptable results. They also showed

[8] See section C in the supplementary material for more details.

that the bootstrap method had some advantages over the MCMC method, especially in terms of computation time.

Computation times can be improved by using the Choleski decomposition for matrices inversion[9] for all the tested worlds (RE, Chamberlain, Hausman-Taylor, CCE). Additionally, multivariate normal random vectors in the common correlated effects (CCE) models could make use of the sparse matrices.[10] Yet, the efficiency gains would be relatively modest given the number of bootstrap draws that need to be generated.

An alternative approach arises if we exploit the intrinsic features of the distributions of the Bayes estimate $\theta_*(b|g_0)$ for the prior distribution $\pi_0(\theta,\tau)$ and the empirical Bayes estimate $\widehat{\theta}_{EB}(b|g_0)$ for the contaminated prior distribution $q(\theta,\tau)$.

We have shown that $\pi_0^*(\theta|g_0)$ is the pdf of a multivariate t-distribution $t_{K_1}\left(\theta_*(b|g_0),\Sigma_{\theta_*(b|g_0)},N(T-1)\right)$ and $\widehat{q}^*(\theta)$ is the pdf of a multivariate t-distribution $t_{K_1}\left(\widehat{\theta}_{EB}(b|g_0),\Sigma_{\widehat{\theta}_{EB}(b|g_0)},N(T-1)\right)$ where the mean vectors $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$ are given by

$$\theta_*(b|g_0)=\frac{\widehat{\theta}(b)+g_0\theta_0\iota_{K_1}}{g_0+1}\ ,\ \widehat{\theta}_{EB}(b|g_0)=\frac{\widehat{\theta}(b)+\widehat{g}_q\widehat{\theta}_q\iota_{K_1}}{\widehat{g}_q+1}.$$

and the variance-covariance matrices $\Sigma_{\theta_*(b|g_0)}$ and $\Sigma_{\widehat{\theta}_{EB}(b|g_0)}$ are given by

$$\begin{aligned}\Sigma_{\theta_*(b|g_0)} &= \left(\frac{\xi_{0,\theta}M_{0,\theta}^{-1}}{N(T-1)-2}\right)\text{ with }M_{0,\theta}=\frac{(g_0+1)}{v(b)}\Lambda_Z\text{ and }\xi_{0,\theta}=1+\left(\frac{g_0}{g_0+1}\right)\left(\frac{R_{\theta_0}^2}{1-R_{\theta_0}^2}\right).\\ \Sigma_{\widehat{\theta}_{EB}(b|g_0)} &= \left(\frac{\xi_{q,\theta}M_{q,\theta}^{-1}}{N(T-1)-2}\right)\text{ with }M_{q,\theta}=\left(\frac{(\widehat{g}_q+1)}{v(b)}\right)\Lambda_Z\text{ and }\xi_{q,\theta}=1+\left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R_{\widehat{\theta}_q}^2}{1-R_{\widehat{\theta}_q}^2}\right).\end{aligned}$$

Thus the ML-II posterior density of θ in (B.17) is a two-component finite mixture of multivariate t-distributions. Its pdf is given by

$$\pi\left(\tilde{\theta}_{ML-II}\right) = \sum_{h=1}^{2}\varrho_h\pi_h\left(\widehat{\theta}_{ML-II},m_h,\Sigma_h,\nu_h\right). \tag{B.19}$$

where $\pi_h\left(\widehat{\theta}_{ML-II},m_h,\Sigma_h,\nu_h\right)$ denotes the h-th pdf of the mixture model with location parameter m_h, scale matrix Σ_h and degrees of freedom ν_h. The mixing proportions satisfy $\varrho_h\geq 0$ $(h=1,2)$ and $\sum_{h=1}^{2}\varrho_h=1$. In our case, $\nu_h=N(T-1)$, $\forall h$, $m_1=\theta_*(b|g_0)$, $m_2=\widehat{\theta}_{EB}(b|g_0)$, $\Sigma_1=\Sigma_{\theta_*(b|g_0)}$, $\Sigma_2=\Sigma_{\widehat{\theta}_{EB}(b|g_0)}$, $\varrho_1=\widehat{\lambda}_{\theta,g_0}$ and $\varrho_2=1-\widehat{\lambda}_{\theta,g_0}$.

Derivations of the location parameter and the scale matrix of a mixture of multivariate t-distributions is a very difficult task (see for instance Walker and Saw (1978), Peel and McLachlan (2000), Kotz and Nadarajah (2004), McLachlan and Peel (2004), Mengersen et al. (2011) among others). Parameter estimates of the mixture of t-distributions is generally obtained via an EM algorithm. McLachlan and Lee (2013) have proposed a `EMMIXuskew R` package for generating and

[9] The `R` function `chol2inv(chol(X))` instead of the standard function `solve(X)` for a definite symmetric positive X matrix.

[10] Using the `rmvn.sparse` command in the `sparseMVN R` package.

fitting mixture of multivariate skewed (and non-skewed) t distributions via the EM Algorithm. Based on the command `rfmmst` of this package, and given the parameters of the two components defined above, one can generate 1000 (or more) random samples of K_1-dimensional multivariate t observations with location parameter m_h, scale matrix Σ_h and degrees of freedom ν_h for $h = 1, 2$, and hence sample from the mixture of these two components to generate as many random vectors of $\tilde{\theta}_{ML-II}$. The variances of the K_1 parameters can then be computed over these 1000 (or more) random samples (see also Baltagi et al. (2021)).

After extensive experimentation, it was found that the estimated variances were slightly underestimated compared to those obtained with the bootstrap method. We therefore propose to correct the variances with the following multiplicative factor: $\sqrt{k_2}\left(1+\sqrt{\hat{r}}\right)^2$. In the RE, Chamberlain and Hausman-Taylor worlds, $\hat{r} = \hat{\sigma}_\mu^2/\left(\hat{\sigma}_\mu^2 + \hat{\sigma}_u^2\right)$ is the fraction of the variance $\left(\hat{\sigma}_\mu^2 + \hat{\sigma}_u^2\right)$ due to the specific effects μ_i, with $\hat{\sigma}_u^2 = \hat{\tau}^{-1}$ and k_2 is the number of covariates in D in eq.(B.16).[11] In the correlated common effects worlds (CCE) worlds,[12] $\hat{\sigma}_\mu^2 = Var\left[\Gamma\right]$ and the correction factor needs to be modified slightly as $\sqrt{mk_2}\left(1+\sqrt{\hat{r}}\right)^2$.

In the case of the dynamic heterogeneous panel data world with common correlated unobserved effects, the correction factor needs to be modified slightly to take into account the average over all individuals. The corrected variance of $\tilde{\theta}_{ML-II}$ is computed as:[13]

$$\begin{aligned} Var\left[\tilde{\theta}_{ML-II}\right] &= \frac{1}{N}\sum_{i=1}^{N} Var\left[\tilde{\theta}_{i,ML-II}\right]_{cor} \\ &\quad \text{with } Var\left[\tilde{\theta}_{i,ML-II}\right]_{cor} = Var\left[\tilde{\theta}_{i,ML-II}\right]\sqrt{mN}\left(1+\sqrt{\hat{r}_i}\right)^2 \\ &\quad \text{and } \hat{r}_i = Var\left[\hat{\gamma}_i\right]/\left(Var\left[\hat{\gamma}_i\right] + Var\left[\hat{u}_i\right]\right). \end{aligned}$$

This additional correction factor, $\sqrt{N}$, is somewhat reminiscent of the results from Theorem 3 of Chudik and Pesaran (2015a), which shows that the convergence rate of the CCEMG estimator $\hat{\theta}_{CCEMG}$ of θ is $\sqrt{N}$ due to the heterogeneity of the coefficients. Moreover, Chudik and Pesaran (2015a) show that the ratio $N/T \to \kappa_1$, for some constant $0 < \kappa_1 < \infty$, is required for the derivation of the asymptotic distribution of $\hat{\theta}_{CCEMG}$ due to the time-series bias and that it is unsuitable for panels with T being small relative to N.[14]

The differences between the estimates for the Bayesian two-step (*B2S2S*) estimator with `se_boot`, or with `se_mixt`, and those obtained with QMLE, two-step QMLE, CCEP or CCEMG are marginal when the sample size is significantly increased. The difference in computation times is impressive. The advantage of our Bayesian two-stage estimator is pretty obvious. And because it has little computing time, it should be valuable for applied econometricians.

[11] In other words, $k_2 = 1$ for the RE, Chamberlain and Hausman-Taylor worlds. $k_2 = m$ for the common trends world and the common correlated effects world.

[12] Γ is given in section 4.1.3 in the main text.

[13] γ_i is given in section 4.1.3 in the main text.

[14] In their simulation study, Chudik and Pesaran (2015a) use $0.2 \leq N/T \leq 5$. They also use a jackknife bias correction and a recursive mean adjustment correction of the CCEMG estimator.

C. Full Bayesian estimator for the random effects world

We also derive the full Bayesian estimator for RE world.

C.1. Gibbs sampling for the RE world

We run full Bayesian estimates (Gibbs sampling) on the RE world following the works of Chib and Carlin (1999), Koop (2003), Chib (2008), Greenberg (2008) and Chan et al. (2019) to mention a few. They have proposed algorithms for the three-stage hierarchical models in a standard RE world and we extend this specification to the dynamic space-time case. Pooling the N individuals for one time period, our initial specification is

$$y_t = \phi y_{t-1} + \rho y_t^* + \delta y_{t-1}^* + X_t\beta + D_t b + u_t \ , \ t = 2, ..., T, \tag{C.20}$$

where y_t is the N-dimensional vector of the dependent variable, $y_t^* = W_N y_t$ and $y_{t-1}^* = W_N y_{t-1}$, X_t is $(N \times K_1)$ and u_t is $(N \times 1)$. In the RE world, the $(N \times K_2)$ matrix D_t is an identity matrix I_N and the $(K_2 \times 1)$ vector b is replaced by the $(N \times 1)$ vector $\mu = (\mu_1, \mu_2, ..., \mu_N)'$ of time-invariant effects. It can be written as the following three-stage hierarchy:

$$\begin{cases} \text{First stage :} & y_t = \phi y_{t-1} + \rho y_t^* + \delta y_{t-1}^* + X_t\beta + \mu + u_t \\ & \text{with } u_t \sim N_N(0, \Sigma_{u_t}) \text{ and } \Sigma_{u_t} = \tau^{-1} I_N \\ \text{Second stage :} & \phi \sim U(-1,1),\ \rho \sim U(-1,1),\ \delta \sim U(-1,1),\ \beta \sim N_{K_1}(\beta_0, B_0) \\ & \text{and } \mu \sim N_N(0, \Sigma_\mu) \text{ with } \Sigma_\mu = \sigma_\mu^2 I_N \\ \text{Third stage :} & \tau \sim G\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \text{and } \sigma_\mu^{-2} \sim G\left(\frac{\gamma_0}{2}, \frac{\eta_0}{2}\right). \end{cases}$$

According to Parent and LeSage (2010) and Debarsy et al. (2012), the dependent parameters ϕ, ρ and δ follow independent uniform distributions $U(-1,1)$. $N_{K_1}(.)$ is the multivariate normal distribution and $G(.)$ is the Gamma distribution.
We can define the conditional posterior distributions within the Gibbs sampler of the previous model for the RE world.[15]

1. We choose diffuse priors with the following hyperparameters $\beta_0 = 0_{K_1}$, $B_0 = 10^2 I_{K_1}$, $\alpha_0 = 2$, $\delta_0 = 200$, $\gamma_0 = 2$, $\eta_0 = 200$ such that the means of the precision τ and σ_μ^{-2} are $E[\tau] = E\left[\sigma_\mu^{-2}\right] = 10^{-2}$ and their variances[16] are $Var[\tau] = Var\left[\sigma_\mu^{-2}\right] = 10^{-4}$.
2. We draw initial values of:

$$\begin{array}{ll} \beta^{(0)} \sim N_{K_1}(\beta_0, B_0) & , \tau^{(0)} \sim G\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \\ \sigma_\mu^{-2(0)} \sim G\left(\frac{\gamma_0}{2}, \frac{\eta_0}{2}\right) & , \mu^{(0)} \sim N_N\left(0, \sigma_\mu^{2(0)} I_N\right) \end{array}$$

[15] See section (C.2) for the derivations.

[16] A random variable x follows a Gamma distribution $G(\alpha, \beta)$ with shape α and scale β if its pdf can be written as

$$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

Its mean and variance are given by $E[x] = \frac{\alpha}{\beta}$ and $Var[x] = \frac{\alpha}{\beta^2}$.

3. At the d^{th} (for $d = 1, ..., \Delta$) draw, we sample:

$$\begin{aligned}
\tau^{(d)} &\sim G\left(\frac{\alpha_1}{2}, \frac{{\delta_1}^{(d)}}{2}\right) \\
{\sigma_\mu^{-2}}^{(d)} &\sim G\left(\frac{\gamma_1}{2}, \frac{{\eta_1}^{(d)}}{2}\right) \\
\mu^{(d)} &\sim N_N\left(\bar{\mu}^{(d)}, B_\mu^{(d)}\right) \\
\beta^{(d)} &\sim N_{K_1}\left(\bar{\beta}^{(d)}, B_\beta^{(d)}\right) \\
\phi^{(d)} &\sim N\left(\bar{\phi}^{(d)}, B_\phi^{(d)}\right) \\
\rho^{(d)} &\sim N\left(\bar{\rho}^{(d)}, B_\rho^{(d)}\right) \\
\delta^{(d)} &\sim N\left(\bar{\delta}^{(d)}, B_\delta^{(d)}\right)
\end{aligned}$$

where

$$\begin{aligned}
{\delta_1}^{(d)} &= \delta_0 + \sum_{t=2}^{T}\left(y_t - Z_t\theta^{(d-1)} - \mu^{(d-1)}\right)' \times \left(y_i - Z_t\theta^{(d-1)} - \mu^{(d-1)}\right) \\
&\quad \text{with } Z_t = \left[y_{t-1}, y_t^*, y_{t-1}^*, X_t\right] \text{ and } \theta^{(d-1)} = \left(\phi^{(d-1)}, \rho^{(d-1)}, \delta_y^{(d-1)}, \beta^{(d-1)'}\right)' \\
\eta_1^{(d)} &= \eta_0 + {\mu'}^{(d-1)}\mu^{(d-1)} \\
B_\mu^{(d)} &= \left[\left((T-1)\tau^{(d)} + {\sigma_\mu^{-2}}^{(d)}\right) I_N\right]^{-1} \\
\overline{\mu}^{(d)} &= B_\mu^{(d)}\left[\tau^{(d)}\sum_{t=2}^{T}\tilde{\tilde{y}}_t^{(d-1)}\right] \\
&\quad \text{where } \tilde{\tilde{y}}_t^{(d-1)} = y_t - \phi^{(d-1)}y_{t-1} - \rho^{(d-1)}y_t^* - \delta_y^{(d-1)}y_{t-1}^* - X_t\beta^{(d-1)} \\
B_\beta^{(d)} &= \left[\sum_{t=2}^{T}\left(X_t' {B_\psi^{-1}}^{(d)} X_t\right) + B_0^{-1}\right]^{-1} \\
&\quad \text{where } B_\psi^{(d)} = (\sigma_\mu^{2^{(d)}} + {\tau^{-1}}^{(d)})I_N \\
\overline{\beta}^{(d)} &= B_\beta^{(d)}\left[\sum_{t=2}^{T}\left(X_t' {B_\psi^{-1}}^{(d)} \tilde{y}_t\right) + B_0^{-1}\beta_0\right] \\
B_\phi^{(d)} &= \left[{\tau^{-1}}^{(d)}\sum_{t=2}^{T} y_{t-1}' y_{t-1}\right]^{-1} \\
\overline{\phi}^{(d)} &= B_\phi^{(d)}\left[{\tau^{-1}}^{(d)}\sum_{t=2}^{T} y_{t-1}' {y_t^{(\phi)}}^{(d)}\right] \\
&\quad \text{where } {y_t^{(\phi)}}^{(d)} = y_t - \rho^{(d-1)}y_t^* - \delta^{(d-1)}y_{t-1}^* - X_t\beta^{(d)} - \mu^{(d)}
\end{aligned}$$

$$
\begin{aligned}
B_{\rho}^{(d)} &= \left[\tau^{-1^{(d)}} \sum_{t=2}^{T} y_t^{*'} y_t^{*}\right]^{-1} \\
\overline{\rho}^{(d)} &= B_{\rho}^{(d)} \left[\tau^{-1^{(d)}} \sum_{t=2}^{T} y_t^{*'} y_t^{(\rho)^{(d)}}\right] \\
&\quad \text{where } y_t^{(\rho)^{(d)}} = y_t - \phi^{(d)} y_{t-1} - \delta_y^{(d-1)} y_{t-1}^{*} - X_t \beta^{(d)} - \mu^{(d)} \\
B_{\delta}^{(d)} &= \left[\tau^{-1^{(d)}} \sum_{t=2}^{T} y_{t-1}^{*'} y_{t-1}^{*}\right]^{-1} \\
\overline{\delta}_y^{(d)} &= B_{\delta}^{(d)} \left[\tau^{-1^{(d)}} \sum_{t=2}^{T} y_{t-1}^{*'} y_t^{(\delta)^{(d)}}\right] \\
&\quad \text{where } y_t^{(\delta)^{(d)}} = y_t - \phi^{(d)} y_{t-1} - \rho^{(d)} y_t^{*} - X_t \beta^{(d)} - \mu^{(d)} \\
\alpha_1 &= \alpha_0 + N(T-1) \\
\gamma_1 &= \gamma_0 + N
\end{aligned}
$$

For the Gibbs sampling, we run $\Delta = 1,000$ draws and we burn the $\Delta_{burn} = 500$ first draws. We store all the vectors β, μ and the scalars ϕ, ρ, δ, σ_{ϵ}^2 and σ_{μ}^2 for the $\Delta^*(= \Delta - \Delta_{burn})$ draws. When the Δ draws are completed, we compute their posterior means, their posterior standard errors, their RMSEs, their 95% HPDIs, their numerical standard errors (nse) and convergence diagnostics (cd) on the Δ^* last draws.

C.2. Derivation of the posterior densities of the Gibbs sampling for the RE world

Since the three-stage hierarchy is written as

$$
\begin{cases}
\text{First stage :} & y_t = \phi y_{t-1} + \rho y_t^{*} + \delta y_{t-1}^{*} + X_t \beta + \mu + u_t \\
& \text{with } u_t \sim N_N(0, \Sigma_{u_t}) \text{ and } \Sigma_{u_t} = \tau^{-1} I_N \\
\text{Second stage :} & \phi \sim U(-1,1),\ \rho \sim U(-1,1),\ \delta \sim U(-1,1),\ \beta \sim N_{K_1}(\beta_0, B_0) \\
& \text{and } \mu \sim N_N(0, \Sigma_{\mu}) \text{ with } \Sigma_{\mu} = \sigma_{\mu}^2 I_N \\
\text{Third stage :} & \tau \sim G\left(\tfrac{\alpha_0}{2}, \tfrac{\delta_0}{2}\right) \text{and } \sigma_{\mu}^{-2} \sim G\left(\tfrac{\gamma_0}{2}, \tfrac{\eta_0}{2}\right).
\end{cases}
$$

and as the known hyperparameters are: β_0, B_0, α_0, δ_0, γ_0 and η_0, then the posterior distribution is proportional to:[17]

$$
\begin{aligned}
\pi\left(\phi, \rho, \delta, \beta, \mu, \tau, \sigma_{\mu}^2 \mid y, y^{*}, y_{-1}^{*}, X\right) \quad \propto \quad & |\Sigma_{u_t}|^{-\frac{(T-1)}{2}} \exp\left[-\frac{1}{2} \sum_{t=2}^{T} \left(y_t - Z_t \theta - \mu\right)' \Sigma_{u_t}^{-1} \left(y_t - Z_t \theta - \mu\right)\right] \\
& \times \exp\left[-\frac{1}{2} \left(\beta - \beta_0\right)' B_0^{-1} \left(\beta - \beta_0\right)\right] \times \tau^{\frac{\alpha_0}{2} - 1} \exp\left[-\frac{\tau \delta_0}{2}\right] \\
& \times \sigma_{\mu}^{\frac{-2N}{2}} \exp\left[-\frac{\sigma_{\mu}^{-2} \mu' \mu}{2}\right] \times \sigma_{\mu}^{-2\left[\frac{\gamma_0}{2} - 1\right]} \exp\left[-\frac{\sigma_{\mu}^{-2} \eta_0}{2}\right]
\end{aligned}
$$

[17] Since the pdf of the dependent parameters are: $p(\phi) = p(\rho) = p(\delta) = 1/2$.

where $Z_t\theta = \phi y_{t-1} + \rho y_t^* + \delta y_{t-1}^* + X_t\beta$.
The posterior distribution of the precision τ is given by:

$$\begin{aligned} & |\Sigma_{u_t}|^{-\frac{(T-1)}{2}} \exp\left[-\frac{1}{2}\sum_{t=2}^{T}(y_t - Z_t\theta - \mu)'\,\Sigma_{u_t}^{-1}\,(y_t - Z_t\theta - \mu)\right] \\ & \quad \times \tau^{\frac{\alpha_0}{2}-1}\exp\left[-\frac{\tau\delta_0}{2}\right] \\ = \quad & \tau^{\frac{\alpha_0+N(T-1)}{2}-1}\exp\left[-\frac{\tau}{2}\left\{\delta_0 + \sum_{t=2}^{T}(y_t - Z_t\theta - \mu)'\,(y_t - Z_t\theta - \mu)\right\}\right] \end{aligned}$$

then

$$\begin{aligned} \tau \quad \sim \quad & G\left(\frac{\alpha_1}{2}, \frac{\delta_1}{2}\right) \\ & \text{with } \alpha_1 = \alpha_0 + N(T-1) \\ & \text{and } \delta_1 = \delta_0 + \sum_{t=2}^{T}(y_t - Z_t\theta - \mu)'\,(y_t - Z_t\theta - \mu) \end{aligned} \tag{C.21}$$

The posterior distribution of the precision σ_μ^{-2} is given by:

$$\begin{aligned} & \sigma_\mu^{\frac{-2N}{2}}\exp\left[-\frac{\sigma_\mu^{-2}\mu'\mu}{2}\right] \times \sigma_\mu^{-2\left[\frac{\gamma_0}{2}-1\right]}\exp\left[-\frac{\sigma_\mu^{-2}\eta_0}{2}\right] \\ = \quad & \sigma_\mu^{-2\left[\frac{N+\gamma_0}{2}-1\right]}\exp\left[-\frac{\sigma_\mu^{-2}}{2}(\eta_0 + \mu'\mu)\right] \end{aligned}$$

then

$$\begin{aligned} \sigma_\mu^{-2} \quad \sim \quad & G\left(\frac{\gamma_1}{2}, \frac{\eta_1}{2}\right) \\ & \text{with } \gamma_1 = \gamma_0 + N \\ & \text{and } \eta_1 = \eta_0 + \mu'\mu \end{aligned} \tag{C.22}$$

Following Chib and Carlin (1999) and Greenberg (2008), it is preferable to sample β and μ in one block as $\pi\left(\beta, \mu \mid y, y^*, y_{-1}^*, X, \phi, \rho, \delta, \tau, \sigma_\mu^2\right)$ rather than in two blocks $\pi\left(\beta \mid y, y_{-1}^*, X, \mu, \tau, \phi, \rho, \delta, \sigma_\mu^2\right)$ and $\pi\left(\mu \mid y, y_{-1}^*, X, \beta, \phi, \rho, \delta, \tau, \sigma_\mu^2\right)$, because of potential correlation between the two. This is done by using:

$$\pi\left(\beta, \mu \mid y, y^*, y_{-1}^*, X, \phi, \rho, \delta, \tau, \sigma_\mu^2\right) \quad = \quad \pi\left(\beta \mid y, y_{-1}^*, X, \mu, \tau, \phi, \rho, \delta, \sigma_\mu^2\right) \times \pi\left(\mu \mid y, y_{-1}^*, X, \beta, \phi, \rho, \delta, \tau, \sigma_\mu^2\right)$$

The first terms on the right-hand side is obtained by integrating out the μ from $\pi\left(\beta, \mu \mid y, y^*, y_{-1}^*, X, \phi, \rho, \delta, \tau, \sigma_\mu^2\right)$. For the second term, set $\tilde{\tilde{y}}_t = \tilde{y}_t - X_t\beta$ where $\tilde{y}_t = y_t - \phi y_{t-1} - \rho y_t^* - \delta y_{t-1}^*$ and complete the square in μ.

$$\begin{aligned} \pi\left(\mu \mid y, y^*, y_{-1}^*, X, \beta, \phi, \rho, \delta, \tau, \sigma_\mu^2\right) \quad \propto \quad & \exp\left[-\frac{1}{2}\sum_{t=2}^{T}\left(\tilde{\tilde{y}}_t - \mu\right)'\,\Sigma_u^{-1}\left(\tilde{\tilde{y}}_t - \mu\right)\right] \\ & \times \exp\left[-\frac{1}{2}\mu'\Sigma_\mu^{-1}\mu\right] \end{aligned}$$

Let us consider the expressions in the exponentiations, ignoring the $\left(-\frac{1}{2}\right)$ terms:

$$\begin{aligned}
& \sum_{t=2}^{T} \left(\tilde{\tilde{y}}_t - \mu\right)' \Sigma_u^{-1} \left(\tilde{\tilde{y}}_t - \mu\right) + \mu' \Sigma_\mu^{-1} \mu \\
= & \sum_{t=2}^{T} \tilde{\tilde{y}}_t' \Sigma_u^{-1} \tilde{\tilde{y}}_t - \sum_{t=2}^{T} \tilde{\tilde{y}}_t' \Sigma_u^{-1} \mu - \sum_{t=2}^{T} \mu' \Sigma_u^{-1} \tilde{\tilde{y}}_t + \sum_{t=2}^{T} \mu' \Sigma_u^{-1} \mu + \mu' \Sigma_\mu^{-1} \mu \\
= & \mu' \left[\sum_{t=2}^{T} \left(\Sigma_u^{-1}\right) + \Sigma_\mu^{-1} \right] \mu - 2\mu' \left[\sum_{t=2}^{T} \left(\Sigma_u^{-1} \tilde{\tilde{y}}_t\right) \right] + \sum_{t=2}^{T} \tilde{\tilde{y}}_t' \Sigma_u^{-1} \tilde{\tilde{y}}_t \\
= & \mu' \left[((T-1)\tau + \sigma_\mu^{-2}) I_N \right] \mu - 2\tau \mu' \sum_{t=2}^{T} \tilde{\tilde{y}}_t + \tau \sum_{t=2}^{T} \tilde{\tilde{y}}_t' \tilde{\tilde{y}}_t
\end{aligned}$$

Since we are only concerned with the distribution of μ, and as $\Sigma_u (= \tau^{-1} I_N)$ is assumed to be known, terms that do not involve μ are all absorbed into the proportionality constant. Applying this idea to the expressions between brackets, then, the posterior distribution of the time-invariant specific effect μ is given by:

$$\begin{aligned}
\mu \quad \sim \quad & N_N \left(\overline{\mu}, B_\mu\right) \\
& \text{with } B_\mu = \left[((T-1)\tau + \sigma_\mu^{-2}) I_N \right]^{-1} \\
& \text{and } \overline{\mu} = B_\mu \left[\tau \sum_{t=2}^{T} \tilde{\tilde{y}}_t \right]
\end{aligned} \tag{C.23}$$

To find the posterior distribution of β, we write:

$$\tilde{y}_t = X_t \beta + (\mu + u_t) = X_t \beta + \psi_t$$

and integrate out μ and u_t. Then

$$E\left[\psi_t \psi_t'\right] \quad = \quad E\left[(\mu + u_t)(\mu + u_t)'\right] = E\left[\mu\mu'\right] + E\left[u_t u_t'\right]$$

As $u_t \sim N(0, \Sigma_{u_t})$ with $\Sigma_{u_t} = \tau^{-1} I_N$ and $\mu \sim N\left(0, \Sigma_\mu\right)$ with $\Sigma_\mu = \sigma_\mu^2 I_N$, then

$$E\left[\psi_t \psi_t'\right] = \Sigma_\mu + \Sigma_u = (\sigma_\mu^2 + \tau^{-1}) I_N = B_\psi$$

which implies $\tilde{y}_t \sim N\left(X_t \beta, B_\psi\right)$. It follows that

$$\begin{aligned}
\pi\left(\beta \mid y, y^*, y_{-1}^*, X, \mu, \phi, \rho, \delta, \tau, \sigma_\mu^2\right) \quad \propto \quad & \exp\left[-\frac{1}{2} \sum_{t=2}^{T} \left(\tilde{y}_t - X_t \beta\right)' B_\psi^{-1} \left(\tilde{y}_t - X_t \beta\right) \right] \\
& \times \exp\left[-\frac{1}{2} \left(\beta - \beta_0\right)' B_0^{-1} \left(\beta - \beta_0\right) \right]
\end{aligned}$$

Completing the expressions between brackets, we get:

$$\begin{aligned}
&\sum_{t=2}^{T}\left(\tilde{y}_t - X_t\beta\right)' B_\psi^{-1}\left(\tilde{y}_t - X_t\beta\right) + \left(\beta-\beta_0\right)' B_0^{-1}\left(\beta-\beta_0\right) \\
= \;& \beta'\left[\sum_{t=2}^{T}\left(X_t' B_\psi^{-1} X_t\right) + B_0^{-1}\right]\beta - 2\beta'\left[\sum_{t=2}^{T}\left(X_t' B_\psi^{-1} y_t\right) + B_0^{-1}\beta_0\right] \\
&+ \sum_{t=2}^{T}\tilde{y}_t' B_\psi^{-1}\tilde{y}_t + \beta_0' B_0^{-1}\beta_0
\end{aligned}$$

from which we have

$$\begin{aligned}
\beta \;\sim\; & N_{K_1}\left(\overline{\beta}, B_\beta\right) \\
& \text{with } B_\beta = \left[\sum_{t=2}^{T}\left(X_t' B_\psi^{-1} X_t\right) + B_0^{-1}\right]^{-1} \\
& \text{where } B_\psi = (\sigma_\mu^2 + \tau^{-1}) I_N \\
& \text{and } \overline{\beta} = B_\beta\left[\sum_{t=2}^{T}\left(X_t' B_\psi^{-1}\tilde{y}_t\right) + B_0^{-1}\beta_0\right]
\end{aligned} \tag{C.24}$$

The posterior distribution of the autoregressive time dependence parameter ϕ is proportional to:

$$\begin{aligned}
& |\Sigma_{u_t}|^{-\frac{(T-1)}{2}} \exp\left[-\frac{1}{2}\sum_{t=2}^{T}\left(y_t^{(\phi)} - \phi y_{t-1}\right)' \Sigma_{u_t}^{-1}\left(y_t^{(\phi)} - \phi y_{t-1}\right)\right] \\
& \text{where } y_t^{(\phi)} = y_t - \rho y_t^* - \delta y_{t-1}^* - X_t\beta - \mu
\end{aligned}$$

Then,

$$\begin{aligned}
\phi \;\sim\; & N\left(\overline{\phi}, B_\phi\right) \\
& \text{with } B_\phi = \left[\tau^{-1}\sum_{t=2}^{T} y_{t-1}' y_{t-1}\right]^{-1} \text{ and } \overline{\phi} = B_\phi\left[\tau^{-1}\sum_{t=2}^{T} y_{t-1}' y_t^{(\phi)}\right] \\
& \text{where } y_t^{(\phi)} = y_t - \rho y_t^* - \delta y_{t-1}^* - X_t\beta - \mu
\end{aligned} \tag{C.25}$$

In the same way, the posterior distribution of the spatial dependence parameter ρ is given by:

$$\begin{aligned}
\rho \;\sim\; & N\left(\overline{\rho}, B_\rho\right) \\
& \text{with } B_\rho = \left[\tau^{-1}\sum_{t=2}^{T} y_t^{*'} y_t^*\right]^{-1} \text{ and } \overline{\rho} = B_\rho\left[\tau^{-1}\sum_{t=2}^{T} y_t^{*'} y_t^{(\rho)}\right] \\
& \text{where } y_t^{(\rho)} = y_t - \phi y_{t-1} - \delta y_{t-1}^* - X_t\beta - \mu
\end{aligned} \tag{C.26}$$

So too, the posterior distribution of the spatio-temporal diffusion parameter δ is given by:

$$\begin{aligned} \delta \quad &\sim \quad N\left(\overline{\delta}_y, B_\delta\right) \qquad \text{(C.27)} \\ &\text{with } B_\delta = \left[\tau^{-1}\sum_{t=2}^{T} y_{t-1}^{*'} y_{t-1}^{*}\right]^{-1} \text{ and } \overline{\delta}_y = B_\delta\left[\tau^{-1}\sum_{t=2}^{T} y_{t-1}^{*'} y_t^{(\delta)}\right] \\ &\text{where } y_t^{(\delta)} = y_t - \phi y_{t-1} - \rho y_t^{*} - X_t\beta - \mu \end{aligned}$$

D. The spatial weighting matrices

We use the census tract data set for Central New York State counties featured in Waller and Gotway (2004) and we work on two subsets of the map consisting of the $N = 63$ census tracts within Syracuse City and the $N = 120$ census tracts within Syracuse City and its neighborhood. We use several weighting matrices $W_N(= \{w_{ij}\})$ which essentially differ in their degree of sparseness. First, we create inverse distance weighting matrices with $w_{ij} = 1/dist(i,j)$ where $dist(i,j)$ is the distance (in km) between two census tracts i and j. The whole matrix W_N is filled with the diagonal elements being zero. Second, we create contiguity neighbors weighting matrices from the census tract rook-style and queen-style contiguities, by analogy with movements on a chessboard. Last, we create k-nearest neighbors weighting matrices with the $k = 4$ or 10 nearest neighbors (see Figures 1, 2, 3). Figure 1 shows the sparsity structure of the row-normalized inverse distance weight matrices for the $N = 63$ and $N = 120$ census tracts within Syracuse City and its neighborhourd. The non-sparsity rate is 98.4% (resp. 99.1%) for the $N = 63$ (resp. $N = 120$) census tracts since the weight matrix is completely filled except its first diagonal. Figure 2 shows the rook-style and queen-style for the $N = 63$ and $N = 120$ census tracts contiguities within Syracuse City and its neighborhourd. But this time, the non-sparsity rates are much lower: 8.7% and 7.7% (resp. 4.9% and 4.5%) for the rook-style and queen-style for the census tracts contiguities within Syracuse City (resp. within Syracuse City and its neighborhood). Figure 3 shows the k-nearest neighbors ($k = 4$, $k = 10$) for the $N = 63$ and $N = 120$ census tracts contiguities within Syracuse City and its neighborhourd. Again, the non-sparsity rates are small: 6.3% and 15.8% (resp. 3.3% and 8.3%) for the 4-nearest neighbors and the 10-nearest neighbors for the census tracts contiguities within Syracuse City (resp. within Syracuse City and its neighborhood). The minimum eigenvalues $\varpi_{\min}$ of the spatial weights matrices W_N change depending on the type and the sparsity of the spatial weights matrices (see Table D.1) while the maximum eigenvalues $\varpi_{\max}$ are always unity. These values allow to check the stationarity conditions given in eq.(4) in the main text. We have generated Monte Carlo DGP which always respect these stationarity conditions except for the explosive case in the random effects world.

Table D.1: Minimum eigenvalues of the spatial weights matrices W_N

	Inverse distance	rook	queen	4 neighbors	10 neighbors
$N = 63$	-0.0963	-0.5906	-0.6556	-0.6471	-0.2793
$N = 120$	-0.0779	-0.6090	-0.6123	-0.6414	-0.2810

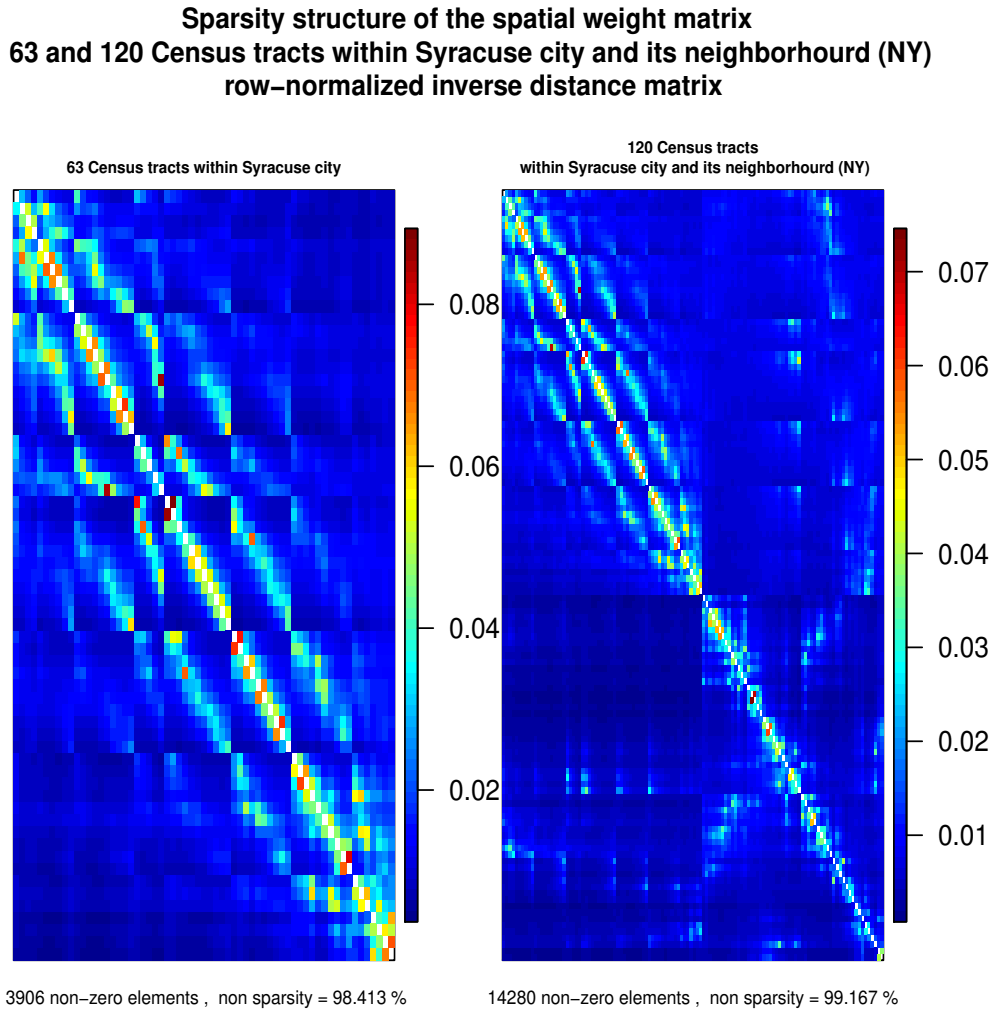


Figure 1: Sparsity structure of the spatial inverse distance weight matrix.

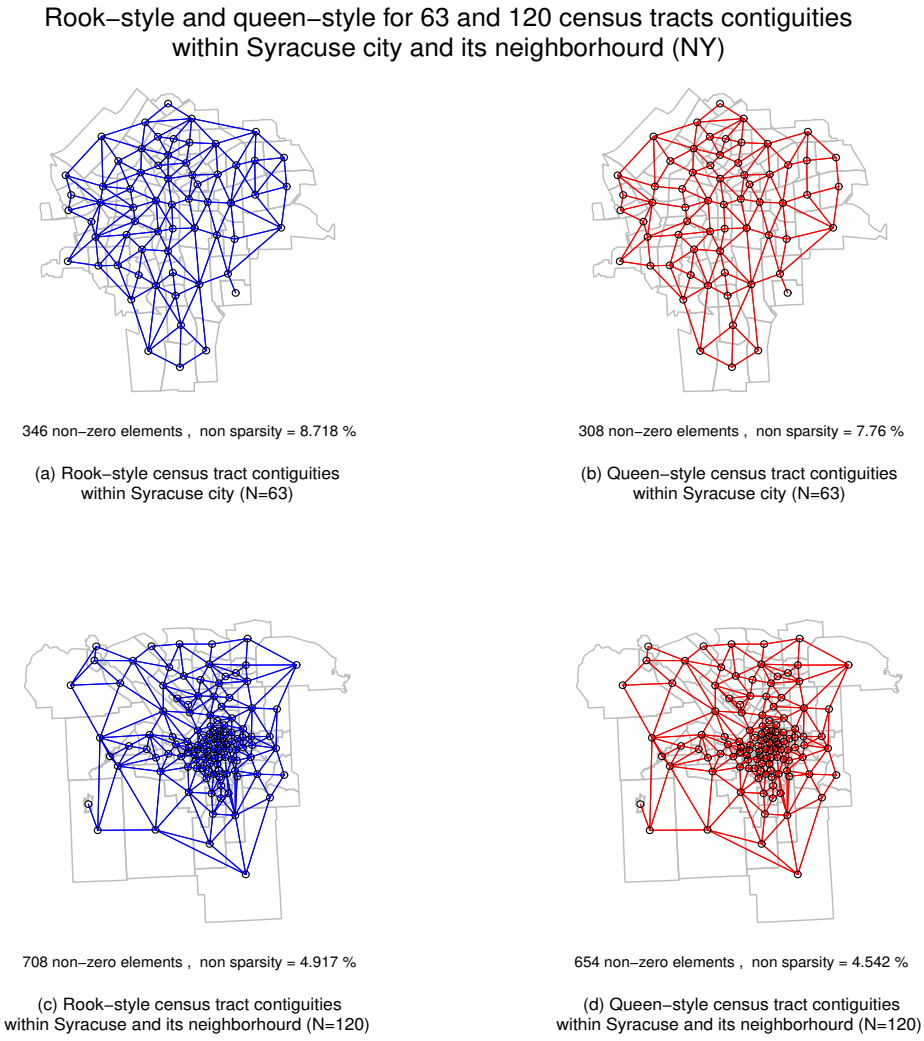
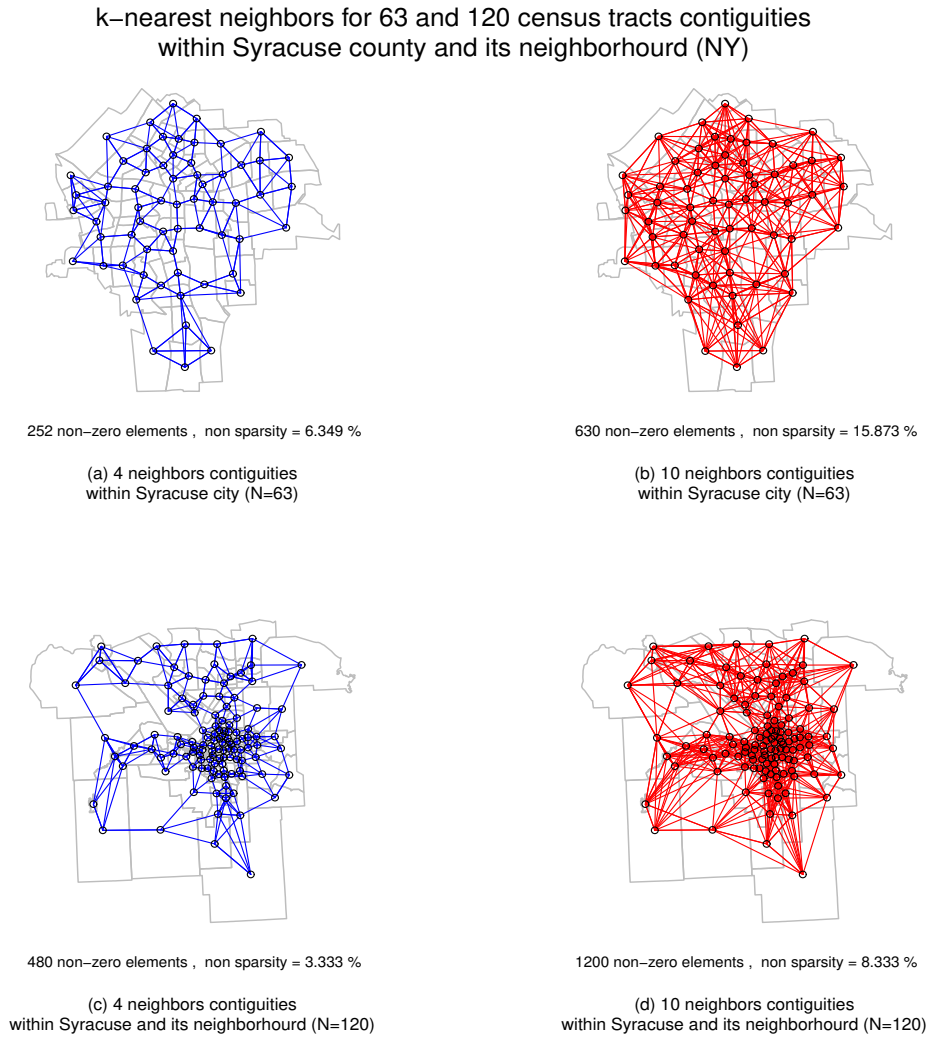


Figure 2: Rook-style and queen-style for the census tracts contiguities.



(a) 4 neighbors contiguities within Syracuse city (N=63)

(b) 10 neighbors contiguities within Syracuse city (N=63)

(c) 4 neighbors contiguities within Syracuse and its neighborhourd (N=120)

(d) 10 neighbors contiguities within Syracuse and its neighborhourd (N=120)

Figure 3: k-nearest neighbors for the census tracts contiguities.

E. A two-stage least squares (2SLS) estimator for the dynamic space-time homogeneous panel data world with correlated common factors.

We propose an extension of the two-stage least squares (2SLS) estimator of Yang (2021) to the case of a dynamic space-time homogeneous panel data world with correlated common factors. Our specification is given by:

$$\begin{aligned} y_{ti} &= \phi y_{t-1,i} + \rho y_{ti}^* + \delta y_{t-1,i}^* + x_{ti}\beta_1 + x_{t-1,i}\beta_2 + \gamma_i' f_t + u_{ti},\ t = 2, ..., T,\ i = 1, ..., N \\ &= \phi y_{t-1,i} + \rho y_{ti}^* + \delta y_{t-1,i}^* + X_{it}'\beta + \gamma_i' f_t + u_{ti} \end{aligned} \tag{E.28}$$

or for the pooled N individuals[18]

$$y_t = \phi y_{t-1} + \rho y_t^* + \delta y_{t-1}^* + X_t\beta + \Gamma f_t + u_t,\ t = 2, ..., T \tag{E.29}$$

with $f_t = (f_{t1}, f_{t2}, \cdots, f_{tm})'$, $\gamma_i = (\gamma_{i1}, \cdots, \gamma_{im})'$ and $\Gamma = (\gamma_1, \cdots, \gamma_N)'$ for m known common trends or m unobserved correlated common factors.
f_t (resp. γ_i, Γ) is of dimension $(m \times 1)$ (resp. $(m \times 1)$, $(N \times m)$). $X_{it}' = (x_{ti}, x_{t-1,i})$ is $(1 \times K_1)$, $X_t = (X_{1t}, \cdots, X_{Nt})'$ is $(N \times K_1)$ and $\beta = (\beta_1, \beta_2)'$ is $(K_1 \times 1)$ with $K_1 = 2$.
Let $Y = (y_2', \cdots, y_T')'$, $Y_{-1} = \left(y_1', \cdots, y_{T-1}'\right)'$, $Y^* = \left(y_2^{*'}, \cdots, y_T^{*'}\right)'$, $Y_{-1}^* = \left(y_1^{*'}, \cdots, y_{T-1}^{*'}\right)'$, $((T-1)N \times 1)$ vectors, $X = (X_2', \cdots, X_T')'$ a $((T-1)N \times K_1)$ matrix.
Let $L = \left(Y_{-1}, Y^*, Y_{-1}^*, X\right) = (L_2', \cdots, L_T')'$ denotes the full set of regressors where $L_t = (l_{t1}, \cdots, l_{tN})'$ with $l_{ti} = \left(y_{t-1,i}, y_{ti}^*, y_{t-1,i}^*, X_{ti}'\right)'$. l_{ti} is of dimension $((3 + K_1) \times 1)$, L_t is $(N \times (3 + K_1))$ and L is $((T-1)N \times (3 + K_1))$. Last, let $F = (f_2', \cdots, f_T')'$ the $((T-1) \times m)$ matrix of correlated common factors.

Following Yang (2021), the de-factoring matrices are defined as:

$$M_f = I_T - F\left(F'F\right)^- F' \text{ and } M_f^b = M_f \otimes I_N \tag{E.30}$$

where $(F'F)^-$ is the generalized inverse of $F'F$ and the 2SLS estimator of $\theta = (\phi, \rho, \delta, \beta_1, \beta_2)'$) is defined as

$$\hat{\theta}_{2SLS} = \left(L'P_Q L\right)' L'P_Q Y \tag{E.31}$$

where

$$P_Q = M_f^b Q \left(Q' M_f^b Q\right)^{-1} Q' M_f^b \tag{E.32}$$

with the IV matrix $Q = (Q_2', \cdots, Q_T')'$ with $Q_t = \left(X_t, W_N X_t, W_N^2 X_t, \cdots, W_N^q X_t\right)$. Q_t is of dimension $(N \times (q+1)K_1)$ and Q is $((T-1)N \times (q+1)K_1)$. In this 2SLS estimator, $M_f^b Q$ can be viewed as instruments.

Yang (2021) (p.15, eq(9)) shows that

$$\sqrt{N(T-1)}\left(\hat{\theta}_{2SLS} - \theta\right) \xrightarrow{d} \left[N\left(0, \Sigma_{\theta,2SLS}\right) + \text{ bias term}\right] \text{ as } (N,T) \to \infty \tag{E.33}$$

[18] We use the notation of Yang (2021) for $\gamma_i' f_t$ and Γf_t.

and, if $T/N \to 0$ when $(N,T) \to \infty$, the bias term vanishes and then the distribution of $\hat{\theta}_{2SLS}$ becomes $\sqrt{N(T-1)}\left(\hat{\theta}_{2SLS}-\theta\right) \sim N\left(0,\Sigma_{\theta,2SLS}\right)$. A consistent estimator for the asymptotic variance matrix $\Sigma_{\theta,2SLS}$ of θ is given by[19]

$$\hat{\Sigma}_{\theta,2SLS} = \hat{\Psi}^{-1}\hat{\Omega}\hat{\Psi}^{-1} \tag{E.34}$$

$$\text{with } \hat{\Psi} = \frac{1}{N(T-1)}L'P_QL \text{ , } \hat{\Omega} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{u}_{ti}^2\hat{l}_{ti}\hat{l}'_{ti} \tag{E.35}$$

where $\hat{L} = P_QL = \left(\hat{L}'_2,\cdots,\hat{L}'_T\right)'$, $\hat{l}_{ti}$ is the i-th row of $\hat{L}_t$, $\hat{u}_{ti}$ is the i-th row of $\hat{u}_t$ and $\hat{u} = M_f^b\left(Y - L\hat{\theta}_{2SLS}\right) = (\hat{u}'_2,\cdots,\hat{u}'_T)'$.

If f_t are m known common trends , we use the de-factoring matrices as in (E.30). On the contrary, for m unobservable common factors, we use observable counterparts of (E.30). We can approximate the $(m\times 1)$ f_t vector with a $((3+K_1)\times 1)$ f_t^* vector of the within time transformation[20] of the covariates:

$$f_t^* = \left(\left(\overline{y}_{-1,t}-\overline{\overline{y}}_{-1}\right),\left(\overline{y}_t^*-\overline{\overline{y}}^*\right),\left(\overline{y}_{-1,t}^*-\overline{\overline{y}}_{-1}^*\right),\left(\overline{x}_t-\overline{\overline{x}}\right),\left(\overline{x}_{-1,t}-\overline{\overline{x}}_{-1}\right)\right)' \tag{E.36}$$
$$\text{with } \overline{x}_t = (1/N)\sum_{i=1}^{N}x_{ti},\ \overline{\overline{x}} = (1/N(T-1))\sum_{i=1}^{N}\sum_{t=2}^{T}x_{ti}$$

or inspired by Chudik and Pesaran (2015a,b), we can approximate the $(m\times 1)$ f_t vector by the $((4+K_1)\times 1))$ f_t vector of the time means of the dependent and explanatory variables:

$$f_t^* = \left(\overline{y}_t,\overline{y}_{-1,t},\overline{y}_t^*,\overline{y}_{-1,t}^*,\overline{x}_t,\overline{x}_{-1,t}\right)' \tag{E.37}$$

Then, F in (E.30) is replaced by $F = (f_2^{*\prime},\cdots,f_T^{*\prime})'$ of dimension $((T-1)\times m)$ with $m = (3+K_1)$ or $m = (4+K_1)$. We will test the two approaches of the within time transformation of the explanatory variables and of the time means of the dependent and explanatory variables.

Contrarily to Yang (2021), we do not use only $q = 2$ in our Monte Carlo simulation study (*e.g.* $Q_t = \left(X_t, W_NX_t, W_N^2X_t\right)$) since it leads to biased estimates and large standard errors. We need to use $q = 7$ (*e.g.* $Q_t = \left(X_t, W_NX_t, W_N^2X_t,\cdots,W_N^7X_t\right)$) to get good results. The larger the dimension $((T-1)N\times(q+1)K_1)$ of the IV matrix Q, the better the estimates, especially in terms of standard errors.

F. A two-stage least squares (2SLS) estimator for the dynamic space-time heterogeneous panel data world with correlated common factors.

We propose an extension of the two-stage least squares (2SLS) estimator of Yang (2021) to the case of a dynamic space-time heterogeneous panel data world with correlated common factors. For

[19] We do not use the Newey-West type robust estimator (see Pesaran (2006), Yang (2021)) since our specification has i.i.d errors.

[20] *i.e.*, the demeaned time means.

an easier formalization, our specification is written in primal form:

$$\begin{aligned} y_{it} &= \phi_i y_{i,t-1} + \rho_i y^*_{i,t} + \delta_i y^*_{i,t-1} + x_{it}\beta_{1,i} + x_{i,t-1}\beta_{2,i} + f'_t\gamma_i + u_{it},\ t = 2, ..., T,\ i = 1, ..., N \\ &= \phi_i y_{i,t-1} + \rho_i y^*_{i,t} + \delta_i y^*_{i,t-1} + X'_{it}\beta_i + f'_t\gamma_i + u_{it} \end{aligned} \tag{F.38}$$

where $X'_{it} = (x_{it}, x_{i,t-1})$ and $\beta_i = (\beta_{1,i}, \beta_{2,i})'$. The pooled $(T-1)$ time periods specification is given by

$$Y_i = \phi_i Y_{i,-1} + \rho_i Y^*_i + \delta_i Y^*_{i,-1} + X_i\beta_i + F\gamma_i + u_i,\ i = 1, ..., N \tag{F.39}$$

where[21] $Y_i = (y_{i,2} \cdots, y_{i,T})'$, $Y_{i,-1} = (y_{i,1} \cdots, y_{i,1})'$, $Y^*_i = \left(y^*_{i,2} \cdots, y^*_{i,T}\right)'$, $Y^*_{i,-1} = \left(y^*_{i,1} \cdots, y^*_{i,T-1}\right)'$, $X_i = (X_{i,2} \cdots, X_{i,T})'$. As previously, the de-factoring matrix is defined as:

$$M_f = I_T - F\left(F'F\right)^- F' \tag{F.40}$$

By following Yang (2021) and being inspired by Pesaran (2006) and Chudik and Pesaran (2015a,b), the individual 2SLS estimator of $\theta_i = (\phi_i, \rho_i, \delta_i, \beta_{1,i}, \beta_{2,i})'$) is defined as

$$\hat{\theta}_{2SLS,i} = \left(L'_i P_{Q_i} L_i\right)' L'_i P_{Q_i} Y_i \tag{F.41}$$

where $L_i = \left(Y_{i,-1}, Y^*_i, Y^*_{i,-1}, X_i\right)$ denotes the full set of regressors for individual i and

$$P_{Q_i} = M_f Q^{(p)}_i \left(Q^{(p)'}_i M_f Q^{(p)}_i\right)^{-1} Q^{(p)'}_i M_f \tag{F.42}$$

where $Q^{(p)}_i$ is the $(T \times (q+1)K_1)$ submatrix of the primal form $Q^{(p)}$ of the IV matrix Q defined in section E. More precisely,

$$Q = \begin{pmatrix} q_{21,1} & \cdots & q_{21,r} \\ q_{22,1} & \cdots & q_{22,r} \\ \vdots & \vdots & \vdots \\ q_{2N,1} & \cdots & q_{2N,r} \\ q_{31,1} & \cdots & q_{31,r} \\ q_{32,1} & \cdots & q_{32,r} \\ \vdots & \vdots & \vdots \\ q_{3N,1} & \cdots & q_{3N,r} \\ \vdots & \vdots & \vdots \\ q_{T1,1} & \cdots & q_{T1,r} \\ q_{T2,1} & \cdots & q_{T2,r} \\ \vdots & \vdots & \vdots \\ q_{TN,1} & \cdots & q_{TN,r} \end{pmatrix},\ Q^{(p)} = \begin{pmatrix} q_{21,1} & \cdots & q_{21,r} \\ q_{31,1} & \cdots & q_{31,r} \\ \vdots & \vdots & \vdots \\ q_{T1,1} & \cdots & q_{T1,r} \\ q_{22,1} & \cdots & q_{22,r} \\ q_{32,1} & \cdots & q_{32,r} \\ \vdots & \vdots & \vdots \\ q_{T2,1} & \cdots & q_{T2,r} \\ \vdots & \vdots & \vdots \\ q_{2N,1} & \cdots & q_{2N,r} \\ q_{3N,1} & \cdots & q_{3N,r} \\ \vdots & \vdots & \vdots \\ q_{TN,1} & \cdots & q_{TN,r} \end{pmatrix} = \begin{pmatrix} Q^{(p)}_1 \\ Q^{(p)}_2 \\ \vdots \\ Q^{(p)}_N \end{pmatrix} \tag{F.43}$$

[21] f_t, γ_i and F are defined as in section E.

where $r = (q+1)K_1$. Inspired by Pesaran (2006) and Chudik and Pesaran (2015a,b), the 2SLS mean group estimator is a simple average of the individual 2SLS estimators $\hat{\theta}_{2SLS,i}$,

$$\hat{\theta}_{2SLS,MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{2SLS,i} \tag{F.44}$$

The distribution of $\hat{\theta}_{2SLS,MG}$ is

$$\sqrt{N}\left(\hat{\theta}_{2SLS,MG} - \theta\right) \quad \xrightarrow{d} \quad N\left(0, \Sigma_{\theta_{2SLS,MG}}\right) \text{ as } (N,T) \to \infty \tag{F.45}$$

and $\Sigma_{\theta_{2SLS,MG}}$ can be consistently estimated non-parametrically by

$$\Sigma_{\theta_{2SLS,MG}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_{2SLS,i} - \hat{\theta}_{2SLS,MG}\right)\left(\hat{\theta}_{2SLS,i} - \hat{\theta}_{2SLS,MG}\right)' \tag{F.46}$$

G. Some Monte Carlo simulation results.

G.1. Some results for the dynamic space-time random effects world

Table G.1: Dynamic Space-Time Random Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.4	-0.3	1	1	1	4			
$N = 63$	B2S2S coef	0.7510	0.4105	-0.3105	1.0032	1.0031	0.9963	3.8719	$< 10^{-4}$	0.4371	
$T = 10$	se_boot	0.0036	0.0424	0.0425	0.0147	0.0103	0.0579	0.7124			191.20
	se_mixt	0.0018	0.0385	0.0385	0.0081	0.0073	0.0579	0.7134			15.63
	rmse	0.0040	0.0460	0.0457	0.0156	0.0109	0.0580	0.7235			
	QMLE coef	0.7501	0.4099	-0.3098	0.9996	1.0000	0.9933	3.8965			430.181
	se	0.0038	0.0458	0.0458	0.0159	0.0111	0.0618	0.7445			
	rmse	0.0040	0.0489	0.0483	0.0155	0.0109	0.0621	0.7513			
	MCMC coef	0.7506	0.4096	-0.3100	0.9993	0.9994	1.0008	4.0054			3977.68
	se	0.0039	0.0500	0.0498	0.0150	0.0105	0.0578	0.7335			
	rmse	0.0039	0.0509	0.0508	0.0150	0.0105	0.0578	0.7331			
	nse	0.0001	0.0013	0.0012	0.0007	0.0005	0.0002	0.0077			
	cd	0.2630	0.1760	0.1590	0.5100	0.4880	0.4160	0.3370			
$N = 120$	B2S2S coef	0.7509	0.4098	-0.3095	1.0024	1.0021	0.9983	3.9507	$< 10^{-4}$	0.4687	
$T = 20$	se_boot	0.0019	0.0249	0.0250	0.0071	0.0049	0.0298	0.5158			790.44
	se_mixt	0.0009	0.0237	0.0237	0.0041	0.0038	0.0298	0.5156			60.32
	rmse	0.0022	0.0286	0.0281	0.0077	0.0056	0.0298	0.5179			
	QMLE coef	0.7500	0.4104	-0.3097	1.0001	0.9998	0.9968	3.9935			1196.266
	se	0.0018	0.0258	0.0259	0.0074	0.0051	0.0307	0.5312			
	rmse	0.0018	0.0286	0.0282	0.0073	0.0051	0.0309	0.5310			

B2S2S : Bayesian two-step two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}\,(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

MCMC: MCMC Gibbs sampling with $1,000$ draws and 500 burnin draws.

Stationarity conditions for B2S2S for $N = 63$, $T = 10$ and for $N = 120$, $T = 20$: $\phi + (\rho + \delta)\,\varpi_{\max} = 0.85(< 1)$ as $\rho + \delta = 0.1(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = 0.05(> -1)$ as $\rho - \delta = 0.7(\geq 0)$.

Due to the excessive computing time, we did not run the MCMC Gibbs sampling for $N = 120$ and $T = 20$.

Table G.2: Dynamic Space-Time Random Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.3	0.8	-0.24	1	1	1	4			
$N = 63$	B2S2S coef	0.2981	0.8094	-0.2446	1.0035	1.0035	0.9935	3.9589	$< 10^{-4}$	0.4332	
$T = 10$	se_boot	0.0076	0.0171	0.0187	0.0152	0.0109	0.0578	0.7281			204.19
	se_mixt	0.0055	0.0152	0.0162	0.0087	0.0079	0.0577	0.7302			17.34
	rmse	0.0083	0.0201	0.0198	0.0160	0.0116	0.0581	0.7289			
	QMLE coef	0.3004	0.8098	-0.2475	0.9992	0.9996	0.9917	3.9303			385.789
	se	0.0082	0.0181	0.0198	0.0164	0.0118	0.06181	0.7460			
	rmse	0.0086	0.0214	0.0219	0.0161	0.0114	0.0623	0.7488			
	MCMC coef	0.3000	0.8094	-0.2468	0.9997	0.9998	0.9998	4.0150			4122.09
	se	0.0081	0.0181	0.0202	0.0156	0.0112	0.0578	0.7372			
	rmse	0.0081	0.0204	0.0213	0.0156	0.0112	0.0577	0.7370			
	nse	0.0003	0.0007	0.0007	0.0007	0.0005	0.0002	0.0062			
	cd	0.3360	0.4870	0.4140	0.4980	0.4450	0.5380	0.4080			
$N = 120$	B2S2S coef	0.2992	0.8081	-0.2454	1.0024	1.0021	0.9970	4.0028	$< 10^{-4}$	0.4673	
$T = 20$	se_boot	0.0037	0.0100	0.0107	0.0074	0.0053	0.0298	0.5243			755.21
	se_mixt	0.0028	0.0093	0.0097	0.0044	0.0041	0.0298	0.5243			58.46
	rmse	0.0041	0.0133	0.0123	0.0080	0.0060	0.0299	0.5240			
	QMLE coef	0.3001	0.8085	-0.2465	0.9998	0.9995	0.9965	4.0008			1116.55
	se	0.0039	0.0103	0.0111	0.0077	0.0056	0.0307	0.5298			
	rmse	0.0040	0.0135	0.0128	0.0078	0.0056	0.0309	0.5295			

B2S2S : Bayesian two-step two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

MCMC: MCMC Gibbs sampling with $1,000$ draws and 500 burnin draws.

Stationarity conditions for B2S2S for $N = 63$, $T = 10$ and for $N = 120$, $T = 20$: $\phi + (\rho + \delta)\,\varpi_{\max} = 0.86(< 1)$ as $\rho + \delta = 0.56(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.74(> -1)$ as $\rho - \delta = 1.04(\geq 0)$.

Due to the excessive computing time, we did not run the MCMC Gibbs sampling for $N = 120$ and $T = 20$.

Table G.3: Dynamic Space-Time Random Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.3	0.4	-0.12	1	1	1	4			
$N = 63$	B2S2S coef	0.2981	0.4123	-0.1253	1.0038	1.0038	0.9939	3.9580	$< 10^{-4}$	0.4327	
$T = 10$	se_boot	0.0075	0.0450	0.0453	0.0152	0.0109	0.0578	0.7278			221.89
	se_mixt	0.0055	0.0404	0.0406	0.0086	0.0079	0.0577	0.7301			18.35
	rmse	0.0082	0.0491	0.0475	0.0161	0.0117	0.0581	0.7287			
	QMLE coef	0.3003	0.4104	-0.1280	0.9996	1.0000	0.9922	3.9310			374.63
	se	0.0081	0.0481	0.0484	0.0164	0.0118	0.0617	0.7475			
	rmse	0.0085	0.0516	0.0505	0.0160	0.0114	0.0622	0.7503			
	MCMC coef	0.2998	0.4105	-0.1279	1.0000	1.0001	1.0002	4.0168			4105.448
	se	0.0080	0.0485	0.0489	0.0155	0.0112	0.0578	0.7378			
	rmse	0.0080	0.0496	0.0495	0.0155	0.0111	0.0577	0.7376			
	nse	0.0003	0.0018	0.0017	0.0007	0.0005	0.0002	0.0062			
	cd	0.3860	0.4490	0.4100	0.4980	0.4690	0.5450	0.4310			
$N = 120$	B2S2S coef	0.2992	0.4125	-0.1285	1.0025	1.0023	0.9972	4.0039	$< 10^{-4}$	0.4669	
$T = 20$	se_boot	0.0037	0.0275	0.0276	0.0074	0.0053	0.0298	0.5241			771.34
	se_mixt	0.0028	0.0256	0.0256	0.0045	0.0040	0.0298	0.5241			56.46
	rmse	0.0041	0.0321	0.0307	0.0081	0.0061	0.0299	0.5238			
	QMLE coef	0.3000	0.4125	-0.1298	1.0001	0.9997	0.9967	4.0025			1114.728
	se	0.0038	0.0286	0.0287	0.0077	0.0056	0.0307	0.5300			
	rmse	0.0040	0.0318	0.0308	0.0078	0.0056	0.0308	0.5297			

B2S2S : Bayesian two-step two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}\,(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

MCMC: MCMC Gibbs sampling with $1,000$ draws and 500 burnin draws.

Stationarity conditions for B2S2S for $N = 63$, $T = 10$ and for $N = 120$, $T = 20$: $\phi + (\rho + \delta)\,\varpi_{\max} = 0.58(< 1)$ as $\rho + \delta = 0.28(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.22(> -1)$ as $\rho - \delta = 0.52(\geq 0)$.

Due to the excessive computing time, we did not run the MCMC Gibbs sampling for $N = 120$ and $T = 20$.

Table G.4: Dynamic Space-Time Random Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=1,000

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.98	0.8	-0.784	1	1	1	4			
$N = 63$	B2S2S coef	0.9805	0.8087	-0.7931	1.0004	1.0002	0.9964	3.7805	$< 10^{-4}$	0.4444	
$T = 10$	se_boot	0.0006	0.0159	0.0161	0.0145	0.0101	0.0580	0.7034			197.56
	se_mixt	0.0002	0.0147	0.0149	0.0074	0.0064	0.0580	0.7025			15.37
	rmse	0.0008	0.0190	0.0193	0.0150	0.0104	0.0581	0.7365			
	95%hpdi lower	0.9789	0.7610	-0.8392	0.9609	0.9720	0.8929	2.5645			
	95%hpdi upper	0.9822	0.8538	-0.7451	1.0419	1.0277	1.1122	5.2077			
	QMLE coef	0.9800	0.8088	-0.7929	0.9991	0.9994	12.0887	2.6163			494.96
	se	0.0006	0.0177	0.0179	0.0161	0.0112	325.1427	36.1711			
	rmse	0.0007	0.0204	0.0207	0.0167	0.0133	325.1692	36.1795			
	MCMC coef	0.9805	0.8087	-0.7932	0.9971	0.9971	1.0004	4.2169			3997.09
	se	0.0007	0.0192	0.0196	0.0149	0.0105	0.0581	0.7333			
	rmse	0.0009	0.0211	0.0216	0.0151	0.0109	0.0580	0.7694			
	95%hpdi lower	0.9792	0.7738	-0.8367	0.9706	0.9767	0.8943	2.6691			
	95%hpdi upper	0.9819	0.8491	-0.7592	1.0299	1.0167	1.1146	5.4582			
	nse	0.0000	0.0001	0.0001	0.0007	0.0005	0.0002	0.0105			
	cd	0.1470	0.0690	0.0810	0.5040	0.4780	0.4910	0.3030			
	true	0.98	0.4	-0.392	1	1	1	4			
$N = 63$	B2S2S coef	0.9805	0.4108	-0.4030	1.0007	1.0005	0.9968	3.7801	$< 10^{-4}$	0.4417	
$T = 10$	se_boot	0.0006	0.0415	0.0419	0.0145	0.0101	0.0579	0.7031			209.28
	se_mixt	0.0002	0.0362	0.0367	0.0073	0.0064	0.0579	0.7021			16.16
	rmse	0.0008	0.0458	0.0463	0.0150	0.0104	0.0580	0.7363			
	95%hpdi lower	0.9789	0.2911	-0.5204	0.9608	0.9724	0.8930	2.5664			
	95%hpdi upper	0.9822	0.5316	-0.2780	1.0417	1.0282	1.1119	5.2233			
	QMLE coef	0.9800	0.4100	-0.4023	0.9991	0.9994	13.3222	2.4739			500.87
	se	0.0006	0.0466	0.0470	0.0161	0.0113	251.2457	27.9639			
	rmse	0.0007	0.0491	0.0497	0.0182	0.0144	251.4222	27.9915			
	MCMC coef	0.9806	0.4104	-0.4028	0.9973	0.9974	1.0005	3.9417			4096.12
	se	0.0007	0.0488	0.0496	0.0148	0.0104	0.0580	0.7179			
	rmse	0.0009	0.0499	0.0507	0.0150	0.0107	0.0580	0.7199			
	95%hpdi lower	0.9793	0.3212	-0.5009	0.9709	0.9776	0.8931	2.6781			
	95%hpdi upper	0.9819	0.5125	-0.3081	1.0299	1.0173	1.1140	5.4108			
	nse	0.0000	0.0003	0.0003	0.0007	0.0005	0.0002	0.0083			
	cd	0.1750	0.0640	0.0670	0.5130	0.4980	0.4960	0.3250			

B2S2S : Bayesian two-step two-step estimation.
se_boot: standard errors computed with individual block resampling bootstrap.
se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
QMLE: quasi-maximum likelihood estimation.
MCMC: MCMC Gibbs sampling with 1,000 draws and 500 burnin draws.
95%hpdi lower, upper: lower and upper bounds of the 95% HPDI.
Stationarity conditions for B2S2S for $\phi = 0.98$, $\rho = 0.8$ and $\delta = -0.784$: $\phi + (\rho + \delta)\,\varpi_{\max} = 0.996(< 1)$ as $\rho + \delta = 0.016(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.604(> -1)$ as $\rho - \delta = 1.584(\geq 0)$.
Stationarity conditions for B2S2S for $\phi = 0.98$, $\rho = 0.4$ and $\delta = -0.392$: $\phi + (\rho + \delta)\,\varpi_{\max} = 0.988(< 1)$ as $\rho + \delta = 0.008(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = 0.188(> -1)$ as $\rho - \delta = 0.792(\geq 0)$.

Table G.5: Dynamic Space-Time Random Effects World

$\varepsilon = 0.5$, $r = 0.8$, Replications=1,000, $N = 63$, $T = 10$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	4			
Rook-style	B2S2S coef	0.7516	0.8115	-0.6120	0.9998	0.9998	0.9931	3.8372	$< 10^{-4}$	0.4341	
	se_boot	0.0037	0.0079	0.0090	0.0149	0.0106	0.0579	0.7101			202.76
	se_mixt	0.0019	0.0069	0.0075	0.0084	0.0076	0.0579	0.7115			15.91
	rmse	0.0043	0.0140	0.0150	0.0154	0.0107	0.0583	0.7282			
	QMLE coef	0.7507	0.8124	-0.6118	0.9959	0.9964	0.9901	3.8643			414.21
	se	0.0039	0.0085	0.0096	0.0161	0.0114	0.0618	0.7450			
	rmse	0.0042	0.0149	0.0150	0.0162	0.0117	0.0626	0.7569			
	MCMC coef	0.7512	0.8120	-0.6121	0.9957	0.9959	0.9974	3.9724			4017.18
	se	0.0043	0.0092	0.0107	0.0153	0.0110	0.0578	0.7326			
	rmse	0.0045	0.0152	0.0161	0.0159	0.0118	0.0579	0.7328			
	nse	0.0001	0.0002	0.0002	0.0007	0.0005	0.0002	0.0073			
	cd	0.1740	0.1350	0.0980	0.4860	0.4410	0.3610	0.2760			
Queen-style	B2S2S coef	0.7516	0.8118	-0.6123	0.9994	0.9993	0.9926	3.8337	$< 10^{-4}$	0.4340	
	se_boot	0.0038	0.0076	0.0087	0.0149	0.0106	0.0578	0.7093			205.86
	se_mixt	0.0019	0.0067	0.0073	0.0084	0.0077	0.0578	0.7106			16.15
	rmse	0.0044	0.0141	0.0152	0.0155	0.0108	0.0583	0.7282			
	QMLE coef	0.7507	0.8127	-0.6122	0.9954	0.9959	0.9896	3.8610			407.09
	se	0.0039	0.0082	0.0094	0.0161	0.0114	0.0618	0.7443			
	rmse	0.0042	0.0151	0.0152	0.0164	0.0119	0.0626	0.7568			
	MCMC coef	0.7513	0.8124	-0.6125	0.9953	0.9954	0.9969	3.9685			3975.79
	se	0.0043	0.0091	0.0105	0.0153	0.0111	0.0578	0.7322			
	rmse	0.0045	0.0153	0.0163	0.0160	0.0120	0.0578	0.7325			
	nse	0.0001	0.0002	0.0002	0.0007	0.0005	0.0002	0.0073			
	cd	0.1590	0.1250	0.1090	0.4900	0.4400	0.3620	0.2880			

B2S2S : Bayesian two-step two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

MCMC: MCMC Gibbs sampling with 1,000 draws and 500 burn-in draws.

Rook-style, Queen-style: rook-style and queen-style weighting matrices for the 63 census tracts contiguities within Syracuse City.

Stationarity conditions for B2S2S with rook-style or queen-style weighting matrix : $\phi + (\rho + \delta)\,\varpi_{\max} = 0.95(< 1)$ as $\rho + \delta = 0.2(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.65(> -1)$ as $\rho - \delta = 1.4(\geq 0)$.

Table G.6: Dynamic Space-Time Random Effects World
$\varepsilon = 0.5$, $r = 0.8$, Replications=1,000, $N = 63$, $T = 10$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	4			
4 neighbors	B2S2S coef	0.7516	0.8104	-0.6110	0.9997	0.9996	0.9929	3.8334	$< 10^{-4}$	0.4339	
	se_boot	0.0037	0.0069	0.0080	0.0149	0.0106	0.0579	0.7066			222.32
	se_mixt	0.0019	0.0060	0.0066	0.0084	0.0077	0.0578	0.7076			16.65
	rmse	0.0043	0.0126	0.0138	0.0155	0.0108	0.0583	0.7256			
	QMLE coef	0.7508	0.8115	-0.6111	0.9955	0.9960	0.9945	3.8600			401.59
	se	0.0039	0.0074	0.0087	0.0161	0.0114	0.1577	0.7467			
	rmse	0.0046	0.0171	0.0186	0.0181	0.0135	0.1578	0.7594			
	MCMC coef	0.7513	0.8108	-0.6110	0.9956	0.9957	0.9972	3.9694			4030.89
	se	0.0043	0.0083	0.0099	0.0154	0.0111	0.0578	0.7315			
	rmse	0.0045	0.0137	0.0148	0.0160	0.0119	0.0579	0.7318			
	nse	0.0001	0.0002	0.0002	0.0007	0.0005	0.0002	0.0072			
	cd	0.1600	0.1160	0.0870	0.4850	0.4320	0.3710	0.2650			
10 neighbors	B2S2S coef	0.7512	0.8088	-0.6093	1.0020	1.0019	0.9952	3.8581	$< 10^{-4}$	0.4354	
	se_boot	0.0037	0.0099	0.0106	0.0148	0.0104	0.0579	0.7139			217.03
	se_mixt	0.0019	0.0088	0.0092	0.0082	0.0074	0.0579	0.7147			18.84
	rmse	0.0041	0.0135	0.0143	0.0154	0.0108	0.0581	0.7275			
	QMLE coef	0.7503	0.8095	-0.6090	0.9983	0.9987	0.9922	3.8842			409.84
	se	0.0038	0.0107	0.0115	0.0160	0.0112	0.0618	0.7459			
	rmse	0.0040	0.0143	0.0145	0.0157	0.0111	0.0623	0.7544			
	MCMC coef	0.7508	0.8092	-0.6092	0.9980	0.9981	0.9996	3.9926			4009.61
	se	0.0040	0.0115	0.0125	0.0151	0.0107	0.0579	0.7363			
	rmse	0.0041	0.0147	0.0155	0.0152	0.0109	0.0579	0.7359			
	nse	0.0001	0.0003	0.0003	0.0007	0.0005	0.0002	0.0075			
	cd	0.2050	0.1410	0.1250	0.4970	0.4750	0.3770	0.2760			

B2S2S : Bayesian two-step two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}\,(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

MCMC: MCMC Gibbs sampling with 1,000 draws and 500 burnin draws.

4, 10 neighbors: 4-nearest and 10-nearest neighbors weighting matrices for the 63 census tracts contiguities within Syracuse City.

Stationarity conditions for B2S2S with 4-nearest or 10-nearest neighbors weighting matrix : $\phi + (\rho + \delta)\,\varpi_{\max} = 0.95(< 1)$ as $\rho + \delta = 0.2(\geq 0)$ and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.65(> -1)$ as $\rho - \delta = 1.4(\geq 0)$.

Table G.7: Dynamic Space-Time Random Effects World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	1.05	0.8	-0.84	1	1	1	4			
$N = 63$	B2S2S coef	1.0500	0.8083	-0.8488	1.0018	1.0016	0.9962	3.8125	$< 10^{-4}$	0.4421	
$T = 10$	se_boot	0.0001	0.0160	0.0168	0.0144	0.0098	0.0582	0.7039			219.85
	se_mixt	0.0000	0.0137	0.0145	0.0068	0.0057	0.0582	0.7031			17.08
	rmse	0.0001	0.0187	0.0196	0.0149	0.0102	0.0583	0.728			
	95%hpdi lower	1.0499	0.7612	-0.8966	0.9616	0.9744	0.8930	2.5214			
	95%hpdi upper	1.0502	0.8539	-0.7992	1.0414	1.0289	1.1117	5.2085			
	QMLE coef	1.0500	0.8087	-0.8492	0.9889	0.9892	62063.263	47004.506			643.16
	se	0.0001	0.0537	0.0565	0.0267	0.0221	157282.95	142750.25			
	rmse	0.0003	0.0757	0.0799	0.0735	0.0704	168989.72	150209.81			
	MCMC coef	1.0500	0.8100	-0.8505	0.9981	0.9981	0.9996	3.9677			4132.92
	se	0.0001	0.0190	0.0200	0.0147	0.0101	0.0582	0.7267			
	rmse	0.0001	0.0214	0.0225	0.0148	0.0103	0.0582	0.7270			
	95%hpdi lower	1.0499	0.7730	-0.8896	0.9712	0.9791	0.8930	2.7001			
	95%hpdi upper	1.0502	0.8453	-0.8130	1.0287	1.0184	1.1124	5.4652			
	nse	0.0000	0.0000	0.0000	0.0007	0.0004	0.0002	0.0076			
	cd	0.2470	0.0730	0.0620	0.5150	0.5000	0.4610	0.3430			

B2S2S : Bayesian two-step two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}\,(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

MCMC: MCMC Gibbs sampling with $1,000$ draws and 500 burnin draws.

95%hpdi lower, upper: lower and upper bounds of the 95% HPDI.

Stationarity conditions for B2S2S : $\phi + (\rho + \delta)\,\varpi_{\min} = 1.05385(<?\,1)$ as $\rho + \delta = -0.04(\geq ?\,0)$
and $\phi - (\rho - \delta)\,\varpi_{\max} = -0.59(> -1)$ as $\rho - \delta = 1.64(\geq 0)$.

Table G.8: Dynamic Space-Time Chamberlain-type Fixed Effects World with row normalized inverse distance weighting matrix
$N = 63$, $T = 10$, $\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	224.3295			
$N = 63$	B2S2S coef	0.7485	0.8072	-0.6055	1.0030	0.9992	0.9913	224.3587	$< 10^{-4}$	0.4842	
$T = 10$	se_boot	0.0038	0.0140	0.0145	0.0104	0.0144	0.0580	42.9346			339.59
	se_mixt	0.0046	0.0272	0.0278	0.0178	0.0280	0.0580	42.7039			22.57
	rmse	0.0041	0.0160	0.0157	0.0110	0.0158	0.0587	42.9133			
	QMLE coef	0.7503	0.8096	-0.6094	0.9986	0.9998	1.0017	220.4818			981.16
	se	0.0037	0.0149	0.0154	0.0113	0.0158	0.2568	42.3261			
	rmse	0.0046	0.0365	0.0385	0.014	0.0164	0.2566	42.4706			

B2S2S : Bayesian two-stage two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}\,(b|g_0)$.

QMLE: quasi-maximum likelihood estimation.

	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
true	0.1678	0.2097	0.2621	0.3277	0.4096	0.5120	0.64	0.8	1
B2S2S coef	0.2182	0.2604	0.3122	0.3813	0.4613	0.5641	0.6925	0.8549	1.0567
se_boot	0.0652	0.0646	0.0639	0.0648	0.0655	0.0647	0.0648	0.0655	0.0663
se_mixt	0.0296	0.0297	0.0297	0.0301	0.0306	0.0310	0.0321	0.0334	0.0351
rmse	0.0806	0.0813	0.0807	0.0831	0.0823	0.0815	0.0810	0.0844	0.085
QMLE coef	0.2163	0.2578	0.3088	0.3777	0.4566	0.5600	0.6867	0.8493	1.0491
se	0.0537	0.0535	0.0533	0.0538	0.0538	0.0539	0.0543	0.0548	0.0555
rmse	0.0772	0.0779	0.0776	0.0794	0.0770	0.0777	0.0765	0.0796	0.0796

Table G.9: Dynamic Space-Time Chamberlain-type Fixed Effects World with row normalized inverse distance weighting matrix
$N = 120$, $T = 20$, $\varepsilon = 0.5$, $r = 0.8$, Replications=1, 000

		ϕ	ρ	δ	β_1	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	222.9121			
$N = 120$	B2S2S coef	0.7491	0.8052	-0.6044	1.0017	0.9997	0.9986	223.0591	$< 10^{-4}$	0.4950	
$T = 20$	se_boot	0.0018	0.0083	0.0085	0.0049	0.0072	0.0280	28.6932			2068.82
	se_mixt	0.0024	0.0166	0.0168	0.0092	0.0146	0.0280	28.6973			100.66
	rmse	0.0021	0.0099	0.0097	0.0051	0.0073	0.0280	28.6793			
	QMLE coef	0.7501	0.8055	-0.6055	0.9997	0.9999	0.9991	220.8988			4221.21
	se	0.0018	0.0087	0.0089	0.0052	0.0076	0.0292	28.4953			
	rmse	0.0019	0.0102	0.0104	0.0049	0.0074	0.0292	28.5521			

B2S2S : Bayesian two-stage two-step estimation.
se_boot: standard errors computed with individual block resampling bootstrap.
se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
QMLE: quasi-maximum likelihood estimation.

	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
true	0.0180	0.0225	0.0281	0.0352	0.0440	0.0550	0.0687	0.0859	0.1074
B2S2S coef	0.0214	0.0265	0.0290	0.0392	0.0485	0.0587	0.0691	0.0893	0.1130
se_boot	0.0383	0.0382	0.0380	0.0382	0.0384	0.0384	0.0383	0.0382	0.0384
se_mixt	0.0153	0.0153	0.0153	0.0153	0.0153	0.0154	0.0153	0.0153	0.0154
rmse	0.0375	0.0372	0.0368	0.0374	0.0394	0.0371	0.0378	0.0370	0.0366
QMLE coef	0.0209	0.0260	0.0289	0.0386	0.0476	0.0585	0.0682	0.0888	0.1122
se	0.0326	0.0326	0.0326	0.0326	0.0327	0.0326	0.0326	0.0326	0.0326
rmse	0.0362	0.0358	0.0357	0.0359	0.0381	0.0356	0.0366	0.0355	0.0357

	π_{11}	π_{12}	π_{33}	π_{14}	π_{15}	π_{16}	π_{17}	π_{18}	π_{19}	π_{20}
true	0.1342	0.1678	0.2097	0.2621	0.3277	0.4096	0.5120	0.64	0.8	1
B2S2S coef	0.1371	0.1700	0.2144	0.2676	0.3298	0.4147	0.5166	0.6443	0.8071	1.0045
se_boot	0.0384	0.0383	0.0382	0.0378	0.0383	0.0385	0.0381	0.0385	0.0388	0.0388
se_mixt	0.0154	0.0154	0.0155	0.0155	0.0157	0.0159	0.0162	0.0166	0.0172	0.0181
rmse	0.0372	0.0374	0.0371	0.0378	0.0383	0.0360	0.0362	0.0388	0.0388	0.0384
QMLE coef	0.1363	0.1691	0.2137	0.2665	0.3282	0.4130	0.5146	0.6419	0.8033	1.0007
se	0.0327	0.0326	0.0327	0.0325	0.0327	0.0326	0.0329	0.0328	0.0332	0.0334
rmse	0.0357	0.0363	0.0357	0.0367	0.0373	0.0355	0.0352	0.0375	0.0372	0.0373

G.3. The dynamic space-time Hausman-Taylor world: results of the Monte Carlo simulation study

The static Hausman-Taylor model (henceforth HT, see Hausman and Taylor (1981)) posits that $y = X\beta + V\eta + Z_\mu\mu + u$, where V is a vector of time-invariant variables, and that subsets of X (*e.g.,* $X'_{2,i}$) and V (*e.g.*, V'_{2i}) may be correlated with the individual effects μ, but leave the correlations unspecified. Hausman and Taylor (1981) proposed a two-step IV estimator.

For our dynamic space-time model: $y = Z\theta + Vb + u = \phi y_{-1} + \rho y^* + \delta y^*_{-1} + X\beta + V\eta + Z_\mu\mu + u$, we assume that ($\overline{X'_{2,i}}$, V'_{2i} and μ_i) are jointly normally distributed:

$$\begin{pmatrix} \mu_i \\ \begin{pmatrix} \overline{X'_{2,i}} \\ V'_{2i} \end{pmatrix} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \begin{pmatrix} E_{\overline{X'_2}} \\ E_{V'_2} \end{pmatrix} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

where $\overline{X'_{2,i}}$ (resp. $E_{\overline{X'_2}}$) is the individual mean (resp. general mean) of $X'_{2,ti}$ (resp. of X'_2). $E_{V'_2}$ is the mean of V'_2. The conditional distribution of $\mu_i \mid \overline{X'_{2,i}}, V'_{2i}$ is given by:

$$\mu_i \mid \overline{X'_{2,i}}, V'_{2i} \sim N\left(\Sigma_{12}\Sigma_{22}^{-1}. \begin{pmatrix} \overline{X'_{2,i}} - E_{\overline{X'_2}} \\ V'_{2i} - E_{V'_2} \end{pmatrix}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

Since we do not know the elements of the variance-covariance matrix Σ_{jk}, we can write:

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right)\theta_X + \left(V'_{2i} - E_{V'_2}\right)\theta_V + \omega_i,$$

where $\omega_i \sim N\left(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$ is uncorrelated with u_{it}, and where θ_X and θ_V are vectors of parameters to be estimated. In order to identify the coefficient vector of V'_{2i} and to avoid possible collinearity problems, we assume that the individual effects are given by:

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right)\theta_X + f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(V'_{2i} - E_{V'_2}\right)\right]\theta_V + \omega_i, \tag{G.47}$$

where $\odot$ is the Hadamard product and $f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(V'_{2i} - E_{V'_2}\right)\right]$ can be a nonlinear function of $\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(V'_{2i} - E_{V'_2}\right)$. The first term on the right-hand side of equation (G.47) corresponds to the Mundlak (1978) transformation while the middle term captures the correlation between V'_{2i} and μ_i. The individual effects, μ, are a function of $PX^{(p)}$ and $\left(f\left[PX^{(p)} \odot V\right]\right)$, *i.e.,* a function of the column-by-column Hadamard product of $PX^{(p)}$ and V where $P = (I_N \otimes J_N/N)$ is the between transformation, $X^{(p)}$ is the primal form of the X matrix (*i.e.*, with now i (resp. t) being the slower (resp. faster) index)[22] and J_N is a $(N \times N)$ matrix of ones.

We can once again concatenate $\left[y_{-1}, y^*, y^*_{-1}, X, \left\{PX^{(p)}\right\}^{(d)}, f\left[\left\{PX^{(p)}\right\}^{(d)} \odot V\right]\right]$ into a single matrix of observables $\tilde{Z}$ where $\left\{PX^{(p)}\right\}^{(d)}$ is the dual form of $PX^{(p)}$ and then the model becomes: $y = \tilde{Z}\theta + D\varpi + u$ with $D = \iota_T \otimes I_N$. As we have assumed that

$$\mu_i = (\overline{x_{2,i}} - E_{\overline{x_2}})\,\theta_X + f\left[(\overline{x_{2,i}} - E_{\overline{x_2}}) \odot (V_{2i} - E_{V_2})\right]\theta_V + \omega_i. \tag{G.48}$$

[22] From the dual form of X, each column k of the $(N(T-1) \times K_x)$ matrix X is rewritten as a $(N \times (T-1))$ matrix A_{X_k} and the primal form of each column k of X is given as: $vec(A'_{X_k})$. We can also use commutation matrices.

We propose adopting the following strategy: If the correlation between μ_i and V_{2i} is quite large (> 0.2), use $f[.] = (\overline{x_{2,i}} - E_{\overline{x_2}})^2 \odot (V_{2i} - E_{V_2})^s$ with $s = 1$. If the correlation is weak, set $s = 2$. In real-world applications, we do not know the correlation between μ_i and V_{2i} *a priori*. We can use a proxy of μ_i defined by the OLS estimation of μ: $\widehat{\mu} = (D'D)^{-1} D'\widehat{y}$ where $\widehat{y}$ are the fitted values of the pooling regression $y = \phi y_{-1} + \rho y^* + \delta y^*_{-1} + x_1\beta_1 + x_2\beta_2 + V_1\eta_1 + V_2\eta_2 + \zeta$. Then, we compute the correlation between $\widehat{\mu}$ and V_2. In our simulation study, it turns out the correlations between μ and V_2 are large: 0.53 (resp. 0.67) when $\phi = 0.75$ and $\rho = 0.8$ for $N = 63$ (resp. $N = 120$). Hence, we choose $s = 1$.

Our B2S2S estimation method is compared with the two-stage quasi-maximum likelihood sequential approach proposed by Kripfganz and Schwarz (2019) and adapted here to the dynamic space-time framework. In the first stage, the coefficients of the time-varying regressors are estimated without relying on coefficient estimates for the time-invariant regressors using the quasi-maximum likelihood (QML) estimator of Hsiao et al. (2002) with the "xtdpdqml" Stata command. Subsequently, the first-stage residuals are regressed on the time-invariant regressors. Identification is achieved by using instrumental variables in the spirit of Hausman and Taylor (1981), and the second-stage standard errors are adjusted to account for the first-stage estimation error.[23] Kripfganz and Schwarz (2019) have proposed a new "xtseqreg" Stata command which implements the standard error correction for two-stage dynamic linear panel data models.[24]

Table G.10 compares results of the B2S2S estimator to those of the two-stage QML sequential approach (TSQML). Once again, the estimates are very close to one another. As soon as N and T increase, the very slight biases observed on the parameters tend to disappear. For the B2S2S, the estimates of the variance of the specific effects, as well as that of the remainder disturbances, do not appear to be biased. If the RMSE of B2S2S and TSQML for ϕ, ρ, δ, β_1 and β_2 are close each other, it is not the same for the coefficient η_2 associated with the time-invariant variable $Z_{2,i}$ which is itself correlated with μ_i. The RMSE of the coefficient η_2 for B2S2S is half the size of TSQML and this ratio remains the same when going from $N = 63$, $T = 10$ to $N = 120$, $T = 20$. Interestingly, the standard errors of that same coefficient η_2 are smaller when using the Bayesian estimator as compared to the two-stage QMLE. Even with a slight bias, the 95% confidence intervals of the Bayesian estimator of η_2 are narrower and entirely nested within those obtained with the two-stage QML sequential approach. We also reached the same conclusion in non-spatial static and dynamic models (see Baltagi et al. (2018, 2021)). Finally, note that the computation times of the two-stage QML sequential approach are 34 (resp. 1.8) times longer than those of the B2S2S with mixture of t-distributions) (resp. with bootstrap).

[23] For the following specification written in primal form: $y_{it} = \phi y_{i,t-1} + \rho y^*_{it} + \delta y^*_{i,t-1} + x'_{it}\beta + V_i'\eta + \mu_i + u_{it}$, the first stage model is $y_{it} = \phi y_{i,t-1} + \rho y^*_{it} + \delta y^*_{i,t-1} + x'_{it}\beta + \overline{\kappa} + e_{it}$, where $e_{it} = \kappa_i - \overline{\kappa} + u_{it}$, $\kappa_i = V_i'\eta + \mu_i$, $\overline{\kappa} = E[\kappa_i]$ and is estimated in first differences. In the second stage, one estimates the coefficients η based on the level relationship: $y_{it} - \widehat{\phi} y_{i,t-1} - \widehat{\rho} y^*_{it} - \widehat{\delta}_y y^*_{i,t-1} - x'_{it}\widehat{\beta} = V_i'\eta + \vartheta_{it}$ where $\vartheta_{it} = \mu_i + u_{it} + (\widehat{\phi} - \phi) y_{i,t-1} - (\widehat{\rho} - \rho) y^*_{it} - (\widehat{\delta}_y - \delta) y^*_{i,t-1} - x'_{it}(\widehat{\beta} - \beta)$ and computes proper standard errors with an analytical correction term (see Kripfganz and Schwarz (2019)).

[24] Following Kripfganz and Schwarz (2019), we use successively these two Stata commands ("xtdpdqml" and "xtseqreg"). Unfortunately, these Stata commands do not give the residual variance of specific effects σ^2_μ but only σ^2_u.

Table G.10: Dynamic Space-Time Hausman-Taylor World with row normalized inverse distance weighting matrix
$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ϕ	ρ	δ	β_1	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1	1	1	4			
$N = 63$	B2S2S coef	0.7482	0.8063	-0.6044	1.0016	1.0005	0.9932	1.0320	0.9929	3.9596	$< 10^{-4}$	0.4996	
$T = 10$	se_boot	0.0047	0.0205	0.0211	0.0146	0.0153	0.7413	0.0491	0.0830	0.8800			315.38
	se_mixt	0.0040	0.0258	0.0262	0.0156	0.0204	0.4994	0.0246	0.0580	0.8590			19.11
	rmse	0.0050	0.0214	0.0215	0.0147	0.0153	0.7413	0.0586	0.0833	0.8805			
	two-stage QML	0.7501	0.8065	-0.6064	0.9996	0.9997	0.9786	0.9913	0.9903	$n.a$			595.37
	se	0.0044	0.0155	0.0163	0.0116	0.0116	0.5682	0.1214	0.0615	$n.a$			
	rmse	0.0044	0.0176	0.0183	0.0120	0.0119	0.6781	0.1244	0.0622	$n.a$			
$N = 120$	B2S2S coef	0.7491	0.8046	-0.6034	1.0008	1.0007	1.0092	1.0190	0.9979	3.9818	$< 10^{-4}$	0.4999	
$T = 20$	se_boot	0.0021	0.0114	0.0117	0.0069	0.0068	0.3214	0.0250	0.0439	0.5706			1486.29
	se_mixt	0.0021	0.0151	0.0153	0.0082	0.0092	0.2249	0.0126	0.0304	0.5656			76.89
	rmse	0.0023	0.0123	0.0122	0.0069	0.0068	0.3215	0.0314	0.0439	0.5706			
	two-stage QML	0.7500	.08048	-0.6044	0.9998	0.9999	1.0042	0.9985	0.9976	$n.a$			1875.13
	se	0.0017	0.0082	0.0084	0.0051	0.0051	0.2564	0.0638	0.0312	$n.a$			
	rmse	0.0017	0.0097	0.0097	0.0053	0.005	0.2770	0.0684	0.0313	$n.a$			

B2S2S : Bayesian two-stage two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}\,(b|g_0)$.

two-stage QML: two-stage quasi-maximum likelihood sequential approach with non available $(n.a)$ estimate of σ_μ^2.

G.4. The dynamic space-time homogeneous panel data world with correlated common effects: results of the Monte Carlo simulation study

Since the m common correlated effects f_t are now unknown, we need to rewrite the general dynamic space-time model as follows:

$$\begin{aligned} y =& Z\theta + Db + u = Z\theta + F\Gamma + u \\ & \text{with } Z'_{ti} = \left[y_{t-1,i}, y^*_{ti}, y^*_{t-1,i}, X'_{ti}\right] \text{ , } \theta' = [\phi, \rho, \delta, \beta']' \text{ and } X'_{ti} = [x_{ti}, x_{t-1,i}], \end{aligned}$$

where in the $(TN \times Nm)$ matrix F of the m unobserved factors, f should be approximated by known variables. Similar to the Hausman-Taylor case (see eq(G.47)), we can approximate the $(T \times m)$ f matrix with a $(T \times K_1)$ f^* matrix of the within time transformation[25] of Z_{ti}:

$$f^* = \begin{pmatrix} f^*_1 \\ \dots \\ f^*_T \end{pmatrix} \qquad \text{where } f^*_t = \left[\left(\overline{y}_{-1,t} - \overline{\overline{y}}_{-1}\right), \left(\overline{y}^*_t - \overline{\overline{y}}^*\right), \left(\overline{y}^*_{-1,t} - \overline{\overline{y}}^*_{-1}\right), \left(\overline{x}_t - \overline{\overline{x}}\right), \left(\overline{x}_{-1,t} - \overline{\overline{x}}_{-1}\right)\right]$$

$$\text{with } \overline{x}_t = (1/N)\sum_{i=1}^{N} x_{ti}, \, \overline{\overline{x}} = (1/NT)\sum_{i=1}^{N}\sum_{t=1}^{T} x_{ti}$$

Then, the product $F\Gamma$ is approximated with the product $F^*\Gamma^*$ where the factor loadings Γ^* is a $(NK_1 \times 1)$ vector and F^* is a $(TN \times NK_1)$ matrix of the within time transformations of Z. As Chudik and Pesaran (2015a), we can approximate the $(T \times m)$ f matrix by the time means of the dependent and explanatory variables. We follow the method of Chudik and Pesaran (2015a,b) by introducing the time means of the dependent and explanatory variables instead of introducing only the within time transformation of the explanatory variables Z'_{ti}. We compare our B2S2S estimator with the 2SLS estimator of Yang (2021) extended to the dynamic space-time case.[26]

Table G.11 shows that the results of B2S2S are very close to those of 2SLS. As for the previous case, we find the same qualities of our estimator (efficiency, computation time saving, absence of bias, ...) as compared to the 2SLS estimator.

[25] *i.e.*, the demeaned time means.

[26] See section E of the supplementary material for more details on the 2SLS estimator of Yang (2021) extended to the dynamic space-time homogeneous case. We use our own R codes for our Bayesian estimator and the 2SLS estimator.

Table G.11: Dynamic Space-Time Homogeneous Panel Data Model with Common Correlated Effects and row normalized inverse distance weighting matrix

$\varepsilon = 0.5$, $r = 0.8$, Replications=1,000

		ϕ	ρ	δ	β_1	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	0.8	-0.6	1	1	1			
$N = 63$	B2S2S coef	0.7495	0.8075	-0.6066	1.0001	1.0010	0.9882	$< 10^{-4}$	0.5924	
$T = 30$	se_boot	0.0021	0.0106	0.0109	0.0093	0.0109	0.0350			3754.26
	se_mixt	0.0027	0.0114	0.0118	0.0126	0.0147	0.0350			153.78
	rmse	0.0022	0.0130	0.0128	0.0093	0.0109	0.0369			
	2SLS coef	0.7508	0.8060	-0.6078	1.0000	0.9976	0.9919			624.72
	se	0.0138	0.0139	0.0230	0.0090	0.0369	0.0897			
	rmse	0.0142	0.0176	0.0261	0.0089	0.0376	0.0900			
$N = 63$	B2S2S coef	0.7497	0.8075	-0.6065	0.9997	1.0008	0.9863	$< 10^{-4}$	0.5325	
$T = 50$	se_boot	0.0016	0.0081	0.0083	0.0073	0.0084	0.0258			4029.75
	se_mixt	0.0021	0.0088	0.0090	0.0099	0.0115	0.0250			210.23
	rmse	0.0016	0.0110	0.0105	0.0073	0.0084	0.0292			
	2SLS coef	0.7512	0.8055	-0.6070	0.9992	0.9968	1.0158			1382.43
	se	0.0131	0.0104	0.0181	0.0091	0.0369	0.0863			
	rmse	0.0134	0.0137	0.0212	0.0091	0.0377	0.0877			
$N = 120$	B2S2S coef	0.7495	0.8097	-0.6081	1.0002	1.0013	0.9934	$< 10^{-4}$	0.6090	
$T = 30$	se_boot	0.0016	0.0104	0.0106	0.0069	0.0080	0.0255			11238.82
	se_mixt	0.0020	0.0111	0.0112	0.0093	0.0108	0.0255			326.90
	rmse	0.0016	0.0142	0.0133	0.0069	0.0081	0.0263			
	2SLS coef	0.7512	0.8101	-0.6125	0.9995	0.9967	0.9851			1720.27
	se	0.0126	0.0120	0.0201	0.0090	0.0350	0.0832			
	rmse	0.0134	0.0188	0.0274	0.0095	0.0371	0.0845			
$N = 120$	B2S2S coef	0.7497	0.8098	-0.6084	1.0002	1.0008	0.9935	$< 10^{-4}$	0.5490	
$T = 50$	se_boot	0.0012	0.0080	0.0081	0.0052	0.0061	0.0204			13014.95
	se_mixt	0.0016	0.0085	0.0086	0.0073	0.0084	0.0204			629.78
	rmse	0.0012	0.0127	0.0117	0.0052	0.0061	0.0214			
	2SLS coef	0.7514	0.8099	-0.6124	0.9996	0.9960	1.0137			4718.60
	se	0.0128	0.0088	0.0163	0.0070	0.0351	0.0950			
	rmse	0.0134	0.0149	0.0221	0.0073	0.0363	0.0959			

B2S2S : Bayesian two-stage two-step estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

2SLS: two-stage least squares estimator of Yang (2021) extended to the case of a dynamic space-time model.

G.5. The dynamic space-time heterogeneous panel data world with correlated common effects: results of the Monte Carlo simulation study

The dynamic space-time heterogeneous panel data world with common factors is defined as:

$$y_{ti} = \phi_i y_{t-1,i} + \rho_i y^*_{ti} + \delta_i y^*_{t-1,i} + x_{ti}\beta_{1i} + x_{t-1,i}\beta_2 + f'_t\gamma_i + u_{ti}$$

This model cannot be estimated using the common correlated effects mean group estimator (CCEMG) (see Pesaran (2006) and Chudik and Pesaran (2015a,b)). But, we propose an extension of the 2SLS estimator of Yang (2021) to the case of a dynamic space-time heterogeneous panel data world with correlated common factors.[27] So we compare the mean coefficients $\widehat{\overline{\theta}} = (1/N)\sum_{i=1}^{N}\widehat{\theta}_i$ of our B2S2S estimator with the 2SLS estimator. While the bottom panel of Table G.12 gives insights on the distribution of ϕ_i , ρ_i, δ_i and β_{1i} for different sample sizes, the top panel of Table G.12 gives the estimated values of the mean coefficients $\overline{\phi}$, $\overline{\rho}$, $\overline{\delta}$ and $\overline{\beta}_1$, the estimated values of β_2 and σ_u^2, their standard deviations and their RMSE's. Table G.12 shows that the results of the B2S2S estimator are close to those of the 2SLS estimator but the RMSEs results of B2S2S are generally smaller than those of 2SLS. Once again, as the computation time of B2S2S with bootstrap is longer, we give only results for B2S2S with mixture of t-distributions whose computation times are very close to those of 2SLS, this time. But the undeniable advantage of our estimator is its better efficiency relative to that of the IV estimator.

[27]See section F of the supplementary material for more details on the 2SLS estimator of Yang (2021) extended to the dynamic space-time heterogeneous case. We use our own R codes for our Bayesian estimator and the 2SLS estimator.

Table G.12: Dynamic Space-Time Heterogeneous Panel Data Model with Common Correlated Effects and row normalized inverse distance weighting matrix, $\varepsilon = 0.5$, $r = 0.8$, Replications=1, 000

		$\overline{\phi}$	$\overline{\rho}$	$\overline{\delta}$	$\overline{\beta}_1$	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.7501	0.7999	-0.6000	0.7498	1	1			
$N = 63$	B2S2S coef	0.7453	0.7745	-0.5853	0.7483	1.0029	0.9960	$< 10^{-4}$	0.5924	
$T = 30$	se_mixt	0.0114	0.0624	0.0604	0.0200	0.0090	0.0377			227.11
	rmse	0.0123	0.0673	0.0621	0.0200	0.0094	0.0379			
	2SLS coef	0.7454	0.7643	-0.5611	0.7473	1.0014	0.9661			58.26
	se	0.0135	0.1002	0.1134	0.0209	0.0120	0.0752			
	rmse	0.0143	0.1063	0.1198	0.0210	0.0121	0.0825			
$N = 63$	B2S2S coef	0.7477	0.7873	-0.5957	0.7488	1.0018	0.9993	$< 10^{-4}$	0.3196	
$T = 50$	se_mixt	0.0112	0.0410	0.0408	0.0189	0.0063	0.0272			238.91
	rmse	0.0115	0.0429	0.0410	0.0190	0.0066	0.0272			
	2SLS coef	0.7487	0.7823	-0.5774	0.7487	1.0008	1.1097			87.01
	se	0.0125	0.0722	0.0900	0.0194	0.0100	0.0701			
	rmse	0.0126	0.0743	0.0928	0.0194	0.0100	0.1302			
	true	0.75	0.8001	-0.6001	0.7495	1	1			
$N = 120$	B2S2S coef	0.7449	0.7968	-0.6059	0.7480	1.0033	1.0010	$< 10^{-4}$	0.1136	
$T = 30$	se_mixt	0.0086	0.0547	0.0523	0.0147	0.0065	0.0274			479.94
	rmse	0.0100	0.0547	0.0526	0.0148	0.0073	0.0275			
	2SLS coef	0.7453	0.7666	-0.5602	0.7477	1.0020	0.9710			106.32
	se	0.0100	0.0990	0.1141	0.0157	0.0084	0.0502			
	rmse	0.0111	0.1045	0.1208	0.0158	0.0087	0.0580			
$N = 120$	B2S2S coef	0.7475	0.8086	-0.6159	0.7492	1.0023	1.0024	$< 10^{-4}$	0.1136	
$T = 50$	se_mixt	0.0083	0.0362	0.0341	0.0138	0.0045	0.0207			555.21
	rmse	0.0086	0.0372	0.0376	0.0138	0.0051	0.0208			
	2SLS coef	0.7485	0.7843	-0.5809	0.7491	1.0010	1.1093			167.59
	se	0.0092	0.0711	0.0863	0.0141	0.0068	0.0544			
	rmse	0.0093	0.0728	0.0884	0.0141	0.0069	0.1220			

B2S2S : Bayesian two-stage two-step estimation.

se_mixt: standard errors of $\theta = (\phi, \rho, \delta, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

2SLS: two-stage least squares estimator of Yang (2021) extended to the case of a dynamic space-time model.

$\overline{\phi} = (1/N)\sum_{i=1}^N \phi_i$, $\overline{\rho} = (1/N)\sum_{i=1}^N \rho_i$, $\overline{\delta} = (1/N)\sum_{i=1}^N \delta_i$ and $\overline{\beta}_1 = (1/N)\sum_{i=1}^N \beta_{1i}$.

Table G.12: (cont'd) Dynamic Space-Time Heterogeneous Panel Data Model with Common Correlated Effects. Distribution of ϕ_i, ρ_i, δ_i and β_{1i} for different sample sizes

	$N = 63$				$N = 120$			
	ϕ_i	ρ_i	δ_i	β_{1i}	ϕ_i	ρ_i	δ_i	β_{1i}
min	0.6047	0.6548	-0.8125	0.5078	0.6026	0.6525	-0.8240	0.5041
mean	0.7501	0.7999	-0.6000	0.7498	0.7500	0.8001	-0.6001	0.7495
sd	0.0865	0.0864	0.0952	0.1444	0.0866	0.0866	0.0953	0.1443
max	0.8953	0.9455	-0.4203	0.9924	0.8976	0.9476	-0.4114	0.9960

H. Application on crop yields and climate change

H.1. The dataset

Keane and Neal (2020) use weather and crop yield data for U.S. counties from 1950 to 2015. They have excluded counties west of the 100th Meridian[28] and counties with less than 30 years of data. This gives $N = 2,209$ corn-growing counties with 30% of unbalanced data for corn yields. Keane and Neal (2020) defined the annual growing (resp. killing) degree days gdd_{ti} (resp. kdd_{ti}) values by summing the daily degree days measures. The hours each day a crop is exposed to one-degree C° temperature intervals is approximated using a sinusoidal function:

$$dd_C = \begin{cases} 0 & \text{if } C > T_{\max}, \\ T_{avg} - C & \text{if } C < T_{\min}, \\ \pi \left[(T_{avg} - C) \cos^{-1}(S) + (T_{\max} - T_{\min}) \sin(S)/2 \right] & \text{otherwise,} \end{cases}$$

where C is the temperature in Celsius, $T_{\max}$, $T_{\min}$ are the daily max/min temperatures, $T_{avg} = (T_{\max} + T_{\min})/2$ and $S = \cos^{-1}\left(\frac{2C - T_{\max} - T_{\min}}{T_{\max} - T_{\min}}\right)$. Then, the daily growing (resp. killing) degree day is $gdd_{id} = dd_0 - dd_{29}$ (resp. $kdd_{id} > dd_{29}$) for each county i and each day d. These values are summing over around 150 days from April 1st to September 30th (see Keane and Neal (2020) p.1406). They are expressed in total hours over the growing season. Precipitation is measured as total inches of rain over the growing season.

Since some values of corn yields are missing for counties, we interpolated these missing values using the inverse distance weighted method. This method uses a weighted average of non-missing values, the weights being reciprocals of the powered distance between values, the power being zero or positive (see Fisher et al. (1993)). We set the power equal to 2. Thus with power 2, values at distance 1 from a point with unknown values have weight 1, values at distance 2 from a point have weight 1/4, distance 3 weight 1/9, and so forth. Missing data concern only the corn yield variable. This variable with missing data was processed using inverse distance weighted method to obtain satisfactory imputations, close to those obtained with cubic B-splines. We also tried multiple imputation, using Bootstrap-based expectation-maximization (EM) algorithms proposed by Honaker and King (2010) and Honaker et al. (2011) but we got implausible values, mainly for the oldest or most recent years. Our choice of using inverse distance weighted smoothing rather than a multiple imputation method (like MICE (multiple imputation by chained equations) or EM, (see White et al. (2011)) is reinforced by the results of Yoon et al. (2017). Many modern missing data methods (*e.g.*, multiple imputation, FIML, EM, ...) assume missing at random. Yoon et al. (2017) have compared the most familiar methods for estimating missing data. They show that recurrent neural networks (RNN), just followed by cubic splines, give the best results (*i.e.*, smallest rmse) as compared to imputation (MICE or EM) (see also Baltagi et al. (2019)). This gives us a balanced dataset of $N = 2,678$ corn-growing counties over $T = 66$ years (1950-2015), *i.e.,* 176,748 observations per variable.

To define the spatial weight matrix, the counties coordinates are taken from an ESRI Shapefile downloaded from the US Census.[29] Using the 2,678 county spatial polygons read from the ESRI Shapefile mentioned above, we first created the county distance matrix. The county distances

[28] The 100th Meridian separates the Great Plains to the east from the semi-arid lands to the west. The western counties are much more reliant on irrigation.

[29] US Census `ftp://ftp2.census.gov/geo/tiger/TIGER2008/tl_2008_us_county00.zip`.

are great-circle distances calculated using the Haversine formula based on internal points in the geographic area. The Haversine formula is given by

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\Phi_2 - \Phi_1}{2}\right) + \cos\Phi_1 \cos\Phi_2 \sin^2\left(\frac{\Lambda_2 - \Lambda_1}{2}\right)}\right)$$

d is the distance between the two points along a great circle of the sphere (Earth). It is the spherical distance (*i.e.*, the shortest distance between two points on the surface of a sphere). r is the radius of the sphere (637.8137 km for Earth). Φ_1 and Φ_2 are the latitudes of point 1 and of point 2 (in radians). Λ_1 and Λ_2 are the longitudes of point 1 and of point 2 (in radians).

For the $2,678$ counties, the minimum distance is 11.59 km between Lancaster County (Virginia) and Middlesex County (Virginia) and the maximum distance is $4,390$ km between Lane County (Oregon) and Suffolk County (New York state).[30] From this county distances matrix, we created a row-normalized inverse square distance spatial weight matrix using the 5 nearest neighbors.[31,32] In the dataset, the missing (continental) states are Alaska, Connecticut, District of Columbia, Maine, Massachusetts, Nevada, New Hampshire, Rhode Island and Vermont.

The specific climatology of each county is defined according to the Köppen climate classification (see below) (see also Kottek et al. (2006) and Aparicio-Ruiz et al. (2018)).[33] Approximately, to the east of the 100th meridian, the climate ranges from humid continental in the north to humid subtropical in the south. The Great Plains west of the 100th meridian is semi-arid. Much of the Western mountains have an alpine climate. The climate is arid in the Great Basin, desert in the Southwest, Mediterranean in coastal California, and oceanic in coastal Oregon and Washington and southern Alaska. Most of Alaska is subarctic or polar. Hawaii and the southern tip of Florida are tropical, as being the populated territories in the Caribbean and the Pacific.

[30] The percentiles (min, 5%, 25%, mean, 75%, 95%, max) of all the county distances (in km) are the following : 11.59, 282.93, 694.73, 1218.29, 1601.30, 2606.24, 4390.60.

[31] The choice of such a number of nearest neighbors is defined to be consistent with Global Climate Models (GCMs) projections. Indeed, these GCMs, used for climate studies and climate projections, are typically run at spatial resolutions of the order of 150 to 200 km (Flato et al. (2014)). But the Keane and Neal (2020) dataset we use, has been converted to county-level projections using the interpolation procedure called "bias-correction and spatial disaggregation" for an appropriate scale of assessing impacts (see Wood et al. (2004), Piani et al. (2010), Li et al. (2015)).

[32] The percentiles (min, 5%, 25%, mean, 75%, 95%, max) of the 5 nearest county distances (in km) are the following: 11.59, 25.82, 34.32, 46.47, 50.30, 89.27, 300.27.

[33] The Köppen-Geiger climate classification for U.S. states and counties has been downloaded from `http://koeppen-geiger.vu-wien.ac.at/data/KoeppenGeiger.UScounty.txt`.

Table H.1: Koppen-Geiger Climate Classification

Climate class	Climate name
Af	Tropical rainforest climate
Am	Tropical monsoon climate
Aw	Tropical wet and dry
Bwh	Warm desert climate
BSh	Warm semi-arid climate
BWk	Cold desert climate
BSk	Cold semi-arid climate
Csa	Warm Mediterranean climate
Csb	Temperate Mediterranean climate
Cfa	Warm oceanic climate/Humid subtropical climate
Cfb	Temperate oceanic climate
Cfc	Subpolar oceanic climate
Cwa	Monsoon-influenced humid subtropical climate
Cwb	Monsoon-influenced temperate oceanic climate
Dfa	Warm/Humid continental climate
Dfb	Temperate/Humid continental climate
Dfc	Cool continental climate/Subarctic climate
Dwa	Warm/Humid continental climate
Dwb	Temperate/Humid continental climate
Dwc	Cool continental climate/Subarctic climate
Dsa	Warm/Mediterranean continental climate
Dsb	Temperate/Mediterranean continental climate
Dsc	Mediterranean-influenced subarctic climate
ET	Tundra climate

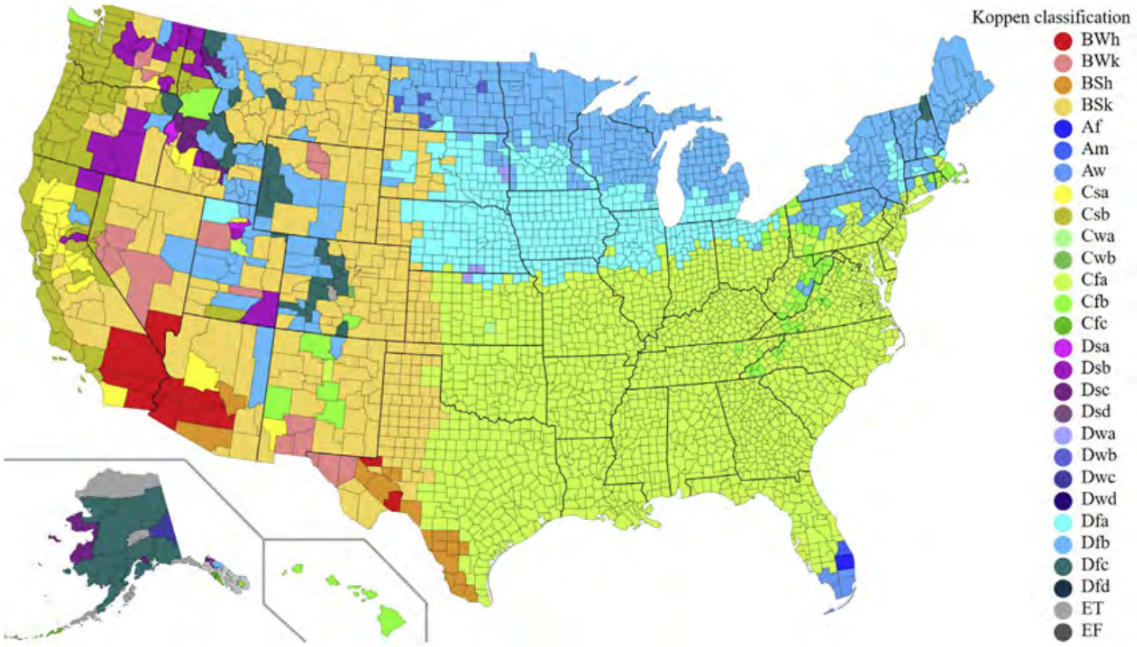


Figure 4: U.S Köppen-Geiger climate classification (*source: Aparicio-Ruiz et al. (2018) pp. 164.*)

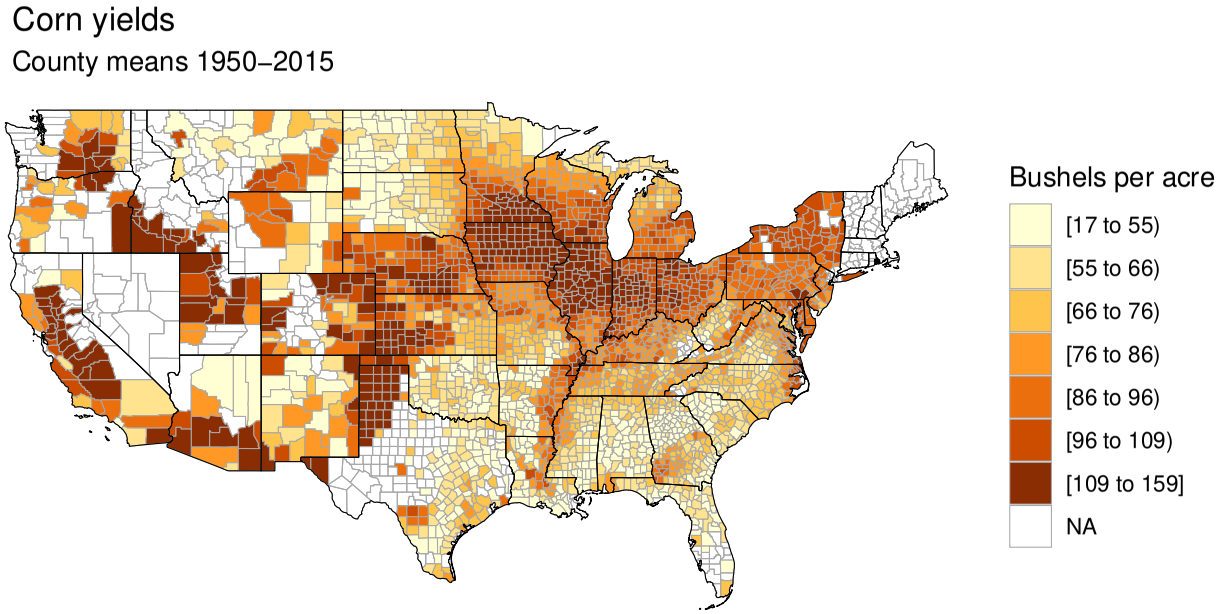


Figure 5: US county means of corn yields over 1950-2015.

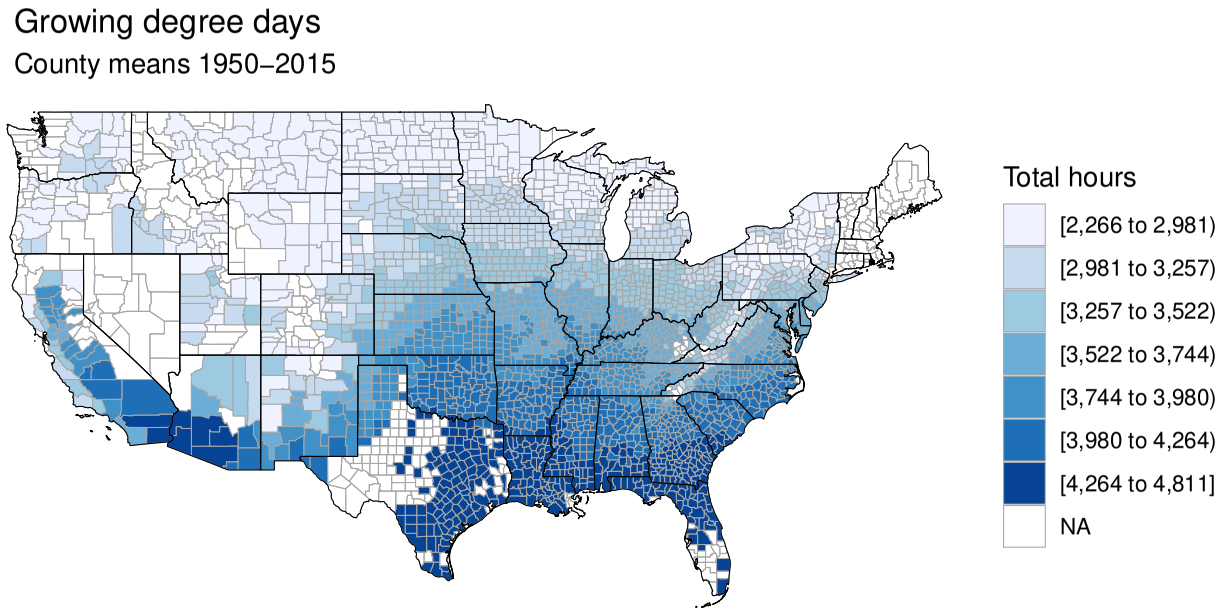


Figure 6: US county means of growing degree days over 1950-2015.

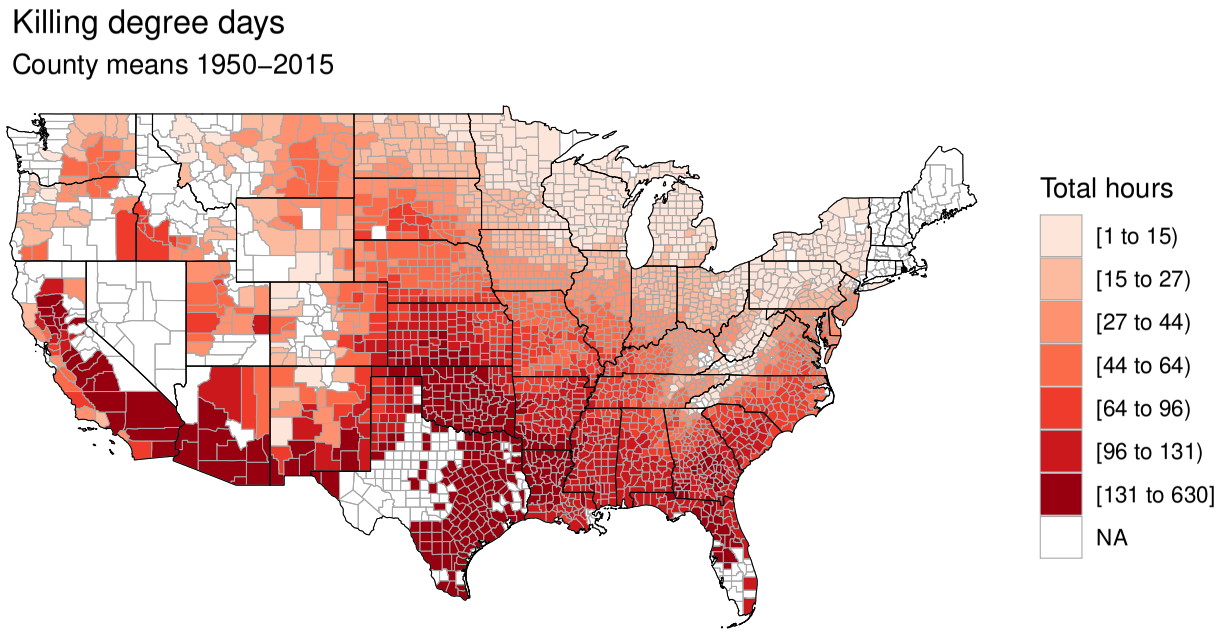


Figure 7: US county means of killing degree days over 1950-2015.

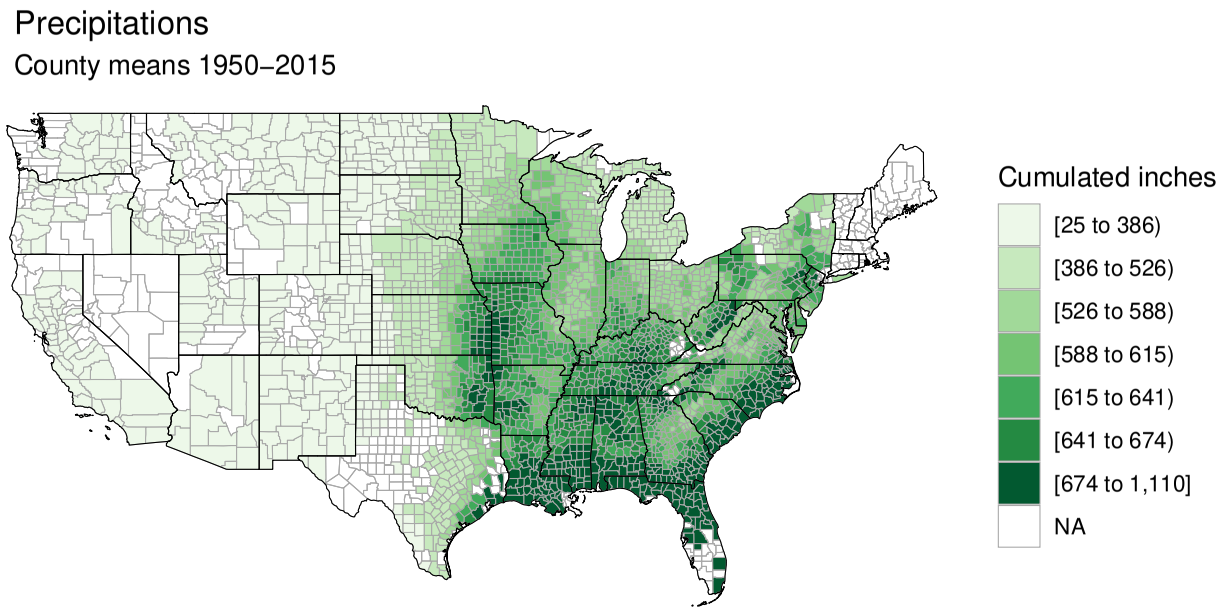


Figure 8: US county means of precipitations over 1950-2015.

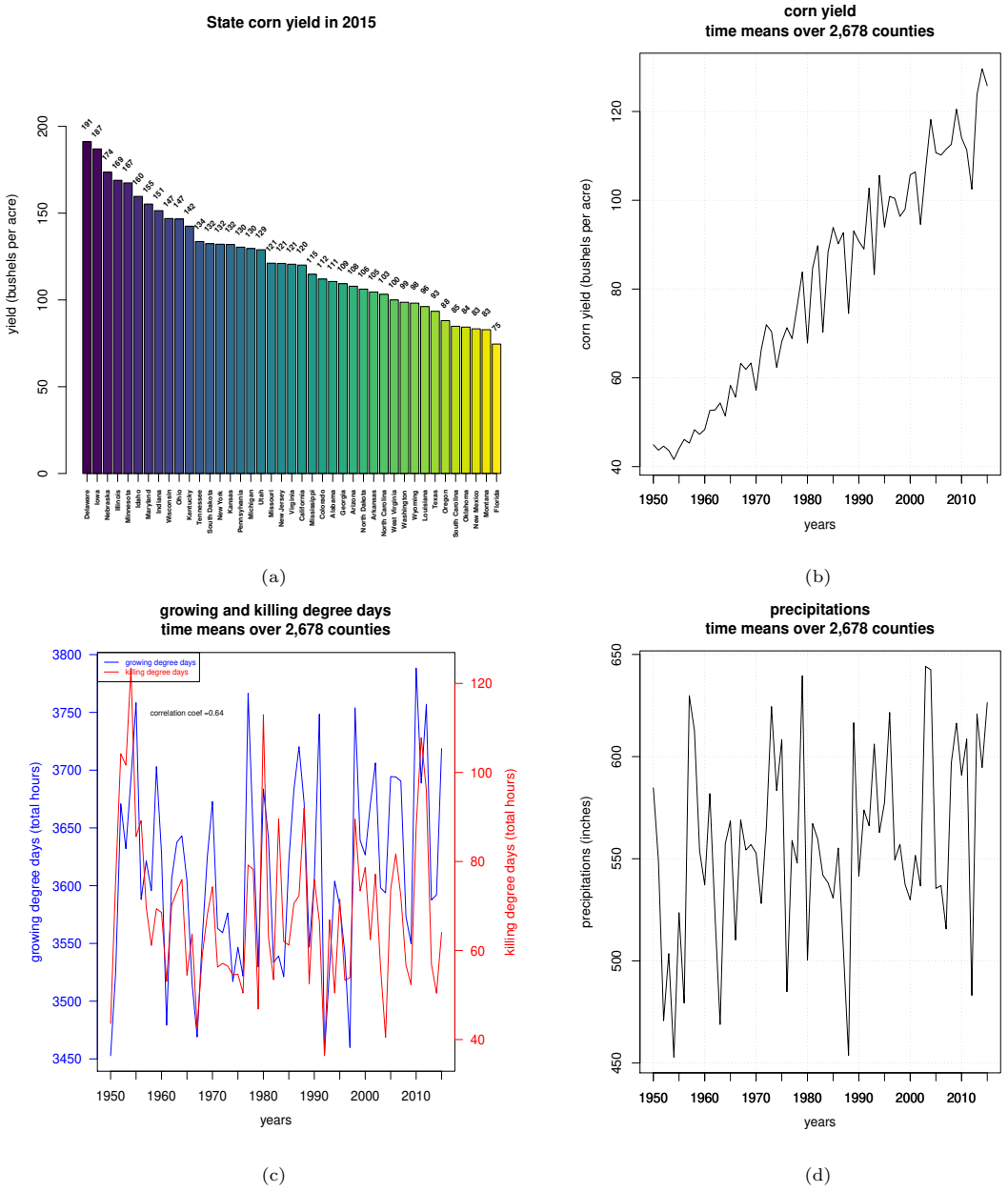


Figure 9: (a) State means of corn yield for 2015. (b) Time means of corn yield over 2,678 counties. (c) Time means of growing and killing degree days over 2,678 counties. (d) Time means of precipitation over 2,678 counties.

Table H.2: Descriptive statistics of corn yield, growing and killing degree days and precipitations for the $N = 2,678$ counties and the $T = 66$ years (1950-2015), ($NT = 176,748$ observations).

		mean	std.dev	min	max
corn yield	overall	80.58	40.76	12	254.44
	between		24.27	16.65	159.37
	within		32.75	-34.60	233.62
growing degree days	overall	3613.76	548.84	2038.88	4931.37
	between		536.88	2265.88	4810.52
	within		114.42	3060.46	4030.19
killing degree days	overall	68.83	67.77	0.4	735.25
	between		60.85	0.5232	629.68
	within		29.85	-86.57	333.87
precipitations	overall	557.17	201.51	5.04	1693.67
	between		159.41	25.29	1110.43
	within		123.29	91.87	1248.43

Table H.3: Köppen-Geiger climate classification dummies (frequency in percent).

Aw	Tropical wet and dry climate	0.07
BSh	Warm semi-arid climate	0.37
BSk	Cold semi-arid climate	8.25
BWh	Warm desert climate	0.18
BWk	Cold desert climate	0.11
Cfa	Warm oceanic climate/Humid subtropical climate	54.29
Cfb	Temperate oceanic climate	1.68
Csa	Warm Mediterranean climate	0.71
Csb	Temperate Mediterranean climate	1.45
Dfa	Warm/Humid continental climate	18.37
Dfb	Temperate/Humid continental climate	13.51
Dfc	Cool continental climate/Subarctic climate	0.37
Dsb	Warm/Humid continental climate	0.37
Dsc	Temperate/Humid continental climate	0.03
Dwa	Cool continental climate/Subarctic climate	0.01
Dwb	Temperate/Mediterranean continental climate	0.11

MO-OLS estimation of the model

$$\log y_{ti} \quad = \quad \beta_{1,ti} gdd_{ti} + \beta_{2,ti} kdd_{ti} + \beta_{3,ti} prec_{ti} + \beta_{4,ti} prec_{ti}^2 + c_{ti} + u_{ti} \ , \ i = 1, ..., N \ , \ t = 1, ..., T$$

Table H.4: MO-OLS estimates of the impacts of temperatures and precipitations on U.S. corn yields for the $N = 2,678$ counties and the $T = 65$ years (1951-2015), ($NT = 174,070$ observations).

	mean	se	min	10th	25th	75th	90th	max
gdd	0.00032	0.00002	-0.00514	-0.00054	0.00005	0.00073	0.00110	0.00285
kdd	-0.00655	0.00014	-0.05383	-0.01279	-0.00970	-0.00252	0.00036	0.01154
$prec$	0.00084	0.00011	-0.01984	-0.00221	-0.00075	0.00231	0.00415	0.04115
$prec^2(\div 10^3)$	-0.00066	0.00017	-0.15584	-0.00396	-0.00199	0.00055	0.00183	0.10422
$intercept$	3.20889	0.07280	-0.80934	1.73755	4.26773	6.21678	7.92130	25.48345
R^2	0.79367							
σ_u^2	0.07153							

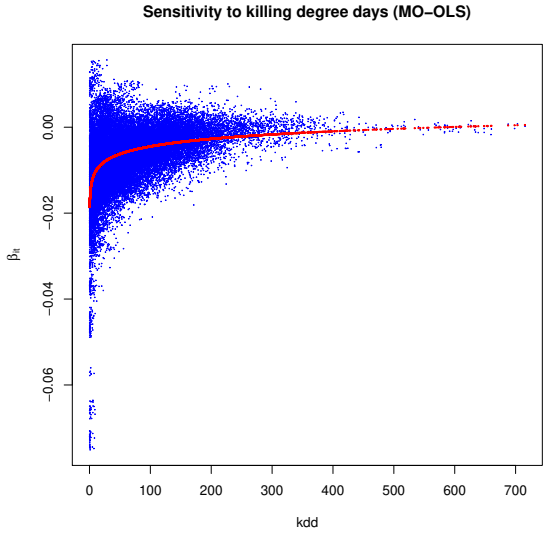


Figure 10: Relationship between the climatology of *kdd*s and the sensitivity of yield to *kdd* ($\hat{\beta}_{2,ti}$). This graph is a scatter plot of a random 20% subsample of $\hat{\beta}_{2,ti}$ against kdd_{ti} in order to reduce the too large size of the postscript file with 100% of the sample. The figure with the full sample is available on request.

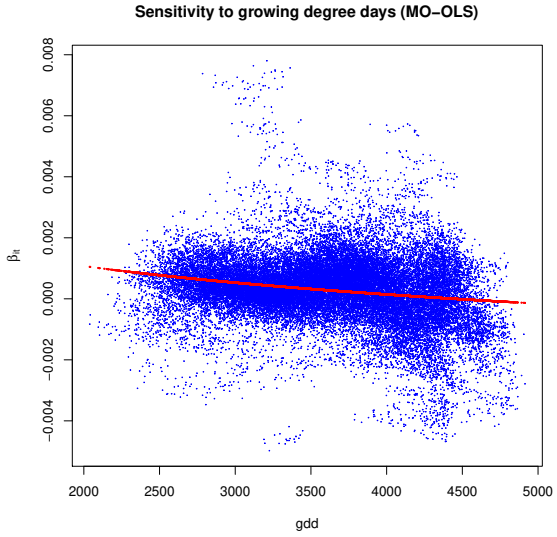


Figure 11: Relationship between the climatology of *gdd*s and the sensitivity of yield to *gdd* ($\hat{\beta}_{1,ti}$). This graph is a scatter plot of a random 20% subsample of $\hat{\beta}_{1,ti}$ against gdd_{ti} in order to reduce the too large size of the postscript file with 100% of the sample. The figure with the full sample is available on request.

H.2. Direct effects, spatial spillover effects and total effects

Pooling the N counties, we can rewrite the estimated specification as

$$\log y_t = \phi \log y_{t-1} + \rho W_N \log y_t + \delta W_N \log y_{t-1} + X_t\beta + (I_N \otimes f_t')\,\Gamma + u_t, \qquad \text{(H.49)}$$

where $\beta = (\beta_1, ..., \beta_6)'$ and $\Gamma = (\gamma_{11}, \cdots, \gamma_{1m}, \cdots, \gamma_{N1}, \cdots, \gamma_{Nm})'$ with $m = 3$.
X is defined as $X = \left(gdd_t, Adapt\_gdd_t, kdd_t, Adapt\_kdd_t, prec_t, prec_t^2, V\right)$ where $Adapt\_gdd_t = \log(gdd_t)\, gdd_t - gdd_t$, $Adapt\_kdd_t = \log(kdd_t)\, kdd_t - kdd_t$ and where V are the Köppen-Geiger climate classification dummies. Then,

$$\log y_t = [I_N - \rho W_N]^{-1} \Big\{ (\phi I_N + \delta W_N) \log y_{t-1} + X_t\beta + (I_N \otimes f_t')\,\Gamma + u_t \Big\}, \qquad \text{(H.50)}$$

Recursively, we get

$$\log y_{t+\tau} = \left(B^{-1}A\right) \log y_{t-1} + \sum_{s=0}^{\tau} D_s \Big\{ X_{t+\tau-s}\beta + \left(I_N \otimes f_{t+\tau-s}'\right)\Gamma + u_{t+\tau-s} \Big\}, \qquad \text{(H.51)}$$

where

$$\begin{aligned} B &= [I_N - \rho W_N] \\ A &= (\phi I_N + \delta W_N) \\ D_s &= \left(B^{-1}A\right)^s B^{-1} \end{aligned}$$

Then, leaving out the dummy variables, the τ-period-ahead (cumulative) multipliers of *gdd*, *kdd* and *prec* are given by

$$\begin{aligned} \frac{\partial \log y_{t+\tau}}{\partial gdd_t'} &= \sum_{s=0}^{\tau} D_s \Big[I_N\beta_1 + \beta_2 diag\left(\overline{\log(gdd)}_i\right)_{i=1,...,N} + \frac{1}{N} diag\left(\gamma_{1i}\right)_{i=1,...,N} \Big], \qquad \text{(H.52)} \\ \frac{\partial \log y_{t+\tau}}{\partial kdd_t'} &= \sum_{s=0}^{\tau} D_s \Big[I_N\beta_3 + \beta_4 diag\left(\overline{\log(kdd)}_i\right)_{i=1,...,N} + \frac{1}{N} diag\left(\gamma_{2i}\right)_{i=1,...,N} \Big], \\ \frac{\partial \log y_{t+\tau}}{\partial prec_t'} &= \sum_{s=0}^{\tau} D_s \Big[I_N\beta_5 + 2\beta_6 diag\left(\overline{prec}_i/10^3\right)_{i=1,...,N} + \frac{1}{N} diag\left(\gamma_{3i}\right)_{i=1,...,N} \Big], \end{aligned}$$

where $diag\left(\overline{\log(gdd)}_i\right)_{i=1,...,N}$ is a $(N \times N)$ diagonal matrix of individual means[34] of $\log(gdd)$ ($\overline{gdd}_i = 1/(T-1)\sum_{t=2}^{T} \log(gdd)_{ti}$) (*idem* for $diag\left(\overline{\log(kdd)}_i\right)_{i=1,...,N}$ and $diag\left(\overline{prec}_i\right)_{i=1,...,N}$). As $(1/N)diag\left(\gamma_{1i}\right)_{i=1,...,N} \to 0$ as N is large and as the γ_{1i} are small values (*idem* for γ_{2i} and γ_{3i}), the

[34] We compute these multipliers at the midpoint of the sample, but we could have selected observations from the last year of the sample (2015).

τ-period-ahead (cumulative) multipliers of *gdd*, *kdd* and *prec* are approximately

$$\begin{aligned}
\frac{\partial \log y_{t+\tau}}{\partial gdd_t'} &\simeq \sum_{s=0}^{\tau} D_s \Big[I_N \beta_1 + \beta_2 diag\left(\overline{\log(gdd)}_i \right)_{i=1,\ldots,N} \Big], \qquad \text{(H.53)} \\
\frac{\partial \log y_{t+\tau}}{\partial kdd_t'} &\simeq \sum_{s=0}^{\tau} D_s \Big[I_N \beta_3 + \beta_4 diag\left(\overline{\log(kdd)}_i \right)_{i=1,\ldots,N} \Big], \\
\frac{\partial \log y_{t+\tau}}{\partial prec_t'} &\simeq \sum_{s=0}^{\tau} D_s \Big[I_N \beta_5 + 2\beta_6 diag\left(\overline{prec}_i / 10^3 \right)_{i=1,\ldots,N} \Big],
\end{aligned}$$

and the impact multiplier ($\tau = 0$) is

$$\begin{aligned}
\frac{\partial \log y_t}{\partial gdd_t'} &\simeq B^{-1} \Big[I_N \beta_1 + \beta_2 diag\left(\overline{gdd}_i \right)_{i=1,\ldots,N} \Big], \qquad \text{(H.54)} \\
\frac{\partial \log y_t}{\partial kdd_t'} &\simeq B^{-1} \Big[I_N \beta_3 + \beta_4 diag\left(\overline{kdd}_i \right)_{i=1,\ldots,N} \Big], \\
\frac{\partial \log y_t}{\partial prec_t'} &\simeq B^{-1} \Big[I_N \beta_5 + 2\beta_6 diag\left(\overline{prec}_i / 10^3 \right)_{i=1,\ldots,N} \Big].
\end{aligned}$$

Since these (impact and cumulative) multipliers are $(N \times N)$ matrices, the (impact or cumulative) direct effect is the $(N \times 1)$ vector of the diagonal elements of these matrices. The (impact or cumulative) indirect effect is the $(N \times 1)$ vector of the row sums of the non-diagonal elements of these matrices. And the (impact or cumulative) total effect is the $(N \times 1)$ vector of the row sums of these matrices. The average direct, indirect and total effects are the averages of these $(N \times 1)$ vectors.

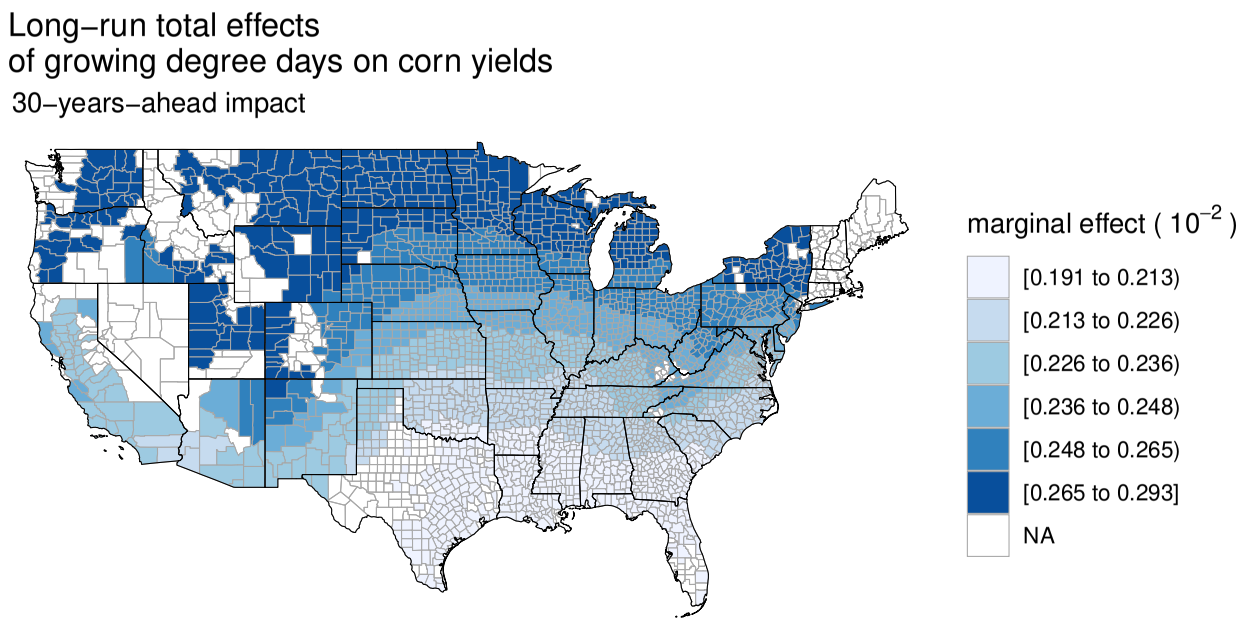


Figure 12: Long-run total effects of growing degree days on corn yields.

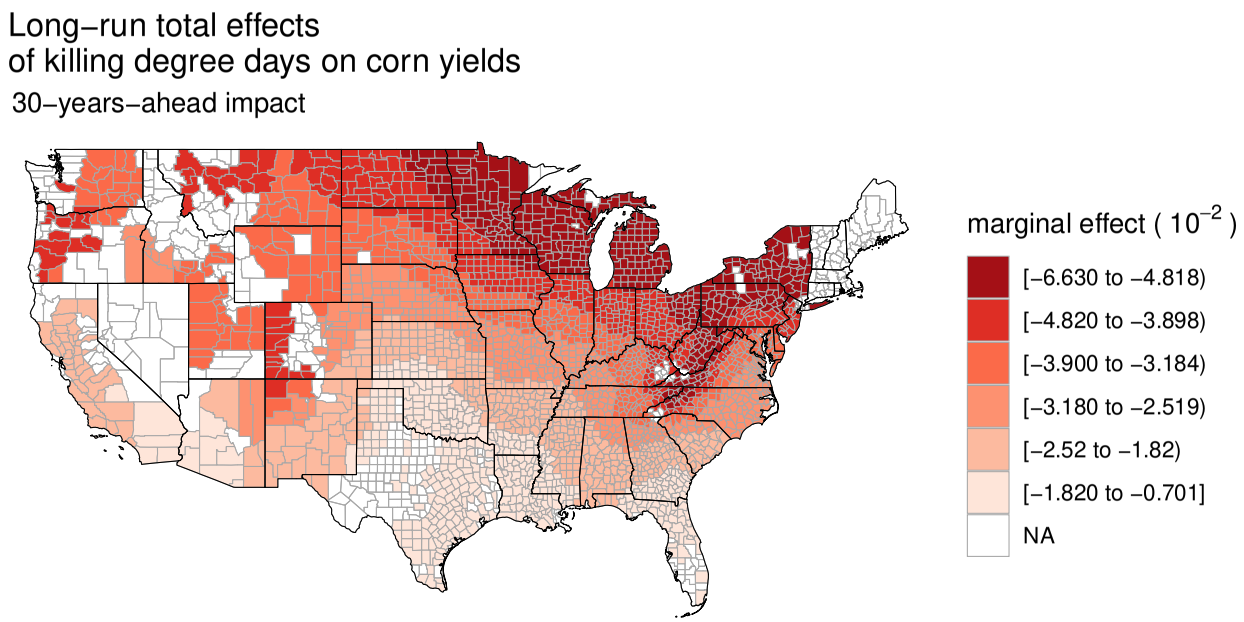


Figure 13: Long-run total effects of killing degree days on corn yields.

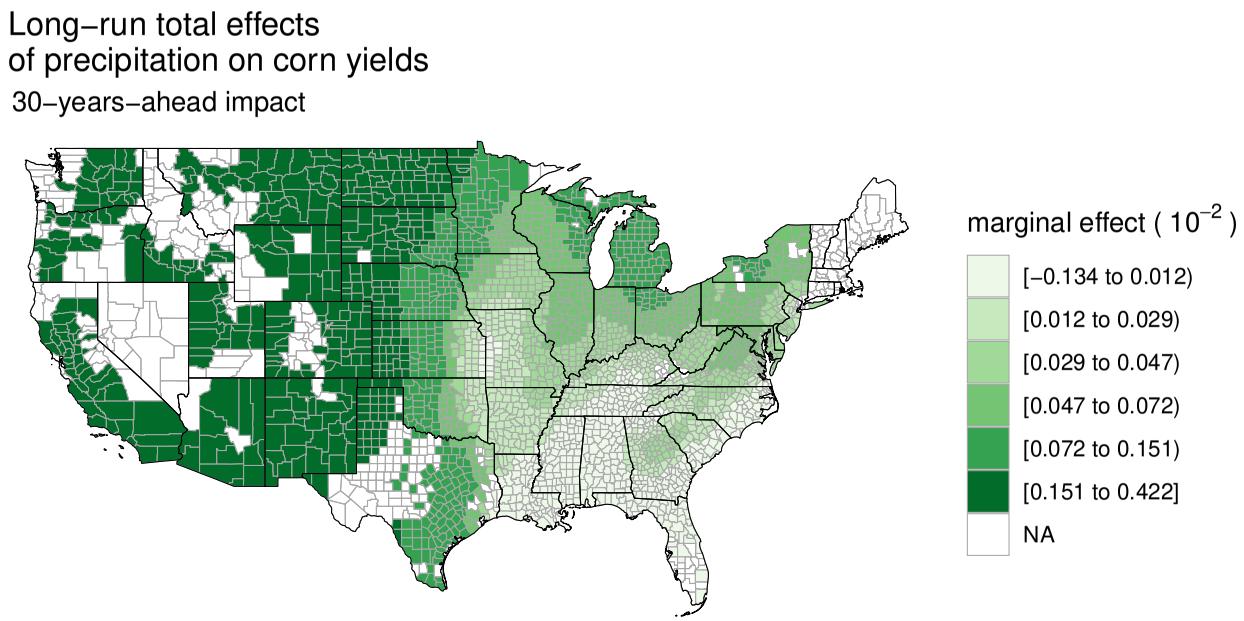


Figure 14: Long-run total effects of precipitations on corn yields.

References

Andersson, M., Karlsson, S., 2001. Bootstrapping error component models. Computational Statistics 16, 221–231.

Aparicio-Ruiz, P., Schiano-Phan, R., Salmerón-Lissén, J.M., 2018. Climatic applicability of downdraught evaporative cooling in the United States of America. Building and Environment 136, 162–176.

Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G., 2018. Robust linear static panel data models using ε-contamination. Journal of Econometrics 202, 108–123.

Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G., 2021. Robust dynamic panel data models using ε-contamination, in: Chudik, A., Hsiao, C., Timmermann, A. (Eds.), Advances in Econometrics, Essays in honor of M. Hashem Pesaran. Emerald Publishing (*forthcoming*).

Baltagi, B.H., Bresson, G., Etienne, J.M., 2019. Carbon dioxide emissions and economic activities: A mean field variational Bayes semiparametric panel data model with random coefficients. Annals of Economics and Statistics , 43–77.

Bellman, L., Breitung, J., Wagner, J., 1989. Bias correction and bootstrapping of error component models for panel data: Theory and applications. Empirical Economics 14, 329–342.

Berger, J., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Berger, J., Berliner, M., 1986. Robust Bayes and empirical Bayes analysis with ε-contaminated priors. Annals of Statistics 14, 461–486.

Chan, J., Koop, G., Poirier, D., Tobias, J., 2019. Bayesian Econometric Methods. 2nd ed., Cambridge University Press.

Chib, S., 2008. Panel data modeling and inference: a Bayesian primer, in: Mátyás, L., Sevestre, P. (Eds.), The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice, pp. 479–516.

Chib, S., Carlin, B.P., 1999. On MCMC sampling in hierarchical longitudinal models. Statistics and Computing 9, 17–26.

Chudik, A., Pesaran, M.H., 2015a. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393–420.

Chudik, A., Pesaran, M.H., 2015b. Large panel data models with cross-sectional dependence: A survey, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 3–45.

Debarsy, N., Ertur, C., LeSage, J.P., 2012. Interpreting dynamic space–time panel data models. Statistical Methodology 9, 158–171.

DeJong, D.N., Whiteman, C.H., 1991. Reconsidering 'trends and random walks in macroeconomic time series'. Journal of Monetary Economics 28, 221–254.

Fisher, N.I., Lewis, T., Embleton, B.J., 1993. Statistical Analysis of Spherical Data. Cambridge University Press.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al., 2014. Evaluation of climate models, in: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 741–866.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. 3rd ed., Chapmam & Hall/CRC press.

Gelman, A., Yao, Y., 2020. Holes in Bayesian statistics. Working Paper, Department of Statistics and Department of Political Science, Columbia University.

Ghosh, M., Heo, J., 2003. Default Bayesian priors for regression models with first-order autoregressive residuals. Journal of Time Series Analysis 24, 269–282.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1997. Markov Chain Monte Carlo in Practice. 2nd ed., Chapman & Hall, London, UK.

Greenberg, E., 2008. Introduction to Bayesian Econometrics. Cambridge University Press, Cambridge, UK.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.

Honaker, J., King, G., 2010. What to do about missing values in time-series cross-section data. American Journal of Political Science 54, 561–581.

Honaker, J., King, G., Blackwell, M., 2011. Amelia II: A program for missing data. Journal of Statistical Software 45, 1–47.

Hsiao, C., Pesaran, M.H., 2008. Random coefficient models, in: The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice. Mátyás, L. and Sevestre, P., pp. 185–214.

Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics 109, 107–150.

Ibazizen, M., Fellag, H., 2003. Bayesian estimation of an AR(1) process with exponential white noise. Statistics 37, 365–372.

Kapetanios, G., 2008. A bootstrap procedure for panel datasets with many cross-sectional units. Econometrics Journal 11, 377–395.

Karakani, H.M., van Niekerk, J., van Staden, P., 2016. Bayesian analysis of AR(1) model. arXiv preprint arXiv:1611.08747 .

Keane, M., Neal, T., 2020. Climate change and US agriculture: Accounting for multidimensional slope heterogeneity in panel data. Quantitative Economics 11, 1391–1429.

Koop, G., 2003. Bayesian Econometrics. Wiley, New York.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift (Contributions to Atmospheric Sciences , 259–263.

Kotz, S., Nadarajah, S., 2004. Multivariate t-distributions and their applications. Cambridge University Press.

Kripfganz, S., Schwarz, C., 2019. Estimation of linear dynamic panel data models with time-invariant regressors. Journal of Applied Econometrics 34, 526–546.

Lemoine, N.P., 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos 128, 912–928.

LeSage, J., Pace, R., 2009. An Introduction to Spatial Econometrics. CRC Press, Taylor-Francis.

LeSage, J.P., Chih, Y.Y., Vance, C., 2019. Markov chain Monte carlo estimation of spatial dynamic panel models for large samples. Computational Statistics & Data Analysis 138, 107–125.

Li, J., Sharma, A., Johnson, F., Evans, J., 2015. Evaluating the effect of climate change on areal reduction factors using regional climate model projections. Journal of Hydrology 528, 419–434.

Liu, L.M., Tiao, G.C., 1980. Random coefficient first-order autoregressive models. Journal of Econometrics 13, 305–325.

McLachlan, G., Lee, S., 2013. EMMIXuskew: An R package for fitting mixtures of multivariate skew t distributions via the EM algorithm. Journal of Statistical Software 55, 1–22.

McLachlan, G., Peel, D., 2004. Finite Mixture Models. John Wiley & Sons.

Mengersen, K., Robert, C., Titterington, M., 2011. Mixtures: Estimation and Applications. Wiley Series in Probability and Statistics, John Wiley & Sons.

Mundlak, Y., 1978. On the pooling of time series and cross-section data. Econometrica 46, 69–85.

Parent, O., LeSage, J.P., 2010. A spatial dynamic panel model with random effects applied to commuting times. Transportation Research Part B: Methodological 44, 633–645.

Parent, O., LeSage, J.P., 2011. A space-time filter for panel data models containing random effects. Computational Statistics & Data Analysis 55, 475–490.

Peel, D., McLachlan, G., 2000. Robust mixture modelling using the t distribution. Statistics and Computing 10, 339–348.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.

Phillips, P.C.B., 1991. To criticize the critics: An objective Bayesian analysis of stochastic trends. Journal of Applied Econometrics 6, 333–364.

Piani, C., Haerter, J., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theoretical and Applied Climatology 99, 187–192.

Robert, C.P., 2007. The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation. 2nd ed., Springer, New York, USA.

Schotman, P.C., Van Dijk, H.K., 1991. On Bayesian routes to unit roots. Journal of Applied Econometrics 6, 387–401.

Simpson, D., Rue, H., Riebler, A., Martins, T., Sørbye, S., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science 32, 1–28.

Sims, C.A., Uhlig, H., 1991. Understanding unit rooters: A helicopter tour. Econometrica , 1591–1599.

Walker, G., Saw, J., 1978. The distribution of linear combinations of t-variables. Journal of the American Statistical Association 73, 876–878.

Waller, L.A., Gotway, C.A., 2004. Applied Spatial Statistics for Public Health Data. John Wiley & Sons.

White, I.R., Royston, P., Wood, A.M., 2011. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine 30, 377–399.

Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic change 62, 189–216.

Yang, C.F., 2021. Common factors and spatial dependence: an application to US house prices. Econometric Reviews 40, 14–50.

Yoon, J., Zame, W.R., van der Schaar, M., 2017. Multi-directional recurrent neural networks: A novel method for estimating missing data, in: Precup, D., Teh, Y.W. (Eds.), Proceedings of Machine Learning Research, Proceedings of Machine Learning Research (PMLR). pp. 3958–3966.