CIRANO
Knowledge into action

ROBUST DYNAMIC SPACE-TIME PANEL DATA MODELS USING E-CONTAMINATION: AN APPLICATION TO CROP YIELDS AND CLIMATE CHANGE

BADI **H. BALTAGI**
GEORGES **BRESSON**
ANOOP **CHATURVEDI**
GUY **LACROIX**

2023s-01
WORKING PAPER
CS

CIRANO
Knowledge into action

Center for Interuniversity Research and Analysis on Organizations

The purpose of the **Working Papers** is to disseminate the results of research conducted by CIRANO research members in order to solicit exchanges and comments. These reports are written in the style of scientific publications. The ideas and opinions expressed in these documents are solely those of the authors.

Les cahiers de la série scientifique visent à rendre accessibles les résultats des recherches effectuées par des chercheurs membres du CIRANO afin de susciter échanges et commentaires. Ces cahiers sont rédigés dans le style des publications scientifiques et n'engagent que leurs auteurs.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

ISSN 2292-0838 (online version)

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*[‡] *and Guy Lacroix*[§]

Abstract/Résumé

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 `a 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.

Keywords/Mots-clés: climate change, crop yields, dynamic model, ε-contamination, panel data, robust Bayesian estimator, space-time / changement climatique, rendement des cultures, modèle dynamique, ε-contamination, données de panel, estimateur bayésien robuste, modèle spatio-temporel

JEL Codes/Codes JEL: C11, C23, C26, Q15, Q54

Pour citer ce document / To quote this document

H. Baltagi B., Bresson G., Chaturvedi A. and Lacroix Guy (2023). Robust dynamic space-time panel data models using ε-contamination: An application to crop yields and climate change (2023s-01, CIRANO). https://doi.org/10.54932/UFYN4045

[*] Department of Economics and Center for Policy Research, Syracuse University, Syracuse, New York, U.S.A.
[†] Department of Economics, Université Paris II, France
[‡] Department of Statistics, University of Allahabad, India
[§] Départment d'économique, Université Laval, Québec, Canada

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

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

[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:

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

[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					

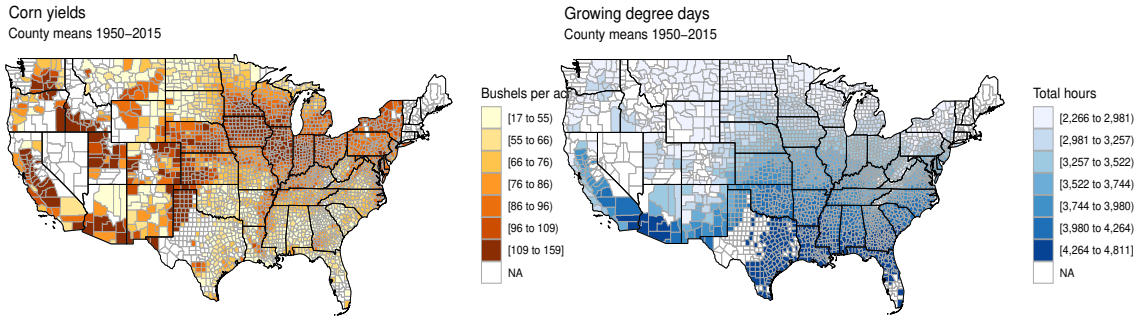
B2S2S: Bayesian two-stage two-step estimation.
se_mixt: standard errors of the parameters θ computed with mixture of
t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.
$f_t' = \left(\overline{gdd}, \overline{kdd}, \overline{prec}\right)$.
Stationarity conditions for B2S2S :
$\phi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

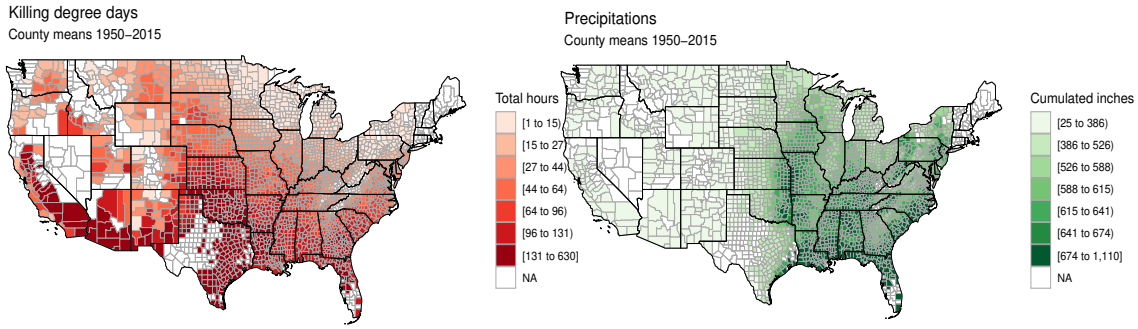
Aw	Tropical wet and dry climate	BSh	Warm semi-arid climate
BSk	Cold semi-arid climate	BWh	Warm desert climate
BWk	Cold desert climate	Cfa	Warm oceanic climate/Humid subtropical climate
Cfb	Temperate oceanic climate	Csa	Warm Mediterranean climate
Csb	Temperate Mediterranean climate	Dfa	Warm/Humid continental 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

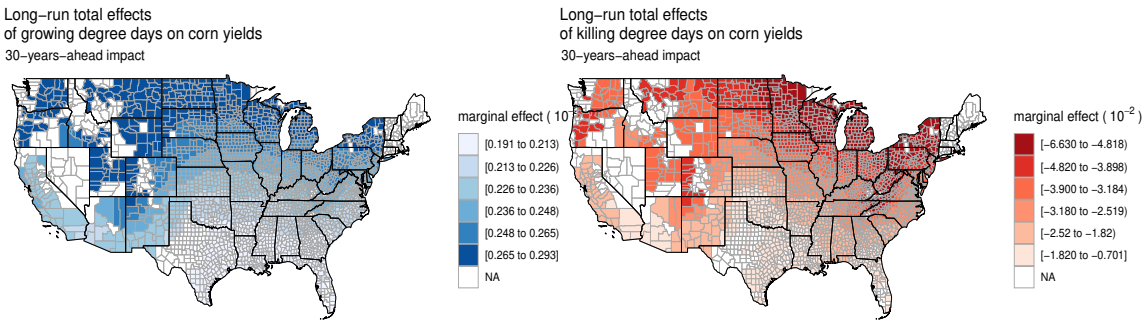


(a) (b)

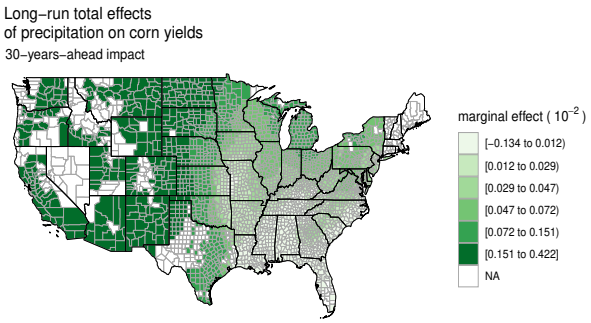


(c) (d)

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.