Unconditional transformed likelihood estimation of time-space dynamic panel data models*

Sebastian Kripfganz†

November 15, 2020

Abstract

Panel data sets allow to account for unobserved unit-specific heterogeneity, as well as time-series and cross-sectional dependence. I derive the unconditional transformed likelihood function and its derivatives for a fixed-effects panel data model with time lags, spatial lags, and spatial time lags that encompasses the pure time dynamic and pure space dynamic models as special cases. In addition, the model can accommodate spatial dependence in the error term. Consistent estimation in short panels requires proper allowance for the influence of the initial observations. I demonstrate that the model-consistent representation of the initial-period distribution involves higher-order spatial lag polynomials. Their order is linked to the minimal polynomial of the spatial weights matrix and, in general, tends to infinity with increasing sample size. An appropriate truncation of these lag polynomials becomes necessary unless the spatial weights matrix has a regular structure. The finite-sample evidence from Monte Carlo simulations shows that the proposed estimator performs well in comparison to a bias-corrected conditional likelihood estimator if parameter proliferation is kept under control. As an application, I use data from the Panel Study of Income Dynamics to estimate a time-space dynamic wage equation that I derive from a bargaining model. I find significant spillover effects among household members that give rise to a positive cohabitation premium. Furthermore, the theoretical bargaining model justifies a particular nonlinear restriction on the spatial time lag that simplifies the analytical derivations considerably and is also empirically supported.

Keywords: Dynamic panel data; Spatial autoregression; Fixed effects; Quasi-maximum likelihood estimation; Intra-household wage spillovers; Cohabitation premium

JEL Classification: C13; C23; J31

*I thank Pooyan Amir Ahmadi, Michael Binder, Sulkhan Chavleishvili, Horst Entorf, Nicola Fuchs-Schündeln, Uwe Hassler, Melanie Krause, Soroosh Soofi Siavash, Rish Singhania, Nathanael Vellekoop, and Jeffrey Wooldridge for helpful comments. I also appreciate valuable suggestions by James LeSage and other participants of several workshops and conferences, as well as seminar participants at Goethe University Frankfurt, University of St. Gallen, Michigan State University, University of Chicago, ENSAI, University of Groningen, University of Exeter, London School of Economics and Political Science, University of Birmingham, University of Hamburg, and University of Cologne. Conference travel support from Vereinigung von Freunden und Förderern der Goethe-Universität and Goethe Money and Macro Association, as well as a stipend from Verein für Socialpolitik / Deutsche Bundesbank is gratefully acknowledged.

†University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter, EX4 4PU, UK. Tel.: +44-1392-722110; E-mail: S.Kripfganz@exeter.ac.uk

1 Introduction

The availability of panel data sets that consist of observations for several economic units over multiple consecutive time periods allows to analyze dynamic adjustment processes over time and spillover effects across units, and to control for various forms of heterogeneity. Modeling cross-sectional dependence has a long-standing tradition in urban and regional economics where cross-sectional units have a fixed location in space. Neighboring cities, districts, regions, or countries are interconnected on various grounds due to migration, trade, technological spillovers, financial links, economic specialization, and fiscal competition. Spatial statistics and econometrics provide the methodological tools for the analysis of such regional dependencies.[1] In recent years, these methods became popular for the analysis of a wide range of economic applications far beyond regional science, including the analysis of financial and social networks. As an example, Bramoullé et al. (2009) discuss the identification of peer effects with spatial econometric methods.

Empirical studies that account both for cross-sectional spillover effects and time-series persistence are still relatively rare, and most of them are located in the regional economics literature. Examples include the estimation of economic growth models in the context of regional convergence (Parent and LeSage, 2012; Yu and Lee, 2012; Ho et al., 2013; Evans and Kim, 2014; Fischer and LeSage, 2015), the analysis of trade patterns (Keller and Shiue, 2007), testing for spatial cointegration in financial liberalization of neighboring countries (Elhorst et al., 2013), and an analysis of the evolution of state-level commuting times (Parent and LeSage, 2010). Despite their dominance, potential applications are not restricted to macroeconomics. In a recent microeconomic study, Verhelst and Van den Poel (2014) assess internal and external habit formation in consumption. In the current paper, I estimate a dynamic wage equation that is motivated by the observation that wages tend to be persistent over time and correlated among cohabiting workers.

Limitations to the analysis of dynamic panel data models emerge in small samples with only a few observations over time. The current paper focuses on panel data models with a short time dimension and a sufficiently large number of units. Following the terminology of Anselin (2001), I consider a time-space dynamic panel data model that allows for pure time lags, contemporaneous spatial lags, and spatial time lags of both the dependent variable and exogenous regressors.

[1] See Anselin (1988), Anselin and Bera (1998), and LeSage and Pace (2009) for a comprehensive overview.

Elhorst (2001) labels this model a general serial and spatial autoregressive distributed lag model and discusses the variety of special cases that result under different parameter restrictions. The unconditional transformed likelihood function derived in this paper encompasses the pure time dynamic and pure space dynamic models considered by Hsiao et al. (2002) and Lee and Yu (2010b), respectively, as special cases. The paper therefore builds both on the literature on short-T dynamic panel data models as well as the literature on spatial autoregressive models.

Linear panel data models with cross-sectional independence and autoregressive components in the time dimension have been studied extensively. The usual way to account for unobserved heterogeneity in these models is to incorporate unit-specific and time-specific effects. When the time dimension is short, ordinary least squares and conditional maximum likelihood estimators fail to deliver consistent estimates in dynamic models with fixed effects as a consequence of the incidental-parameters problem discussed by Neyman and Scott (1948). Nickell (1981) characterizes the resulting bias. Treating the unit-specific effects as random does not solve the problem due to the nonnegligible impact of the initial observations. Over the past decades, several solutions have been proposed. They include bias-corrected estimators (Kiviet, 1995; Hahn and Kuersteiner, 2002; Bun and Carree, 2005; Dhaene and Jochmans, 2016), instrumental variable estimators (Anderson and Hsiao, 1981), generalized method of moments (GMM) estimators (Arellano and Bond, 1991; Ahn and Schmidt, 1995; Arellano and Bover, 1995; Blundell and Bond, 1998), and unconditional likelihood-based estimators (Bhargava and Sargan, 1983; Hsiao et al., 2002; Binder et al., 2005; Moral-Benito, 2013), to name only a few. In this paper, I extend the transformed likelihood approach of Hsiao et al. (2002) to general time-space dynamic models, carefully modeling the distribution of the initial observations.

The assumption of cross-sectional independence can be relaxed in various ways.[2] An extensive part of the literature puts the attention on properly accounting for cross-sectional error dependence.[3] I focus on models with cross-sectional dependence in the form of spatial lags of the dependent variable and the exogenous regressors while still allowing for spatial dependence in the error term. Lee (2002, 2004) provides a rigorous asymptotic framework for the analysis of such

[2]Anselin (1988, 2001) and Anselin and Bera (1998) summarize the early developments of spatial econometrics, while Lee and Yu (2010c) report recent developments in spatial panel data models.

[3]Sarafidis and Wansbeek (2012) give a topical overview of the literature on weak and strong error cross-sectional dependence with reference to spatial and factor models. Kapoor et al. (2007) and Baltagi et al. (2013) consider panel data models with different forms of spatially correlated error components.

mixed-regressive spatial autoregressive models, and Lee and Yu (2010b) derive asymptotic properties of quasi-maximum likelihood estimators for spatial autoregressive panel data models with fixed effects.[4] In recent years, increasing emphasis is put on jointly modeling the dynamics in the time and the spatial dimension. Lee and Yu (2014) propose an efficient GMM estimator, while Yu et al. (2008) and Lee and Yu (2010a) investigate the asymptotic properties of conditional quasi-maximum likelihood estimators under asymptotics where both the time and the cross-sectional dimension become large. They propose a bias reduction procedure to correct for the incidental-parameters bias. Elhorst (2005, 2010) and Parent and LeSage (2011, 2012) derive unconditional likelihood functions for panel data models that are dynamic both in time and space.

The present paper is closely related to the work of Elhorst (2010) and Parent and LeSage (2012). They obtain an unconditional likelihood function for panel data models with a time lag and a contemporaneous spatial lag of the dependent variable. The latter authors additionally allow for a restricted or unrestricted spatial time lag. When the time dimension is fixed, appropriate conditions need to be imposed on the initial observations to obtain consistent estimates. Both Elhorst (2010) and Parent and LeSage (2012) approximate the marginal distribution of the initial observations by observed values of the exogenous regressors following Bhargava and Sargan (1983) and Hsiao et al. (2002) without fully accounting for the complex spatial structure. Accordingly, Elhorst (2010) finds considerable bias in the estimate of the spatial lag coefficient. In this paper, I derive the model-consistent distribution of the initial observations. It involves higher-order spatial lag polynomials that potentially require an appropriate truncation for consistent estimation. Simplifications may occur if the spatial weights matrix has a regular structure, or under a suitable restriction on the coefficient of the spatial time lag.

The next section motivates the formulation of a time-space dynamic wage equation based on a bargaining model with intra-household spillover effects. The general time-space dynamic panel data model is described in Section 3 together with the derivation of the model-consistent representation of the initial observations. The unconditional transformed likelihood function is formulated in Section 4 for the unrestricted model, while Section 5 discusses several restricted model specifications. Monte Carlo simulation results are provided in Section 6, and the empirical

[4]Kelejian and Prucha (1998, 1999), Lee (2003, 2007), and Baltagi et al. (2014), among others, propose instrumental variable and GMM estimators as computationally simple alternatives.

results are presented in Section 7. Finally, Section 8 concludes.

2 Time-space dynamic wage equation

2.1 Intra-household wage spillovers

In this section, I motivate the formulation of a time-space dynamic wage equation. The term *space* might be misleading here as it does not refer to the geographic location of the units but to different individuals within the same household. More generally, *space* is used in this paper as a synonym for the cross-sectional dimension of the panel. In the original meaning, it might refer to geographical units such as districts, states, or countries. However, the methodology applies equally well to situations where the cross-sectional units are firms, households, or individuals, as well as any kind of financial asset or other unit of observation with an ascertainable dependence structure. The spatial lag then captures spillover effects in industrial, social, or financial networks instead of regional spillovers.[5]

Andini (2007) and Semykina and Wooldridge (2013) argue in favor of a time dynamic earnings equation that income is correlated across subsequent years. In addition, I argue that intra-household spillover effects give rise to an earnings equation that is dynamic also across space. In the literature on the marriage premium, it is often argued that a significant marital wage premium for men results from economies of scale and specialization within a household (Stratton, 2002) or the positive impact of the wife's presence on the productivity of the husband (or vice versa). Alternatively, marriage might coincide with changes in behavior due to preference changes that also affect labor market outcomes (Choi et al., 2008). Similar arguments can be made to rationalize a cohabitation premium which is likely to be smaller than the marriage premium because marriage is a more stable relationship than cohabitation (Stratton, 2002). However, the existence of these premiums does not require that both partners are working. Conversely, the effects due to specialization might be stronger when either of the partners devotes more time to household work.

A positive link between the earnings of housemates can be justified on different grounds. First, the literature on assortative mating argues that individuals choose spouses with similar charac-

[5]For example, Bramoullé et al. (2009) adopt a spatial econometric modeling approach for the identification of peer effects in social networks.

teristics. Consequently, this could on average also lead to similar wage profiles.[6] Second, labor market outcomes are potentially correlated among partners as a result of collective household preferences. Third, a wage increase of the partner potentially carries the information content that the macroeconomic circumstances might be favorable to ask the own employer for a pay rise. Fourth, social norms and peer pressure not to fall behind the spouse's wage go into the same direction. Fifth, the income of the cohabitant can be regarded as an outside option in wage negotiations if the household income is pooled or if there are intra-household wage transfers to compensate for income differentials. The better the outside option the higher is the expected wage outcome in an employer-employee bargaining problem.

2.2 Employer-employee wage bargaining model

While many different reasons can be adduced to rationalize a positive relationship between partner's wages, in the following I focus exemplarily on the last argument and derive a time-space dynamic wage equation from a theoretical bargaining model. The starting point is the human capital model of Mincer (1974) that relates net potential earnings, p_{it}, of a worker i at a point in time t to his productivity determinants $\mathbf{w}_{1it}$ that usually include labor market experience and years of schooling:

$$\ln p_{it} = \mathbf{w}_{1it}' \boldsymbol{\psi}_1 + \xi_i. \tag{1}$$

ξ_i denotes other individual-specific characteristics such as ability that are unobserved to the econometrician.

In perfectly competitive labor markets, equilibrium wages y_{it}^* then equal net potential earnings. Andini (2013) derives a time dynamic wage equation from a bargaining model where the outside option is affected by unemployment benefits that depend on past income. Andini (2013) instead considers a framework of imperfect competition by embedding the above human capital model into a simple wage-bargaining model between the worker and an employer. In his model, unemployment benefits are available as an outside option for the worker. Because unemployment benefits are a function of past income, this eventually gives rise to a time-dynamic wage equation. I extend his work to motivate a time-space dynamic wage equation by adding intra-household transfers as a

[6] Henz and Sundström (2001) find empirical support in favor of this hypothesis for Swedish couples.

second outside option.

At a given time t, the worker maximizes his earnings utility, $U_{it}^w = \ln y_{it}$, subject to his outside option $\bar{U}_{it}^w = \phi \ln b_{it} + \delta \ln t_{it}$, where $b_{it} = y_{i,t-1} e^{-\mathbf{w}_{2it}' \boldsymbol{\psi}_2}$ denotes unemployment benefits that depend on the worker's wage history,[7] and $t_{it} = p_{jt} e^{-\mathbf{w}_{3it}' \boldsymbol{\psi}_3}$ are (expected) intra-household transfers from worker i's partner j. More precisely, unemployment benefits are modeled as a fraction $e^{-\mathbf{w}_{2it}' \boldsymbol{\psi}_2} \in (0, 1]$ of the wage in the last period, where $\mathbf{w}_{2it}$ can be worker-specific or institutional variables that determine this fraction. Regarding the intra-household transfers, when both partners i and j bargain with their employers at the same time they cannot condition on the wage realization of the other. However, they can form naive expectations based on the observed net potential earnings p_{jt}. Similar to the unemployment benefits, additional individual-specific or household-specific variables $\mathbf{w}_{3it}$ can affect the fraction of the partner's expected income that is shared within the household, $e^{-\mathbf{w}_{3it}' \boldsymbol{\psi}_3} \in (0, 1]$.

The parameters ϕ and δ are semi-elasticities that measure the relative utility gains from an increase in the respective outside option to a corresponding increase in the labor income. If receiving unemployment benefits is associated with a stigma, we would expect ϕ to be smaller than unity.[8] Similarly, it can be discomfortable to rely on the partner's income. The social pressure to accept a paid job is potentially larger in the latter case than with unemployment benefits such that $\delta < \phi$ can be expected.

On the labor demand side, the employer maximizes his utility from offering a wage contract to the worker, $U_{it}^e = \ln(p_{it}/y_{it})$. His outside option is characterized by the revenue per dollar paid, $r_{it} = e^{\mathbf{w}_{4it}' \boldsymbol{\psi}_4} \geq 1$, from hiring an alternative worker, $\bar{U}_{it}^e = \ln r_{it}$. The variables $\mathbf{w}_{4it}$ can be macroeconomic factors related to the business cycle that affect the labor market.

We can then obtain the equilibrium wage y_{it}^* by maximizing the Nash bargaining function with respect to the wage y_{it}:

$$U_{it} = (U_{it}^w - \bar{U}_{it}^w)^\theta (U_{it}^e - \bar{U}_{it}^e)^{1-\theta}, \tag{2}$$

[7]For simplicity, I model the unemployment benefits as a function of the previous period's wage only. Andini (2013) allows them to be a function of additional time lags.

[8]An argument could also be made for $\phi > 1$ if the preference for leisure is high relative to spending the same amount of time at work.

where $\theta \in (0,1)$ is the bargaining weight of the worker. The first-order condition yields

$$\ln y^*_{it} = (1-\theta)(\phi \ln b_{it} + \delta \ln t_{it}) + \theta(\ln p_{it} - \ln r_{it}). \tag{3}$$

By symmetry of the wage-bargaining problem for i's partner j, we can rearrange the corresponding first-order condition:

$$\ln p_{jt} = \frac{1}{\theta} \ln y^*_{jt} - \frac{1-\theta}{\theta} (\phi \ln b_{jt} + \delta \ln t_{jt}) + \ln r_{jt}, \tag{4}$$

which we can substitute in equation (3) by using the fact that t_{it} is a function of p_{jt}. We then obtain

$$\begin{aligned}
\ln y^*_{it} = {} & (1-\theta)\phi \ln y_{i,t-1} + \frac{1-\theta}{\theta}\delta \ln y^*_{jt} - \frac{(1-\theta)^2}{\theta}\phi\delta \ln y_{j,t-1} \\
& + \left[\theta - \frac{(1-\theta)^2}{\theta}\delta^2\right] (\mathbf{w}'_{1it}\boldsymbol{\psi}_1 + \xi_i) - (1-\theta)\phi\mathbf{w}'_{2it}\boldsymbol{\psi}_2 - (1-\theta)\delta\mathbf{w}'_{3it}\boldsymbol{\psi}_3 - \theta\mathbf{w}'_{4it}\boldsymbol{\psi}_4 \\
& + \frac{(1-\theta)^2}{\theta}\phi\delta\mathbf{w}'_{2jt}\boldsymbol{\psi}_2 + \frac{(1-\theta)^2}{\theta}\delta^2\mathbf{w}'_{3jt}\boldsymbol{\psi}_3 + (1-\theta)\delta\mathbf{w}'_{4jt}\boldsymbol{\psi}_4.
\end{aligned} \tag{5}$$

Finally, adding a stochastic error term leads us to the following time-space dynamic wage equation that we can bring to the data:

$$\ln y_{it} = \lambda \ln y_{i,t-1} + \rho_0 \ln y_{jt} + \rho_1 \ln y_{j,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}'_{jt}\boldsymbol{\gamma} + \alpha_i + u_{it}, \tag{6}$$

where $\mathbf{x}_{it} = \bigcup_{k=1}^{4} \mathbf{w}_{kit}$. The coefficients in the econometric model (6) are linked to the model parameters in the equilibrium relationship (5), in particular $\lambda = (1-\theta)\phi$, $\rho_0 = (1-\theta)\delta/\theta$, $\rho_1 = -(1-\theta)^2\phi\delta/\theta$. Notice that this implies the restriction $\rho_1 = -\lambda\rho_0$. Similar implications of the bargaining model could be exploited to restrict the coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

2.3 Restriction on the spatial time lag

The restriction $\rho_1 = -\lambda\rho_0$ on the coefficients in equation (6) results from the symmetry of the bargaining problem for the two partners. Because the model predicts that λ and ρ_0 are both positive, the coefficient ρ_1 of the spatial time lag should have a negative sign. The intuition can be found in the rearranged first-order condition (4). In the equilibrium, past wages are negatively related to net potential earnings. Consequently, an increase in past wages can be seen as a proxy for lower net potential earnings today which in turn reduce the outside option of the partner and therefore his equilibrium wage outcome.

This nonlinear restriction deserves special attention beyond this particular model. Tao and Yu (2012) derive the same restriction from a model of intertemporal consumption and investment decisions with external habits, as well as a model of governmental policy adjustments with policy inertia and spillovers from neighboring jurisdictions. Yu and Lee (2012) and Evans and Kim (2014) introduce technological spillovers into the neoclassical growth model and obtain an estimation equation with the same nonlinear restriction for which they also find empirical support.[9] As shown by Parent and LeSage (2010, 2011, 2012) and further down in the current paper, it simplifies the computational complexity of likelihood-based estimation procedures considerably. It also disentangles the time effect, λ, and the spatial effect, ρ_0, which enhances the interpretation of marginal effects and the computation of dynamic responses.

Equation (6) is a formulation of a time-space dynamic panel data model in the terminology of Anselin (2001) that features a pure time lag of the dependent variable, $y_{i,t-1}$, a contemporaneous spatial lag, y_{jt}, and a spatial time lag, $y_{j,t-1}$. In the following sections, I discuss unconditional quasi-maximum likelihood estimation of general time-space dynamic panel data models.

3 Time-space dynamic panel data model

3.1 Unrestricted model

Consider the following time-space dynamic autoregressive distributed lag model for the cross-sectional units $i = 1, 2, \ldots, N$, and a fixed number of time periods $t = 1, 2, \ldots, T$:

$$y_{it} = \lambda y_{i,t-1} + \beta_0 x_{it} + \beta_1 x_{i,t-1} + \sum_{j=1}^{N} w_{ij}(\rho_0 y_{jt} + \rho_1 y_{j,t-1} + \gamma_0 x_{jt} + \gamma_1 x_{j,t-1}) + e_{it}, \tag{7}$$

[9] In the context of regional convergence, the empirical results of Parent and LeSage (2012), Ho et al. (2013), and Fischer and LeSage (2015) provide further evidence for this particular restriction, as do the estimates of Keller and Shiue (2007) in their analysis of trade patterns.

where x_{it} is a strictly exogenous regressor,[10] and the spatial weights w_{ij} describe the structural dependencies between the cross-sectional units. The initial observations y_{i0} and x_{i0} are assumed to be observed, and the error term e_{it} has a spatial error components structure following Kapoor et al. (2007):[11]

$$e_{it} = \sum_{j=1}^{N} w_{ij} \rho_2 e_{jt} + \alpha_i + u_{it}, \tag{8}$$

where α_i is a unit-specific intercept that is allowed to be freely correlated with x_{it}, and u_{it} is an independently and identically distributed error term. We can write model (7) and (8) in more compact form as

$$\mathbf{S}_N \mathbf{y}_t = \mathbf{A}_N \mathbf{y}_{t-1} + \mathbf{B}_{0N} \mathbf{x}_t + \mathbf{B}_{1N} \mathbf{x}_{t-1} + \mathbf{e}_t, \tag{9}$$

$$\mathbf{R}_N \mathbf{e}_t = \boldsymbol{\alpha} + \mathbf{u}_t, \tag{10}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $\mathbf{y}_{t-1} = (y_{1,t-1}, y_{2,t-1}, \ldots, y_{N,t-1})'$, and $\mathbf{x}_t$, $\mathbf{x}_{t-1}$, $\mathbf{e}_t$, $\boldsymbol{\alpha}$, and $\mathbf{u}_t$ are stacked accordingly.[12] The coefficient matrices are $\mathbf{S}_N = (\mathbf{I}_N - \rho_0 \mathbf{W}_N)$, $\mathbf{A}_N = (\lambda \mathbf{I}_N + \rho_1 \mathbf{W}_N)$, $\mathbf{B}_{lN} = (\beta_l \mathbf{I}_N + \gamma_l \mathbf{W}_N), l = 0, 1$, and $\mathbf{R}_N = (\mathbf{I}_N - \rho_2 \mathbf{W}_N)$, where $\mathbf{W}_N$ is the $N \times N$ spatial weights matrix with w_{ij} as the (i, j)-th element, and $\mathbf{I}_N$ is the identity matrix of dimension N.

The following conditions are assumed to hold:

Assumption 1: The regressor x_{it} is strictly exogenous with respect to the error term u_{it} such that $E[\mathbf{u}_t | \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T] = \mathbf{0}$.

Assumption 2: The sequence of error terms $\{u_{it}\}$, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, are i.i.d. with $E[\mathbf{u}_t \mathbf{u}_t' | \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T] = \sigma_u^2 \mathbf{I}_N$ and $E[\mathbf{u}_t \mathbf{u}_s' | \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T] = \mathbf{0}$ for all $t \neq s$.[13] The fourth moment $E[u_{it}^4 | \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T]$ exists.

[10] To avoid notational complications, I restrict the exposition to a single exogenous regressor x_{it}.

[11] Baltagi et al. (2013) allow the error components α_i and u_{it} to follow separate spatial autoregressive processes, while Lee and Yu (2010b) model the spatial error dependence in u_{it} only. For the current paper, this distinction is not of relevance because the fixed effects α_i will drop out by applying a suitable model transformation. Lee and Yu (2010b) also allow the spatial weights to be different in equations (7) and (8). I abstract from this additional complication.

[12] For convenience, in defining these vectors I suppress the subscript N that denotes the dependence on the sample size.

[13] Allowing for cross-sectional heteroscedasticity is possible by adopting the approach of Hayakawa and Pesaran (2015).

3.2 Spatial weights matrix

The spatial weights matrix captures the cross-sectional dependency structure. In the regional science literature, these structural dependencies typically reflect the relative location of the regional units. $\mathbf{W}_N$ can be an inverse distance matrix with spatial weights w_{ij} that measure the inverse of the (geographical) distance between two regions i and j. In a broader sense, the distance between two units can also be defined on economic grounds by considering, for example, the bilateral trade intensity or the degree of financial connectedness. Alternatively, $\mathbf{W}_N$ may be constructed as a contiguity matrix where nonzero weights indicate a common border among two units. Similar to Lee (2004) and Lee and Yu (2010b), I make the following regularity assumptions:

Assumption 3: The spatial weights w_{ij} are constant over time and at most of order h_N^{-1} uniformly in all i, j, with $w_{ii} = 0$. The sequence $\{h_N\}$ is bounded or divergent, and the ratio $h_N/N \to 0$ as $N \to \infty$.

Assumption 4: The matrices $\mathbf{S}_N$ and $\mathbf{R}_N$ are invertible for all $\rho_0, \rho_2 \in (-1/|\omega_{\min}|, 1/\omega_{\max})$, where $\omega_{\min}$ and $\omega_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{W}_N$, respectively.

Assumption 5: The matrices $\mathbf{W}_N$, $\mathbf{S}_N^{-1}$, and $\mathbf{R}_N^{-1}$ are uniformly bounded both in row and column sum.

Assumptions 3 to 5 limit the cross-sectional dependence to a manageable degree.[14] To simplify the exposition, I further impose the following normalization:

Assumption 6: The spatial weights matrix $\mathbf{W}_N$ is normalized to have a spectral radius of unity, that is $||\mathbf{W}_N|| = 1$ for the spectral matrix norm $|| \cdot ||$.

Assumption 6 is neither necessary nor restrictive. A spectral radius of unity can be achieved by dividing the spatial weights matrix by its largest eigenvalue in absolute value. The spatial coefficients ρ_0, ρ_1, ρ_2, γ_0, and γ_1 are rescaled accordingly. Alternatively, the normalization of $\mathbf{W}_N$ can be achieved by row standardization, that is by dividing all elements of the initial weights matrix

[14] See Lee (2004) for a detailed discussion.

by the respective row sum such that $\sum_{j=1}^{N} w_{ij} = 1$. Notice that if the weights matrix was symmetric before standardization the former approach preserves symmetry while row standardization may not. In both cases, $\omega_{\max} = 1$ after the normalization.

For the time-space dynamic wage equation (6) with intra-household spillovers derived in Section 2 and a household size of $M = 2$ workers, the weights are easily constructed as $w_{ij} = 1$ if i and j, with $i \neq j$, are members of the same household, and $w_{ij} = 0$ otherwise. The resulting spatial weights matrix is a special case of a matrix used by Lee (2004) for illustrative purposes. Inspired by Case (1991), he considers a sample with R regions (or households) that each contain M members such that $N = RM$. The members within a region are spatially dependent with equal weights but there is no dependence across regions. The spatial weights matrix thus has the symmetric and block-diagonal form $\mathbf{W}_N = \mathbf{I}_R \otimes \mathbf{B}_M$ with $\mathbf{B}_M = (\boldsymbol{\iota}_M \boldsymbol{\iota}_M' - \mathbf{I}_M)/(M-1)$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{\iota}_M$ is an M-dimensional column vector of ones. $\mathbf{W}_N$ is diagonalizable and has only two distinct eigenvalues, unity with multiplicity R and $(1-M)^{-1}$ with multiplicity $R(M-1)$. In this case, $h_N/N = (M-1)/(RM)$ such that Assumption 3 rules out asymptotics where R is fixed and $M \to \infty$.

As discussed by Kripfganz (2015), the Cayley-Hamilton theorem implies that every power $K > Q_N$ of an N-dimensional square matrix can be expressed by a polynomial of order Q_N when its minimal polynomial is of order $Q_N + 1$, where $Q_N < N$ may or may not depend on the dimension N. For diagonalizable matrices, Q_N equals the number of distinct eigenvalues less one. Therefore, $Q_N = 1$ for this particular spatial weights matrix and we can express all powers of $\mathbf{W}_N$ as first-order polynomials, namely $\mathbf{W}_N^k = \zeta_{k0}\mathbf{I}_N + \zeta_{k1}\mathbf{W}_N$ for all $k \geq 0$, where $\zeta_{kl} = [1 - (1-M)^{1-k-l}](M-1)^l/M$ for $l = 0, 1$. This result also simplifies the computation of inverse matrices that are polynomials in $\mathbf{W}_N$:

$$(a_0\mathbf{I}_N - a_1\mathbf{W}_N)^{-1} = \frac{1}{a_0}\sum_{k=0}^{\infty}\left(\frac{a_1}{a_0}\mathbf{W}_N\right)^k = b_0\mathbf{I}_N + b_1\mathbf{W}_N, \tag{11}$$

with

$$b_0 = \frac{1}{a_0 - a_1} - b_1, \quad b_1 = \frac{a_1(M-1)}{(a_0 - a_1)[a_0(M-1) + a_1]},$$

for some scalar constants a_0 and a_1 such that the inverse has a convergent series representation.[15]

[15]See Kripfganz (2015) for a generalization of these results and additional examples.

With these insights, we can already infer the following statement on the identification of the model parameters:

Proposition 1: A necessary condition for identification of all parameters in the unrestricted model (9) and (10) is $Q_N > 1$.

Proof. Consider equation (9) for a generic period t premultiplied by $\mathbf{R}_N$:

$$\mathbf{R}_N \mathbf{S}_N \mathbf{y}_t = \mathbf{R}_N \mathbf{A}_N \mathbf{y}_{t-1} + \mathbf{R}_N \mathbf{B}_{0N} \mathbf{x}_t + \mathbf{R}_N \mathbf{B}_{1N} \mathbf{x}_{t-1} + \boldsymbol{\alpha} + \mathbf{u}_t. \tag{12}$$

Now suppose $Q_N = 1$ such that $\mathbf{W}_N^k = \zeta_{k0}\mathbf{I}_N + \zeta_{k1}\mathbf{W}_N$ for all $k \geq 0$. Thus,

$$\begin{aligned}
\mathbf{R}_N \mathbf{S}_N &= (1 + \zeta_{20}\rho_0\rho_2)\mathbf{I}_N + (\zeta_{21}\rho_0\rho_2 - \rho_0 - \rho_2)\mathbf{W}_N = \sigma^*\mathbf{I}_N - \rho_0^*\mathbf{W}_N, \\
\mathbf{R}_N \mathbf{A}_N &= (\lambda - \zeta_{20}\rho_1\rho_2)\mathbf{I}_N + (\rho_1 - \lambda\rho_2 - \zeta_{21}\rho_1\rho_2)\mathbf{W}_N = \lambda^*\mathbf{I}_N + \rho_1^*\mathbf{W}_N, \\
\mathbf{R}_N \mathbf{B}_{lN} &= (\beta_l - \zeta_{20}\gamma_l\rho_2)\mathbf{I}_N + (\gamma_l - \beta_l\rho_2 - \zeta_{21}\gamma_l\rho_2)\mathbf{W}_N = \beta_l^*\mathbf{I}_n + \gamma_l^*\mathbf{W}_N, \quad l = 0, 1.
\end{aligned}$$

After normalizing the estimation equation by $1/\sigma^*$, we obtain a reformulation of the initial time-space dynamic panel data model with first-order spatial lags only but without spatial error dependence. Consequently, we cannot identify all spatial lag parameters ρ_0, ρ_1, γ_0, and γ_1 jointly with ρ_2 when $Q_N = 1$. □

Therefore, a spatial weights matrix with $Q_N = 1$ requires identifying restrictions on the model parameters, for example $\rho_0 = 0$ or $\rho_2 = 0$. However, the models under these two restrictions are still observationally equivalent. Differentiating between spatial lag dependence, $\rho_0 \neq 0$ and $\rho_2 = 0$, and spatial error dependence, $\rho_0 = 0$ and $\rho_2 \neq 0$, is only possible if at least one additional coefficient of the right-hand side variables is restricted to zero. When $Q_N > 1$, identification is in general feasible but might be weak if the second-order spatial lags $\mathbf{W}_N^2\mathbf{y}_t$ and $\mathbf{W}_N^2\mathbf{x}_t$ are highly correlated with $(\mathbf{y}_t, \mathbf{W}_N\mathbf{y}_t)$ and $(\mathbf{x}_t, \mathbf{W}_N\mathbf{x}_t)$, respectively.

3.3 Model transformation and initial observations

To remove the time-invariant incidental parameters $\boldsymbol{\alpha}$ from equation (9), we can apply a linear filter Δ_t such that $\Delta_t\boldsymbol{\alpha} = \mathbf{0}$. When L denotes the time lag operator, Δ_t may inter alia be the operator that creates first differences, $\Delta_t = \Delta = 1-L$, the forward-orthogonal deviations operator, $\Delta_t = \sqrt{(T-t)/(T-t+1)}[1 - \sum_{s=1}^{T-t} L^{-s}/(T-t)]$, or the operator that creates deviations from the time means, $\Delta_t = 1 - \sum_{s=1}^{T} L^{t-s}/T$.[16] In this paper, I follow Hsiao et al. (2002) and focus on the first-difference transformation that yields

$$\mathbf{S}_N\Delta\mathbf{y}_t = \mathbf{A}_N\Delta\mathbf{y}_{t-1} + \mathbf{B}_{0N}\Delta\mathbf{x}_t + \mathbf{B}_{1N}\Delta\mathbf{x}_{t-1} + \mathbf{R}_N^{-1}\Delta\mathbf{u}_t. \tag{13}$$

The model in equation (13) is well defined for $t = 2, 3, \ldots, T$ but not for $t = 1$ because $\Delta\mathbf{y}_0$ and $\Delta\mathbf{x}_0$ are unobserved. By continuous substitution we can write

$$\begin{aligned}\mathbf{S}_N\Delta\mathbf{y}_1 = {} & (\mathbf{A}_N\mathbf{S}_N^{-1})^p\mathbf{S}_N\Delta\mathbf{y}_{1-p} \\ & + \sum_{s=0}^{p-1}(\mathbf{A}_N\mathbf{S}_N^{-1})^s(\mathbf{B}_{0N}\Delta\mathbf{x}_{1-s} + \mathbf{B}_{1N}\Delta\mathbf{x}_{-s}) + \sum_{s=0}^{p-1}(\mathbf{A}_N\mathbf{S}_N^{-1})^s\mathbf{R}_N^{-1}\Delta\mathbf{u}_{1-s},\end{aligned} \tag{14}$$

which still depends on unobservables. Let $\Delta\mathbf{x} = (\Delta\mathbf{x}_1', \Delta\mathbf{x}_2', \ldots, \Delta\mathbf{x}_T')'$. Similar to Hsiao et al. (2002), I make either of the following two assumptions:

Assumption 7.1: The process started in the infinite past, $p \to \infty$, and reached stationarity, $||\mathbf{A}_N\mathbf{S}_N^{-1}|| < 1$.

Assumption 7.2: The process was initiated at some finite period in the past with $E[\Delta\mathbf{y}_{1-p}|\Delta\mathbf{x}] = \varphi_0\varphi(\mathbf{W}_N)\boldsymbol{\iota}_N$, where $|\varphi_0| < \infty$, $\varphi(\mathbf{W}_N) = \mathbf{I}_N + \sum_{k=1}^{\infty}\varphi_k\mathbf{W}_N^k$, and there exists a real constant q_φ, $0 < q_\varphi < 1$, such that $|\varphi_k| \leq q_\varphi^k$ for all $k \geq 1$.[17]

The stationarity condition $||\mathbf{A}_N\mathbf{S}_N^{-1}|| < 1$ implies a restriction on the coefficients, namely

[16] As shown by Arellano and Bover (1995) and Bun and Kiviet (2006), the resulting least squares estimator conditional on the initial observations is invariant to the particular transformation. Hsiao et al. (2002) derive a similar result for the unconditional transformed likelihood estimator.

[17] Consequently, the sequence of coefficients $\{\varphi_k\}$ decays to zero at an exponential rate, and $||\varphi(\mathbf{W}_N)|| \leq 1 + \sum_{k=1}^{\infty}|\varphi_k| \cdot ||\mathbf{W}_N||^k < \infty$ since $\mathbf{W}_N$ is normalized to have a spectral radius of unity.

$|(\lambda+\rho_1\omega_j)(1-\rho_0\omega_j)^{-1}| < 1$ for all eigenvalues $\omega_j, j = 1, 2, \ldots, N$, of $\mathbf{W}_N$.[18] Moreover, I generalize the stationarity assumption for the regressor x_{it} made by Hsiao et al. (2002) to allow for spatial dependence:

Assumption 8: $\mathbf{x}_t$ is generated either by a trend stationary or a first-difference stationary process such that

$$\Delta\mathbf{x}_t = \varsigma(\mathbf{W}_N)\left(g\boldsymbol{\iota}_N + \sum_{s=0}^{\infty} d_s\boldsymbol{\epsilon}_{t-s}\right), \tag{15}$$

where $\varsigma(\mathbf{W}_N) = \mathbf{I}_N + \sum_{k=1}^{\infty}\varsigma_k\mathbf{W}_N^k$ is bounded away from zero, and there exists a real constant q_ς, $0 < q_\varsigma < 1$, such that $|\varsigma_k| \leq q_\varsigma^k$ for all $k \geq 1$. Also, $\sum_{s=0}^{\infty}|d_s| < \infty$, and $\boldsymbol{\epsilon}_t$ are independently and identically distributed with $E[\boldsymbol{\epsilon}_t] = \mathbf{0}$, and $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \sigma_\epsilon^2\mathbf{I}_N$.

For example, if $\varsigma_k = \varsigma^k$ for all k and $|\varsigma| < 1$ such that $\varsigma(\mathbf{W}_N) = (\mathbf{I}_N - \varsigma\mathbf{W}_N)^{-1}$ then $\Delta\mathbf{x}_t$ follows a first-order spatial autoregressive process. Under normality of $\boldsymbol{\epsilon}_t$, Assumption 8 implies[19]

$$E[\Delta\mathbf{x}_{1-s}|\Delta\mathbf{x}] = b_{s0}\varsigma(\mathbf{W}_N)\boldsymbol{\iota}_N + \sum_{l=1}^{T} b_{sl}\Delta\mathbf{x}_l, \tag{16}$$

such that we can rewrite equation (14) as

$$\mathbf{S}_N\Delta\mathbf{y}_1 = \psi(\mathbf{W}_N)\boldsymbol{\iota}_N + \sum_{l=1}^{T}\pi_l(\mathbf{W}_N)\Delta\mathbf{x}_l + \tilde{\boldsymbol{\nu}}_1, \tag{17}$$

where

$$\psi(\mathbf{W}_N) = (\mathbf{A}_N\mathbf{S}_N^{-1})^p\mathbf{S}_N\varphi_0\varphi(\mathbf{W}_N) + \pi_0(\mathbf{W}_N)\varsigma(\mathbf{W}_N), \tag{18}$$

$$\pi_l(\mathbf{W}_N) = \sum_{s=0}^{p-1}(\mathbf{A}_N\mathbf{S}_N^{-1})^s(b_{sl}\mathbf{B}_{0N} + b_{s+1,l}\mathbf{B}_{1N}), \quad l = 0, 1, \ldots, T. \tag{19}$$

Under Assumption 7.1, expression (18) reduces to $\psi(\mathbf{W}_N) = \pi_0(\mathbf{W}_N)\varsigma(\mathbf{W}_N)$. Also, $\pi_0(\mathbf{W}_N) = \mathbf{0}$

[18] Compare Elhorst (2001), Yu et al. (2008), and Parent and LeSage (2011, 2012).

[19] If $\boldsymbol{\epsilon}_t$ are not normally distributed, equation (16) can be seen as a linear projection. See Hsiao et al. (2002).

if $g = 0$ in equation (15). The initial observations error term turns out to be

$$\tilde{\boldsymbol{\nu}}_1 = (\mathbf{A}_N \mathbf{S}_N^{-1})^p \mathbf{S}_N \mathbf{q}_{y,1-p} + \sum_{s=0}^{p-1} (\mathbf{A}_N \mathbf{S}_N^{-1})^s (\mathbf{B}_{0N} \mathbf{q}_{x,1-s} + \mathbf{B}_{1N} \mathbf{q}_{x,-s} + \mathbf{R}_N^{-1} \Delta \mathbf{u}_{1-s}), \tag{20}$$

where $\mathbf{q}_{y,1-p} = \Delta \mathbf{y}_{1-p} - E[\Delta \mathbf{y}_{1-p} | \Delta \mathbf{x}]$ and $\mathbf{q}_{x,s} = \Delta \mathbf{x}_s - E[\Delta \mathbf{x}_s | \Delta \mathbf{x}]$ are projection errors with variances $\tau_{y,1-p} \varphi(\mathbf{W}_N) \varphi(\mathbf{W}_N)'$ and $\tau_{x,s} \varsigma(\mathbf{W}_N) \varsigma(\mathbf{W}_N)'$, respectively. To simplify the formulation of the likelihood function, notice that the distribution of the error term $\tilde{\boldsymbol{\nu}}_1$ is observationally equivalent to that of $\phi^{-1}(\mathbf{W}_N) \mathbf{R}_N^{-1} \boldsymbol{\nu}_1$, where

$$\begin{aligned}\phi^{-1}(\mathbf{W}_N) &= \sqrt{\frac{\tau_{y,1-p}}{\sigma_\nu^2}} (\mathbf{A}_N \mathbf{S}_N^{-1})^p \mathbf{S}_N \varphi(\mathbf{W}_N) \mathbf{R}_N \\ &+ \sum_{s=0}^{p-1} (\mathbf{A}_N \mathbf{S}_N^{-1})^s \left[\sqrt{\frac{\tau_{x,1-s}}{\sigma_\nu^2}} \mathbf{B}_{0N} \varsigma(\mathbf{W}_N) + \sqrt{\frac{\tau_{x,-s}}{\sigma_\nu^2}} \mathbf{B}_{1N} \varsigma(\mathbf{W}_N) \right] \mathbf{R}_N + \sqrt{\frac{2}{\tau}} \sum_{s=0}^{p-1} (\mathbf{A}_N \mathbf{S}_N^{-1})^s, \end{aligned} \tag{21}$$

and $\boldsymbol{\nu}_1$ is a composite error with the properties $E[\boldsymbol{\nu}_1 | \Delta \mathbf{x}] = \mathbf{0}$, $E[\boldsymbol{\nu}_1 \boldsymbol{\nu}_1'] = \sigma_\nu^2 \mathbf{I}_N$ with $\sigma_\nu^2 / \sigma_u^2 = \tau$, $E[\boldsymbol{\nu}_1 \Delta \mathbf{u}_t'] = \mathbf{0}$ for all $t = 3, 4, \ldots, T$, and $E[\boldsymbol{\nu}_1 \Delta \mathbf{u}_2'] = -\sigma_u^2 \phi(\mathbf{W}_N)$.[20] Finally notice that the functions $\psi(\mathbf{W}_N)$, $\pi_l(\mathbf{W}_N)$, and $\phi(\mathbf{W}_N)$ can be written as polynomials in $\mathbf{W}_N$ such that

$$\psi(\mathbf{W}_N) = \sum_{k=0}^{Q_N} \psi_k \mathbf{W}_N^k, \tag{22}$$

$$\pi_l(\mathbf{W}_N) = \sum_{k=0}^{Q_N} \pi_{kl} \mathbf{W}_N^k, \quad l = 1, 2, \ldots, T, \tag{23}$$

$$\phi(\mathbf{W}_N) = \mathbf{I}_N + \sum_{k=1}^{Q_N} \phi_k \mathbf{W}_N^k, \tag{24}$$

where the order Q_N equals the order of the minimal polynomial of the spatial weights matrix less one, as discussed in Section 3.2. The first coefficient in the polynomial $\phi(\mathbf{W}_N)$ is standardized to unity by appropriately scaling $\phi^{-1}(\mathbf{W}_N)$ in equation (21).

In the case of the spatial weights matrix $\mathbf{W}_N = \mathbf{I}_R \otimes \mathbf{B}_M$ from Section 3.2, the polynomials $\phi(\mathbf{W}_N)$ and $\pi_l(\mathbf{W}_N)$ reduce both to the first order irrespective of the sample size, and $\psi(\mathbf{W}_N)$ becomes a scalar multiple of the identity matrix due to row standardization. Moreover, if the

[20] The variance of $\boldsymbol{\nu}_1$ follows from a suitable scaling of $\phi^{-1}(\mathbf{W}_N)$, and the covariance term $E[\boldsymbol{\nu}_1 \Delta \mathbf{u}_2'] = -\sigma_u^2 \phi(\mathbf{W}_N)$ results from $E[\phi^{-1}(\mathbf{W}_N) \mathbf{R}_N^{-1} \boldsymbol{\nu}_1 \Delta \mathbf{u}_2'] = E[\tilde{\boldsymbol{\nu}}_1 \Delta \mathbf{u}_2'] = -\sigma_u^2 \mathbf{R}_N^{-1}$.

exogenous regressor is a region-specific variable such that $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Rt})' \otimes \boldsymbol{\iota}_M$, the order of $\pi_l(\mathbf{W}_N)$ shrinks to zero because $\mathbf{W}_N \Delta\mathbf{x}_t = \Delta\mathbf{x}_t$ in this case.[21]

Remark 1: At this stage, we shall compare the formulation (17) of the initial observations with that of Elhorst (2010). He considers a restricted version of model (9) with $\rho_1 = \beta_1 = \gamma_0 = \gamma_1 = 0$ such that $\mathbf{A}_N = \lambda\mathbf{I}_N$, $\mathbf{B}_{0N} = \beta_0\mathbf{I}_N$, and $\mathbf{B}_{1N} = \mathbf{0}$. His projection is then characterized by $\psi(\mathbf{W}_N) = \psi\mathbf{I}_N$, and $\pi_l(\mathbf{W}_N) = \pi_l\mathbf{I}_N$ for $l = 1, 2, \ldots, T$. However, even with a row-standardized spatial weights matrix a model-consistent representation demands $\pi_l(\mathbf{W}_N) = \beta\sum_{s=0}^{p-1}\lambda^s\mathbf{S}_N^{-s}b_{sl}$, $l = 0, 1, \ldots, T$, which is not free of spatial dependence irrespective of whether $p \to \infty$ or not. It also does not depend on the choice of $\varphi(\mathbf{W}_N)$ and $\varsigma(\mathbf{W}_N)$. Similar arguments apply to the projection of Parent and LeSage (2012) in the context of an unrestricted random effects model without a first-difference transformation. A set of restrictions that removes this kind of spatial dependence is discussed in Section 5.

3.4 Covariance matrix

The variance-covariance matrix of the stacked vector of errors $\Delta\mathbf{u} = (\boldsymbol{\nu}_1', \Delta\mathbf{u}_2', \ldots, \Delta\mathbf{u}_T')'$ is given by

$$\tilde{\Omega} = \sigma_u^2 \begin{pmatrix} \tau\mathbf{I}_N & -\phi(\mathbf{W}_N) & \mathbf{0} & \cdots & \mathbf{0} \\ -\phi(\mathbf{W}_N)' & 2\mathbf{I}_N & -\mathbf{I}_N & & \\ \mathbf{0} & -\mathbf{I}_N & 2\mathbf{I}_N & & \\ \vdots & & & \ddots & -\mathbf{I}_N \\ \mathbf{0} & & & -\mathbf{I}_N & 2\mathbf{I}_N \end{pmatrix}. \tag{25}$$

The cross-sectional dependence in the covariance of $\boldsymbol{\nu}_1$ and $\Delta\mathbf{u}_2$ is an unfortunate characteristic of this model as it considerably complicates both the numerical and the analytical inversion of $\tilde{\Omega}$.[22] As a computationally tractable circumvention of this problem, I propose the approximation

[21] See also Kripfganz (2015).

[22] It can be shown that using forward-orthogonal deviations instead of first differences does not solve this problem. While in that case $E[\Delta_s\mathbf{u}_s\Delta_t\mathbf{u}_t'] = \mathbf{0}$ for all $s \neq t$, it holds $E[\Delta_0\mathbf{u}_{-s}\Delta_{T-s}\mathbf{u}_{T-s}'] = -\sqrt{s/[(s+1)T(T+1)]}\sigma_u^2\mathbf{I}_N$ for all $s > 0$, which creates a nonzero correlation between the transformed errors $\Delta_t\mathbf{u}_t$, $t = \min(1, T-p), \ldots, T-1$, and the composite initial observations error.

$\tilde{\Omega} \approx \Omega \otimes \mathbf{I}_N$, with

$$\Omega = \sigma_u^2 \Omega^* = \sigma_u^2 \begin{pmatrix} \tau & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix}, \tag{26}$$

which corresponds to a truncation of the polynomial $\phi(\mathbf{W}_N)$ in equation (25) at order zero. The covariance matrix (26) is identical to that of the pure time dynamic model analyzed by Hsiao et al. (2002) who show that $|\Omega| = \sigma_u^{2T}[1 + T(\tau - 1)]$. As I demonstrate in Section 6, the transformed likelihood estimator performs well under this approximation.

4 Quasi-maximum likelihood estimation

With

$$\Delta \mathbf{y}_t = \begin{cases} \tilde{g}(\boldsymbol{\nu}_1) = E[\Delta \mathbf{y}_1 | \Delta \mathbf{x}] + \mathbf{S}_N^{-1} \phi^{-1}(\mathbf{W}_N) \mathbf{R}_N^{-1} \boldsymbol{\nu}_1 & , t = 1 \\ g(\Delta \mathbf{u}_t) = E[\Delta \mathbf{y}_t | \Delta \mathbf{y}_{t-1}, \ldots, \Delta \mathbf{y}_1, \Delta \mathbf{x}] + \mathbf{S}_N^{-1} \mathbf{R}_N^{-1} \Delta \mathbf{u}_t & , t = 2, 3, \ldots T \end{cases}, \tag{27}$$

we can write the joint density of $\Delta \mathbf{y} = (\Delta \mathbf{y}_1', \Delta \mathbf{y}_2', \ldots, \Delta \mathbf{y}_T')'$ conditional on $\Delta \mathbf{x}$ as

$$\begin{aligned} f_y(\Delta \mathbf{y} | \Delta \mathbf{x}) =& f_{y1}(\Delta \mathbf{y}_1 | \Delta \mathbf{x}) \prod_{t=2}^{T} f_{yt}(\Delta \mathbf{y}_t | \Delta \mathbf{y}_{t-1}, \ldots, \Delta \mathbf{y}_1, \Delta \mathbf{x}^*) \\ =& f_{\nu 1}\left(\tilde{g}^{-1}(\Delta \mathbf{y}_1) | \Delta \mathbf{x}\right) \operatorname{abs}\left(|\tilde{\mathbf{J}}(\Delta \mathbf{y}_1)|\right) \\ & \times \prod_{t=1}^{T} f_{ut}\left(g^{-1}(\Delta \mathbf{y}_t) | \Delta \mathbf{y}_{t-1}, \ldots, \Delta \mathbf{y}_1, \Delta \mathbf{x}\right) \operatorname{abs}\left(|\mathbf{J}(\Delta \mathbf{y}_t)|\right), \end{aligned} \tag{28}$$

where f_{yt}, f_{ut}, and $f_{\nu 1}$ denote the marginal density functions of $\Delta \mathbf{y}_t$, $\Delta \mathbf{u}_t$, and $\boldsymbol{\nu}_1$, respectively. Notice the appearance of the two terms $\operatorname{abs}(|\mathbf{J}(\Delta \mathbf{y}_t)|)$ and $\operatorname{abs}(|\tilde{\mathbf{J}}(\Delta \mathbf{y}_1)|)$ that denote the absolute value of the determinant of the Jacobian matrix of g^{-1} with respect to $\Delta \mathbf{y}_t$ and $\tilde{g}^{-1}$ with respect to $\Delta \mathbf{y}_1$, respectively. Assuming for simplicity that all eigenvalues ω_l of $\mathbf{W}_N$ are real-valued, it is readily seen from equation (27) that

$$|\mathbf{J}(\Delta\mathbf{y}_t)| = |\mathbf{R}_N\mathbf{S}_N| = \prod_{l=1}^{N}(1-\rho_2\omega_l)(1-\rho_0\omega_l), \tag{29}$$

which is independent of t and strictly positive as a consequence of Assumption 6.[23] Similarly,

$$|\tilde{\mathbf{J}}(\Delta\mathbf{y}_1)| = |\mathbf{R}_N\mathbf{S}_N\phi(\mathbf{W}_N)| = \prod_{l=1}^{N}(1-\rho_2\omega_l)(1-\rho_0\omega_l)\left(1+\sum_{k=1}^{Q_N}\phi_k\omega_l^k\right). \tag{30}$$

When f_{ut} and $f_{\nu 1}$ describe the density function of the normal distribution, the log-likelihood function is given by

$$\begin{aligned}\ln\mathcal{L} = &-\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\ln|\tilde{\Omega}| - \frac{1}{2}\Delta\mathbf{u}'\tilde{\Omega}^{-1}\Delta\mathbf{u} \\ &+ T\sum_{l=1}^{N}\ln(1-\rho_2\omega_l) + T\sum_{l=1}^{N}\ln(1-\rho_0\omega_l) + \sum_{l=1}^{N}\ln\left(1+\sum_{k=1}^{Q_N}\phi_k\omega_l^k\right),\end{aligned} \tag{31}$$

with

$$\Delta\mathbf{u} = (\mathbf{I}_T\otimes\mathbf{R}_N)\mathbf{\Phi}_{NT}[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}^*\boldsymbol{\theta}^*], \tag{32}$$

where $\boldsymbol{\theta}^* = (\tilde{\boldsymbol{\theta}}', \boldsymbol{\theta}')'$, $\tilde{\boldsymbol{\theta}} = (\psi_0, \psi_1, \dots, \psi_{Q_N}, \boldsymbol{\pi}_0', \boldsymbol{\pi}_1', \dots, \boldsymbol{\pi}_{Q_N}')'$, $\boldsymbol{\theta} = (\lambda, \rho_1, \beta_0, \gamma_0, \beta_1, \gamma_1)'$, $\boldsymbol{\pi}_k = (\pi_{k1}, \pi_{k2}, \dots, \pi_{kT})'$,

$$\mathbf{\Phi}_{NT} = \begin{pmatrix}\phi(\mathbf{W}_N) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{N(T-1)}\end{pmatrix}, \quad \Delta\mathbf{Z}^* = \begin{pmatrix}\Delta\tilde{\mathbf{Z}} & \mathbf{0}\\ \mathbf{0} & \Delta\mathbf{Z}\end{pmatrix},$$

$\Delta\tilde{\mathbf{Z}} = (\boldsymbol{\iota}_N, \mathbf{W}_N\boldsymbol{\iota}_N, \dots, \mathbf{W}_N^{Q_N}\boldsymbol{\iota}_N, \Delta\mathbf{X}, \mathbf{W}_N\Delta\mathbf{X}, \dots, \mathbf{W}_N^{Q_N}\Delta\mathbf{X})$, $\Delta\mathbf{X} = (\Delta\mathbf{x}_1, \Delta\mathbf{x}_2, \dots, \Delta\mathbf{x}_T)$, and

$$\Delta\mathbf{Z} = \begin{pmatrix}\Delta\mathbf{y}_1 & \mathbf{W}_N\Delta\mathbf{y}_1 & \Delta\mathbf{x}_2 & \mathbf{W}_N\Delta\mathbf{x}_2 & \Delta\mathbf{x}_1 & \mathbf{W}_N\Delta\mathbf{x}_1\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \Delta\mathbf{y}_{T-1} & \mathbf{W}_N\Delta\mathbf{y}_{T-1} & \Delta\mathbf{x}_T & \mathbf{W}_N\Delta\mathbf{x}_T & \Delta\mathbf{x}_{T-1} & \mathbf{W}_N\Delta\mathbf{x}_{T-1}\end{pmatrix}.$$

[23]Compare Ord (1975). If $\mathbf{W}_N$ has complex eigenvalues, it holds that $\mathrm{abs}(|\mathbf{S}_N|) = \prod_{l=1}^{N}[(1-\rho_0\mathrm{Re}(\omega_l))^2 + (\rho_0\mathrm{Im}(\omega_l))^2]^{1/2}$, where $\mathrm{Re}(\omega_l)$ and $\mathrm{Im}(\omega_l)$ denote the real and imaginary part of ω_l, respectively (Lee, 2002, Note 7). In the remainder of the paper, I assume for simplicity that all eigenvalues are real. The latter is guaranteed if $\mathbf{W}_N$ is symmetric, or if $\mathbf{W}_N$ is similar to a symmetric matrix which holds if $\mathbf{W}_N$ was symmetric before row standardization.

Under the approximation $\tilde{\Omega} \approx \Omega \otimes \mathbf{I}_N$ we get the simplifications $\tilde{\Omega}^{-1} = \Omega^{-1} \otimes \mathbf{I}_N$ and $\ln|\tilde{\Omega}| = NT \ln(\sigma_u^2) + N \ln(1 + T(\tau - 1))$.

When the order Q_N increases with the sample size, an incidental-parameters problem might occur depending on the rate at which Q_N grows to infinity. In that case, a solution lies in the truncation of the polynomials $\phi(\mathbf{W}_N)$, $\psi(\mathbf{W}_N)$, and $\pi_l(\mathbf{W}_N)$ such that the truncation order K_N satisfies the following condition:

Assumption 9: The truncation order K_N satisfies $K_N \leq Q_N$ for all N, $K_N \to Q$ as $Q_N \to Q$ for $0 < Q \leq \infty$, and $K_N^3/N \to 0$ as $N \to \infty$.[24]

Q_N might be bounded or not. If $Q < \infty$ it is in most cases safe to set $K_N = Q_N$ unless there are not enough degrees of freedom in small samples. In the case of $\mathbf{W}_N = \mathbf{I}_R \otimes \mathbf{B}_M$ as before, $Q_N = 1$ for all N and no truncation is needed to obtain consistent estimates. Now consider a generalization with a potentially differing number of members in each group. As shown by Kripfganz (2015), Q_N is then determined by the number of distinct group sizes. If there is an upper bound $\bar{M}$ for the group size, Q_N may still depend on the sample size but is bounded by $\bar{M}$. When there is no upper bound on the group size, we have to distinguish between increasing-domain asymptotics, $R \to \infty$, and fixed-domain asymptotics, $M \to \infty$ while R is fixed. In the first case, $Q_N \to \infty$ as $R \to \infty$ and a truncation becomes necessary unless further restrictions are imposed on the distribution of the group sizes. In the second case, Q_N is bounded above by R. However, fixed-domain asymptotics are incompatible with Assumption 3 as discussed by Lee (2004).

Maximization of the log-likelihood function can be done with an iterative procedure such as Newton-Raphson. Appendix A provides the analytical first-order and second-order derivatives. The initial parameter estimates can be obtained from a consistent GMM estimation. Alternatively, we can obtain initial estimates for $(\lambda, \beta_0, \beta_1, \psi_0, \boldsymbol{\pi}_0', \sigma_u^2, \tau)$ with the transformed likelihood estimator of Hsiao et al. (2002) assuming absence of spatial dependence. The remaining parameters are initialized as zero. We can further improve the quality of the initial estimates by applying a sequential iterative procedure similar to the algorithm proposed by Anderson and Hsiao (1982). By substituting the initial estimate of τ into the closed-form solutions for the parameters in $\boldsymbol{\theta}^*$ we

[24] The condition $K_N^3/N \to 0$ originates from the work of Berk (1974) and Said and Dickey (1984). Chudik and Pesaran (2013) assume $K_N^3/T \to c$, $0 < c < \infty$, as $T \to \infty$ such that $K_N^3/(NT) \to 0$ as both $(T, N) \to \infty$.

can update the latter, keeping fixed the spatial parameters ρ_0, ρ_2, and $\phi_1, \phi_2, \ldots, \phi_{K_N}$ at zero (or any other initial value). These new values can then be used to obtain a new estimate of τ, and so on. In contrast to the pure time dynamic model, this sequential iterative approach does not yield the quasi-maximum likelihood estimates because closed-form solutions for the spatial parameters ρ_0, ρ_2, and $\phi_1, \phi_2, \ldots, \phi_{K_N}$ are not available.

5 Restricted model specifications

In the literature, several restricted versions of model (9) are considered.[25] In this section, I focus on some of these restrictions that have particular implications for econometric modeling. All of the following restrictions can be tested by means of a likelihood ratio test. The test statistic has a χ^2 distribution with r degrees of freedom that are determined by the number of imposed restrictions:

$$\mathcal{LR} = 2(\ln \mathcal{L} - \ln \mathcal{L}_0) \sim \chi^2_r, \tag{33}$$

where $\ln \mathcal{L}$ denotes the maximized value of the unrestricted log-likelihood function (31) and $\ln \mathcal{L}_0$ that of the restricted log-likelihood function.

5.1 Restricted time-space dynamic panel data model

Parent and LeSage (2012) consider the particularly interesting nonlinear restriction $\rho_1 = -\lambda\rho_0$ that leaves the model time-space dynamic but enhances the interpretability of the model parameters as it disentangles the time and spatial effects.[26] The same restriction emerges also from the theoretical bargaining model with intra-household wage spillovers discussed in Section 2. In the presence of distributed spatial lags, I extend this idea and impose the additional nonlinear restriction $\beta_1\gamma_0 = \beta_0\gamma_1$.

The first restriction implies $\mathbf{A}_N = \lambda\mathbf{S}_N$, and the second restriction yields $\mathbf{B}_{1N} = \kappa\mathbf{B}_{0N}$ with

[25] Elhorst (2001) summarizes the resulting models under certain parameter restrictions.

[26] Alternatively, both effects can be separated by setting $\rho_1 = 0$. This restriction does not deserve particular attention because it leaves the derivations in Sections 3 and 4 unaffected.

$\kappa = \beta_1/\beta_0$, assuming that $\beta_0 \neq 0$. As a consequence, model (9) becomes

$$\mathbf{S}_N \mathbf{y}_t = \lambda \mathbf{S}_N \mathbf{y}_{t-1} + \mathbf{B}_{0N}(\mathbf{x}_t + \kappa \mathbf{x}_{t-1}) + \mathbf{e}_t. \tag{34}$$

The stationarity condition imposed by Assumption 7.1 simplifies to $|\lambda| < 1$. Under Assumption 7.2, we may assume $\varphi(\mathbf{W}_N) = \mathbf{S}_N^{-1}\mathbf{R}_N^{-1}$ to simplify the distribution of the initial observations. Similarly, we shall assume $\varsigma(\mathbf{W}_N) = \beta_0 \mathbf{B}_{0N}^{-1}\mathbf{R}_N^{-1}$ in Assumption 8, provided that $\mathbf{B}_{0N}$ is nonsingular and $|\gamma_0/\beta_0| < 1$. Under these two restrictions, $\psi(\mathbf{W}_N) = \psi \mathbf{R}_N^{-1}$ in equation (18), and the projection error for the initial observations has the same spatial dependence structure as the error terms for the other periods, $\phi^{-1}(\mathbf{W}_N)\mathbf{R}_N^{-1}\boldsymbol{\nu}_1 = \mathbf{R}_N^{-1}\boldsymbol{\nu}_1$, since $\phi(\mathbf{W}_N)$ collapses to the identity matrix. Finally, $\pi_l(\mathbf{W}_N)$ in equation (19) does not depend on the particular form of $\varphi(\mathbf{W}_N)$ and $\varsigma(\mathbf{W}_N)$ and becomes a first-order polynomial. The restriction that is imposed on $\varsigma(\mathbf{W}_N)$ may be disputable. Particularly in the absence of distributed spatial lags, that is $\gamma_0 = \gamma_1 = 0$, it boils down to the assumption that the spatial dependence structure of x_{it} coincides with the spatial error dependence of y_{it}.

If we accept all of the above restrictions, the initial observations satisfy

$$\mathbf{S}_N \Delta \mathbf{y}_1 = \psi \mathbf{R}_N^{-1}\boldsymbol{\iota}_N + \sum_{l=1}^{T}(\pi_{0l}\mathbf{I}_N + \pi_{1l}\mathbf{W}_N)\Delta \mathbf{x}_l + \mathbf{R}_N^{-1}\boldsymbol{\nu}_1. \tag{35}$$

Notice that the approximation of the variance-covariance matrix proposed in Section 3.4 becomes obsolete as well. Consequently, the restricted log-likelihood function is given as

$$\begin{aligned} \ln \mathcal{L}_r = & -\frac{NT}{2}\ln(2\pi\sigma_u^2) - \frac{N}{2}\ln\left(1 + T(\tau - 1)\right) - \frac{1}{2}\Delta \mathbf{u}_r'(\Omega^{-1} \otimes \mathbf{I}_N)\Delta \mathbf{u}_r \\ & + T\sum_{l=1}^{N}\ln(1 - \rho_2\omega_l) + T\sum_{l=1}^{N}\ln(1 - \rho_0\omega_l), \end{aligned} \tag{36}$$

with

$$\Delta \mathbf{u}_r = (\mathbf{I}_T \otimes \mathbf{R}_N)[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta \mathbf{y} - \Delta \mathbf{Z}_r^* \boldsymbol{\theta}_r^*] - \psi \mathbf{s}_1, \tag{37}$$

where $\mathbf{s}_1 = (\boldsymbol{\iota}_N', \mathbf{0}, \ldots, \mathbf{0})'$, and $\boldsymbol{\theta}_r^* = (\tilde{\boldsymbol{\theta}}_r', \boldsymbol{\theta}_r')'$ is partitioned in a similar fashion as in the unrestricted case, with $\tilde{\boldsymbol{\theta}}_r = (\boldsymbol{\pi}_0', \boldsymbol{\pi}_1')'$, $\boldsymbol{\theta}_r = (\lambda, \beta_0, \gamma_0)'$, and $\boldsymbol{\pi}_k = (\pi_{k1}, \pi_{k2}, \ldots, \pi_{kT})', k = 0, 1.$

Moreover,

$$\Delta\mathbf{Z}_r^* = \begin{pmatrix} \Delta\tilde{\mathbf{Z}}_r & \mathbf{0} \\ \mathbf{0} & \Delta\mathbf{Z}\mathbf{H} \end{pmatrix},$$

where $\Delta\tilde{\mathbf{Z}}_r = (\Delta\mathbf{X}, \mathbf{W}_N\Delta\mathbf{X})$, and $\mathbf{H}$ is a matrix that contains the parameter restrictions:

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_\rho & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_\kappa \end{pmatrix}, \quad \mathbf{h}_\rho = \begin{pmatrix} 1 \\ -\rho_0 \end{pmatrix}, \quad \mathbf{H}_\kappa = \begin{pmatrix} \mathbf{I}_2 \\ \kappa\mathbf{I}_2 \end{pmatrix}.$$

The analytical derivatives for a gradient-based optimization are provided in Appendix A.

The degrees of freedom for the likelihood ratio test exceed the two restrictions $\rho_1 = -\lambda\rho_0$ and $\beta_1\gamma_0 = \beta_0\gamma_1$ because the additional coefficients for the higher-order spatial lags in equation (17) compared to equation (35) are treated as free parameters. The total number of restrictions is $r = 2+(T+2)K_N - T$. If $\mathbf{W}_N$ is row standardized, r is reduced by K_N. If more than one regressor is included in the model, each additional regressor adds another $T(K_N - 1)$ restrictions.[27]

5.2 Pure time dynamic panel data model

The time dynamic model without spatial lags is nested in model (9) by restricting $\rho_0 = \rho_1 = \gamma_0 = \gamma_1 = 0$ such that

$$\mathbf{y}_t = \lambda\mathbf{y}_{t-1} + \beta_0\mathbf{x}_t + \beta_1\mathbf{x}_{t-1} + \mathbf{e}_t. \tag{38}$$

Without loss of generality, we can leave the spatial error coefficient ρ_2 unrestricted. Under the Assumption 7.1, the stationarity condition simplifies again to $|\lambda| < 1$. If the process started in the finite past and $\mathbf{W}_N$ is not row standardized, we need to restrict $\varphi(\mathbf{W}_N) = \mathbf{R}_N^{-1}$ in Assumption 7.2 unless $\varphi_0 = 0$. Otherwise, the polynomial $\phi(\mathbf{W}_N)$ does not vanish. As in the previous subsection, the projection error $\phi^{-1}(\mathbf{W}_N)\mathbf{R}_N^{-1}\boldsymbol{\nu}_1$ still suffers from cross-sectional correlation if we continue to allow for spatial dependence in the data generating process of $\mathbf{x}_t$ by leaving $\varsigma(\mathbf{W}_N)$ unconstrained. We thus have to additionally impose $\varsigma(\mathbf{W}_N) = \mathbf{R}_N^{-1}$. $\pi_l(\mathbf{W}_N)$ becomes a scalar multiple of the identity matrix irrespective of any restriction on $\varphi(\mathbf{W}_N)$ or $\varsigma(\mathbf{W}_N)$.

The absence of spatial dependence can be tested again with a likelihood ratio test. With a

[27] When $\gamma_0 = \gamma_1 = 0$ both in the unrestricted and the restricted model, as considered by Parent and LeSage (2012), we have $r = 2 + (T+2)K_N$ and each additional regressor adds TK_N restrictions.

similar argument as in the previous subsection, the degrees of freedom are $r = 4 + (T + 2)K_N$ because the spatial lag coefficients in the initial observations are treated as free parameters in the unrestricted model and need to be restricted to zero as well. If $\mathbf{W}_N$ is row standardized the number of restrictions reduces to $r = 4 + (T + 1)K_N$. If more than one regressor is included in the model, each additional regressor adds another TK_N restrictions.

5.3 Pure space dynamic panel data model

The space dynamic model without time lags is nested as well in model (9) by considering the restrictions $\lambda = \rho_1 = \beta_1 = \gamma_1 = 0$. In this case, the distribution of the initial observations does not depend on unobservables and can be obtained directly from the model. Therefore, it is irrelevant for the specification of the likelihood function if the exogenous regressors are cross-sectionally correlated or not. Technically, we have to impose the additional $(T + 2)(K_N + 1) - 1$ restrictions $\pi_{10} = \beta_0$, $\pi_{11} = \gamma_0$, $\pi_{sk} = 0$ if $s > 1$ or $k > 1$, as well as $\phi_k = \psi_k = 0$ for all $k \geq 1$. Also, the initial observations error term now has the same distribution as the errors for the other time periods. This implies the final restriction $\tau = 2$. The likelihood ratio test statistic (33) is thus based on $r = 4 + (T + 2)(K_N + 1)$ degrees of freedom that are again reduced by K_N if $\mathbf{W}_N$ is row standardized. Each additional regressor adds $T(K_N + 1)$ restrictions.

Notice that by imposing these linear parameter restrictions, the unconditional QML estimator obtained in Section (4) collapses to the QML estimator proposed by Lee and Yu (2010b).[28] Furthermore, with the two additional restrictions $\rho_0 = \gamma_0 = 0$, the least squares dummy variables estimator for the static model can also be viewed as a constrained estimator within the general time-space dynamic framework.

[28]Lee and Yu (2010b) consider the forward-orthogonal model transformation instead of first differences to remove the incidental fixed effects parameters. However, the estimator is invariant to the choice of the transformation matrix as demonstrated by Arellano and Bover (1995).

6 Monte Carlo simulation

6.1 Simulation design

To analyze the finite-sample performance of the estimator, I conduct the following Monte Carlo experiments. The data generating process of the dependent variable, y_{it}, has the same time-space dynamic structure as in the simulation exercise of Parent and LeSage (2012):

$$\mathbf{y}_t = \lambda_y \mathbf{y}_{t-1} + \rho_{y0}\mathbf{W}_N\mathbf{y}_t + \rho_{y1}\mathbf{W}_N\mathbf{y}_{t-1} + \beta\mathbf{x}_t + \boldsymbol{\alpha}_y + \mathbf{u}_t, \quad \mathbf{u}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_u^2\mathbf{I}_N). \tag{39}$$

The regressor x_{it} features similar autoregressive dynamics in both dimensions:

$$\mathbf{x}_t = \lambda_x \mathbf{x}_{t-1} + \rho_{x0}\mathbf{W}_N\mathbf{x}_t + \rho_{x1}\mathbf{W}_N\mathbf{x}_{t-1} + \boldsymbol{\alpha}_x + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}_N), \tag{40}$$

such that $\mathbf{x}_t$ is strictly exogenous with respect to $\mathbf{u}_t$. The unobserved unit-specific effects $\boldsymbol{\alpha}_y$ and $\boldsymbol{\alpha}_x$ are generated from a joint normal distribution:

$$\begin{pmatrix} \alpha_{yi} \\ \alpha_{xi} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha y}^2 & \sigma_{\alpha yx} \\ \sigma_{\alpha yx} & \sigma_{\alpha x}^2 \end{pmatrix} \right). \tag{41}$$

When $\sigma_{\alpha yx} \neq 0$ it is inappropriate to consider the unit-specific effects in equation (39) as random with respect to $\mathbf{x}_t$. The spatial weights matrix $\mathbf{W}_N$ is generated following Lee (2004). There are R regions with M members each. All members have equal weight within the same region and are spatially unrelated to members of other regions. Therefore, $\mathbf{W}_N = \mathbf{I}_R \otimes \mathbf{B}_M$ with $\mathbf{B}_M = (\boldsymbol{\iota}_M\boldsymbol{\iota}_M' - \mathbf{I}_M)/(M-1)$.

I set the unconditional long-run effect of $\mathbf{x}_t$ on $\mathbf{y}_t$ to unity. That is $\beta = 1 - \lambda_y - \rho_{y0} - \rho_{y1}$.[29] The processes are initialized at $t = -5$ with their long-run means conditional on the realizations

[29] Parent and LeSage (2012) restrict the long-run effect of $\mathbf{x}_t$ on $(\mathbf{I}_N - \rho_{y0}\mathbf{W}_N)\mathbf{y}_t$ to unity such that $\beta = (1 - \lambda_y - \rho_{y0} - \rho_{y1})/(1 - \rho_{y0})$. However, they set down the reciprocal of that fraction which is presumably a typographical error.

of the unit-specific effects:

$$\mathbf{y}_{-5} = [(1-\lambda_y)\mathbf{I}_N - (\rho_{y0}+\rho_{y1})\mathbf{W}_N]^{-1}(\beta\mathbf{x}_{-5} + \boldsymbol{\alpha}_y) \tag{42}$$

$$\mathbf{x}_{-5} = [(1-\lambda_x)\mathbf{I}_N - (\rho_{x0}+\rho_{x1})\mathbf{W}_N]^{-1}\boldsymbol{\alpha}_x. \tag{43}$$

The matrix inverses can be computed efficiently with the inversion formula (11). The first 5 observations are discarded. Furthermore, I fix $\lambda_y = \lambda_x = 0.4$, $\rho_{y0} = \rho_{x0} = 0.2$, $\rho_{y1} = \rho_{x1} = -0.08$, $\sigma_u^2 = \sigma_\epsilon^2 = 1$, $\sigma_{\alpha y}^2 = \sigma_{\alpha x}^2 = 3$, and $\sigma_{\alpha yx} = 1.5$. The time span is $T = 9$ and the cross-sectional dimension varies along $R \in \{10, 20, 50\}$ and $M \in \{2, 5, 10, 20, 50\}$. For each simulation I perform 1000 replications.

I estimate the model with three versions of the unconditional transformed likelihood estimator developed in this paper. For the first estimator, I impose the nonlinear restriction $\rho_{y1} = -\lambda_y\rho_{y0}$ which is valid given the parameter values in the data generating process. As discussed in Section 5, this restriction simplifies the distribution of the initial observations by removing the spatial lags of the exogenous regressor. However, since x_{it} itself is spatially correlated the initial observations error term will still exhibit cross-sectional dependence, and by the Cayley-Hamilton theorem the order of $\phi(\mathbf{W}_N)$ equals one. To analyze the impact of a different choice for the truncation order, I further compare the performance of two otherwise unrestricted estimators. The truncation order K_N is either set to zero or unity for all polynomials $\pi_l(\mathbf{W}_N)$ and $\phi(\mathbf{W}_N)$. Finally, as a benchmark I add the bias-corrected conditional transformed likelihood estimator of Yu et al. (2008) to the simulation study. The optimization is performed with a Newton-Raphson algorithm, and analytical derivatives of the log-likelihood function are used to sizeably increase the computation speed of this gradient-based optimization.[30]

6.2 Simulation results

Tables 1 and 2 show the average bias and the root mean square error (RMSE) of the parameter estimates for different combinations of R and M. With increasing R and fixed M the spatial weights matrix becomes relatively more sparse, while with increasing M and fixed R it becomes

[30]First-order and second-order derivatives can be found in Appendix A. The bias-corrected estimator of Yu et al. (2008) is briefly sketched in Appendix B. *Stata* estimation files are available from the author upon request.

relatively more dense. The simulation results reveal that this distinction is important for the spatial lag coefficients ρ_{y0} and ρ_{y1}. When M increases, neither the bias nor the RMSE of the estimators show a clear tendency to go down for these coefficients. This observation is in line with the argumentation of Lee (2004) that the quasi-maximum likelihood estimators are inconsistent under fixed-domain asymptotics.[31] Intuitively, by holding the number of regions R fixed and increasing the number of members M the spatial structure changes and there is no informational gain about the spatial dependence parameters. To the contrary, with increasing domain R both the bias and the RMSE of all estimators shrink for the spatial lag coefficients. Clearly, the statistical inference can be improved by drawing additional regions R from the population that have the same dimension as the previously sampled data. For the coefficients λ_y and β, the direction of the asymptotics does not matter. In fact, the bias and RMSE have about the same magnitude for a given sample size N irrespective of the ratio M/R.

[Table 1 about here.]

[Table 2 about here.]

Another observation concerns the unconditional transformed likelihood estimator with truncation order $K_N = 1$. For small sample sizes, in particular when $R = 10$, it shows unexpectedly large biases and dispersions for the parameters that capture the dynamics across time and space. In the extreme case, $N = RM = 20$, the estimator even breaks down completely as the Newton-Raphson algorithm fails to converge. This is likely to be a consequence of overparameterizing the distribution of the initial observations. With more observations at hand convergence can be achieved although large distortions remain when the sample increases only in the direction of M. I conjecture that the source of these distortions can be found in the vicinity of the too many instruments problem in instrumental variable estimations. As a remedy serves the restriction of the truncation order to $K_N = 0$. The resulting estimator convinces with a reasonably small bias and RMSE even in very small samples. When R increases, the performance of both estimators becomes comparable for the parameters λ_y and β while the truncated version, $K_N = 0$, retains its lead for the spatial lag coefficients ρ_{y0} and ρ_{y1}.

[31] The Monte Carlo findings of Bao (2013) also confirm this result.

Not surprisingly, the unconditional likelihood estimator with the valid nonlinear restriction $\rho_{y1} = -\lambda_y \rho_{y0}$ shows by far the lowest root mean square error for the restricted coefficient ρ_{y1}. The gains spill over to the coefficient λ_y, in particular when M is small, while it seems to come at the cost of a slightly increased RMSE for the coefficient ρ_{y0}.

Leaving aside the specific problems when R is small of the estimator without truncation, $K_N = Q_N = 1$, the unconditional likelihood estimators perform well in comparison to the bias-corrected conditional likelihood estimator of Yu et al. (2008). For the coefficient λ_y, the latter reveals a larger bias than the unconditional estimators but can convince with the smallest RMSE under most combinations of R and M. Interestingly, for the spatial parameters ρ_{y0} and ρ_{y1} this picture flips upside down. Here, the bias-corrected estimator shows a smaller bias than the unconditional estimators but a larger RMSE than its unrestricted competitor with truncation order $K_N = 0$. Eventually, for $R = 50$ even the unrestricted estimator with $K_N = 1$ passes by the bias-corrected estimator in terms of RMSE.

[Table 3 about here.]

While studying the estimator performance for the individual parameters provides useful insights, the main quantities of interest in empirical studies are the marginal effects of x_{it} on y_{it}. In space dynamic models we can divide the total marginal effect into a direct and an indirect effect. The time dynamics in addition allow to distinguish between short-run and long-run effects. Table 3 provides the simulation results for the marginal effects for the sample dimensions $R = 50$ and $M = 2$. For the short-run effects, all the estimators under comparison are essentially indistinguishable. Notable differences emerge for the long-run effects. The unconditional likelihood estimator that exploits the valid restriction $\rho_{y1} = -\lambda\rho_{y0}$ convinces in particular with a RMSE for the indirect long-run effect that is less than half of the RMSE for the other estimators. This comparative advantage carries over to the total long-run effect. However, also the unconditional estimators without this restriction show lower RMSEs than the bias-corrected conditional likelihood estimator. Here, the advantages of modeling the distribution of the initial observations under a short time horizon as opposed to a bias-correction that is derived under large-T asymptotics become evident.

7 Application

7.1 Data description

For the estimation of the time-space dynamic wage equation that was derived in Section 2, I use biennial data from the Panel Study of Income Dynamics (PSID) for the years from 2001 to 2011. According to the Mincerian theory of wage determination, labor market experience plays a key role as an explanatory factor of net potential earnings. I proxy experience by the age of the individuals. Because age is linearly growing over time, as is experience in the absence of career gaps, in first differences it reduces to a constant term. Therefore, its coefficient cannot be identified separately from the set of time-specific effects that are included to capture fluctuations and trends in macroeconomic conditions. However, the coefficient of squared age is still identifiable. Furthermore, I include dummy variables for union coverage, marriage, and the number of children living in the same family unit. When the model includes a spatial lag, the marriage variable is supposed to capture a true marriage premium that is not confounded with a mere cohabitation premium. As additional control variables, I include three industry dummies (primary sector, manufacturing, and public administration) and three regional dummies (northeast, north central, and south), such that the baseline group is workers in the services sector living in the western part of the United States. Other explanatory variables such as education, gender, race, or family background are time-invariant characteristics that are removed by the first-difference transformation.[32] Further details on the variables can be found in Appendix C.

The PSID does not only allow to track individuals over time but also to link them to other individuals living in the same household. I restrict the sample to individuals that report a nonzero hourly wage rate for all periods. Furthermore, I dismiss all individuals that earned less than the federal minimum wage in at least one of the years.[33] To avoid time-varying spatial weights, I dismiss workers that did not stay together with the same cohabitant over the sample period. This leads to a strongly balanced sample of $N = 738$ individuals of which 138 have a working housemate. Consequently, there is a total of $R_1 = 600$ households with workers that are either single or have a

[32]The effects of time-invariant regressors could be estimated under potentially restrictive exogeneity assumptions by adjusting the two-stage procedure proposed by Kripfganz and Schwarz (2019).

[33]The federal minimum wage was raised from 5.15 USD to 7.25 USD per hour on July 24, 2009. I selected the individuals based on the former for the years 2001 to 2009, and based on the latter for the final year 2011.

partner that is not working (and thus not part of the sample), and $R_2 = 69$ dual-earner households. By appropriately ordering the individuals, the spatial weights matrix has the following symmetric and block-diagonal structure:

$$\mathbf{W}_N = \begin{pmatrix} \mathbf{I}_{R_1} \otimes \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{R_2} \otimes \mathbf{B}_2 \end{pmatrix}, \quad \mathbf{B}_1 = 0, \quad \mathbf{B}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This matrix is a special case of the spatial weights matrix discussed in Section 3.2 with $\tilde{R} = \tilde{M} = 2$. $\mathbf{W}_N$ is diagonalizable and has three distinct eigenvalues, namely unity and minus unity with multiplicity R_2 each, and zero with multiplicity $R_1 = N - 2R_2$. As discussed by Kripfganz (2015), we can thus write higher-order powers of $\mathbf{W}_N$ as second-order polynomials which determines the order $Q_N = 2$ of the polynomials $\phi(\mathbf{W}_N)$ and $\pi_l(\mathbf{W}_N)$ in Section 3.3. Moreover, the row sum of $\mathbf{W}_N$ is unity for the last $2R_2$ rows and zero for the remaining rows. It is therefore only partially row standardized which requires $\psi(\mathbf{W}_N)$ to be a first-order polynomial.[34]

In the current analysis, the labor market participation decision is left aside. In general, an increase in the relative wage of the cohabitant may have an adverse effect on the own participation as specialization in household labor becomes comparatively more advantageous. However, if a worker decides to reduce the number of hours worked this does not have to be reflected in an adjustment of hourly wage rates. If the hours are reduced to zero in any period the worker drops out from the estimation sample for all periods. The resulting sample selection is therefore a time-invariant characteristic and thus captured by the individual-specific effects.

7.2 Marginal effects and the cohabitation premium

For parsimony, consider model (9) with the restrictions $\beta_1 = \gamma_0 = \gamma_1 = 0$. To obtain marginal effects of the exogenous regressor variable, the following result for this particular spatial weights matrix is helpful:[35]

$$(\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} = \mathbf{I}_N + \frac{\rho_0}{1-\rho_0^2}\mathbf{W}_N + \frac{\rho_0^2}{1-\rho_0^2}\mathbf{W}_N^2, \tag{44}$$

[34] A similar argument applies to $\pi_l(\mathbf{W}_N)$ if x_{it} is a household-specific regressor that takes on the same value for both partners, as it is the case for marriage, children, and the regional dummies.

[35] See LeSage and Pace (2009) for a discussion of direct, indirect, and total marginal effects in spatial econometric models with an arbitrary spatial weights matrix.

where $\mathbf{W}_N^2$ equals the identity matrix with the first R_1 main diagonal elements replaced by zero. We then have the short-run direct marginal effect

$$\frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{it}} = \begin{cases} \beta & , i \leq R_1 \\ \beta \left(1 + \frac{\rho_0^2}{1-\rho_0^2}\right) & , i > R_1 \end{cases}, \tag{45}$$

and the short-run indirect marginal effect

$$\frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{jt}} = \begin{cases} 0 & , i \leq R_1 \\ \beta \frac{\rho_0}{1-\rho_0^2} & , i > R_1 \end{cases}, \tag{46}$$

where j refers to the cohabitant of individual i. The short-run total effect equals the sum of both:

$$\frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{it}} + \frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{jt}} = \begin{cases} \beta & , i \leq R_1 \\ \beta \frac{1}{1-\rho_0} & , i > R_1 \end{cases}. \tag{47}$$

For single-earner households, $i \leq R_1$, there are no within-household spillover effects such that the total effect equals the direct effect, and both are just given by the coefficient β. The immediate spillover effect in dual-earner households manifests itself in the nonzero indirect effect, but also the direct effect is larger than for single earners because of second-round effects. The total effect for dual-earner households is then obtained by scaling β with the spatial multiplier $1/(1-\rho_0)$. We can thus compute a regressor-specific short-run cohabitation premium that individuals obtain from sharing a household with another worker as the difference of the total effects for dual-earner and single-earner households:

$$\frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}; i > R_1]}{\partial x_{it}} + \frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}; i > R_1]}{\partial x_{jt}} - \frac{\partial E[y_{it}|\mathbf{y}_{t-1}, \mathbf{x}_t, \boldsymbol{\alpha}; i \leq R_1]}{\partial x_{it}} = \beta \frac{\rho_0}{1-\rho_0}. \tag{48}$$

Corresponding long-run effects depend on potential restrictions imposed on the parameter ρ_1.

For the unrestricted model, the long-run direct marginal effects are obtained as

$$\frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{it}} = \begin{cases} \beta \frac{1}{1-\lambda} & , i \leq R_1 \\ \beta \frac{1}{1-\lambda} \left(1 + \frac{(\rho_0+\rho_1)^2}{(1-\lambda)^2-(\rho_0+\rho_1)^2}\right) & , i > R_1 \end{cases}, \tag{49}$$

and the long-run indirect marginal effects become

$$\frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{jt}} = \begin{cases} 0 & , i \leq R_1 \\ \beta \frac{\rho_0+\rho_1}{(1-\lambda)^2-(\rho_0+\rho_1)^2} & , i > R_1 \end{cases}. \tag{50}$$

Adding them up yields the long-run total marginal effects

$$\frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{it}} + \frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}]}{\partial x_{jt}} = \begin{cases} \beta \frac{1}{1-\lambda} & , i \leq R_1 \\ \beta \frac{1}{1-\lambda-\rho_0-\rho_1} & , i > R_1 \end{cases}. \tag{51}$$

Under the restriction $\rho_1 = -\lambda\rho_0$ the spatial long-run multiplier $1/(1 - \lambda - \rho_0 - \rho_1)$ becomes the product of the simple long-run multiplier $1/(1 - \lambda)$ and the short-run spatial multiplier $1/(1 - \rho_0)$. Finally, we can compute the unrestricted long-run cohabitation premium again as the difference between the dual-earner and single-earner total effects:

$$\frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}; i > R_1]}{\partial x_{it}} + \frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}; i > R_1]}{\partial x_{jt}} - \frac{\partial E[y_{it}|\mathbf{x}_t, \boldsymbol{\alpha}; i \leq R_1]}{\partial x_{it}} = \beta \frac{\rho_0 + \rho_1}{(1 - \lambda)(1 - \lambda - \rho_0 - \rho_1)}, \tag{52}$$

which simplifies to $[\beta/(1 - \lambda)] \cdot [\rho_0/(1 - \rho_0)]$ under the nonlinear restriction $\rho_1 = -\lambda\rho_0$. Standard errors for all marginal effects and cohabitation premiums can be obtained with the Delta method.

7.3 Estimation results

Table 4 presents the estimation results for a set of wage equations with different dynamic model components. All models are nested within the general time-space dynamic panel data model (7) as outlined in Section 5. In line with the dynamic wage equation derived from the bargaining model in Section 2, I do not consider distributed time or spatial lags of the regressors. Also, to avoid identification problems I assume absence of spatial error dependence. Thus, $\beta_1 = \gamma_0 = \gamma_1 = \rho_2 = 0$ in all specifications.

[Table 4 about here.]

I compare several restricted versions of the model. The static model does neither control for time dependence nor for within-household dependence. The space dynamic model adds a spatial lag of the dependent variable in line with the argumentation that wage rates are correlated among household members.[36] In contrast, the empirical observation of correlated earnings over time is accommodated in the time dynamic model. For this specification, I assume absence of spatial dependence both in the dependent variable and the exogenous regressors. Thus, the resulting QML estimators for the space dynamic and the time dynamic models are those of Lee and Yu (2010b) and Hsiao et al. (2002), respectively. The time-space dynamic model merges the space dynamic and the time dynamic models. Besides the unrestricted version, I consider two restrictions on the spatial time lag, $\rho_1 = 0$ or $\rho_1 = -\lambda\rho_0$, that both disentangle the time and spatial effects. The latter is in line with the bargaining model derived above. For the time-space dynamic specifications, I allow for an unrestricted spatial correlation of the regressors according to Assumption 8, but set the truncation order to $K_N = 0$ to avoid an overparameterization of the initial observations equation. This truncation has been shown in Section 6 to yield satisfactory results.

The static model delivers the familiar results. The statistically significantly negative coefficient of age squared signals a hump-shaped wage-age profile provided that the unidentified linear age component enters with a positive sign. The aggregate time effects (relative to the initial period) are all positive and growing over time.[37] In the absence of other excluded trending components, this would indeed imply a positive influence of the linear age component. Moreover, there is a strongly statistically significant wage premium of about 6.5 percent for jobs covered by a union contract. The marriage premium of 4.1 percent is statistically significant only at the 10 percent level, as is the effect of the number of children in the household.

Adding a contemporaneous spatial lag to the model has almost no effect on the remaining

[36] The unconditional correlation coefficient of the hourly wage rates among cohabitants in the sample is 0.55. For the corresponding biennial growth rates it is still 0.22.

[37] To economize on space, the estimates of the industry, regional, and time effects are not reported in the table.

coefficient estimates even though the spatial lag coefficient is highly statistically significant. The picture changes when adding a time lag. Both in the pure time dynamic and in the time-space dynamic models, the time lag coefficient is statistically and economically highly significant. If interpreted as a speed of convergence, it implies that wage rate deviations from their long-run trend are reduced by 54 percent within one period (2 years). While the coefficient of union coverage rises due to the inclusion of the time lag, the effects of the marital status and the presence of children shrink towards zero and turn statistically insignificant. When comparing the unrestricted time-space dynamic model to the specification with the nonlinear restriction $\rho_1 = -\lambda\rho_0$, it is evident that the point estimate of the unrestricted coefficient of the spatial time lag comes close to the restricted estimate. Importantly, the restriction reduces the corresponding standard error by factor 2.5 and turns the spatial time lag statistically significant at the 1 percent level while it is insignificant without the restriction.[38]

Concerning the optimal model choice, the time-space dynamic models tend to be preferred. Both the time lag and the contemporaneous spatial lag are statistically significant predictors of the wage outcome. A model comparison on the basis of a likelihood ratio test points in the same direction. We can reject the null hypothesis that the imposed restrictions are valid for the static, space dynamic, and time dynamic model. However, we cannot reject the two restricted versions of the time-space dynamic model against the unrestricted model which is evidence that the underlying dynamics can be characterized by just two instead of three coefficients that capture the time and spatial effect, respectively. In particular, we cannot reject the time-space dynamic wage equation with the nonlinear restriction that results from the bargaining model in Section 2. Also observe that the static model is rejected against any alternative which strongly suggests to consider dynamic model specifications in the analysis of wage determinants.

For comparison of the different models' implications, the marginal effects are more meaningful than pure coefficient estimates. Table 5 presents these effects for dual earners.[39] For the static model, the results are presented under the assumption that the underlying data generating process is static as well. In this case, short-run and long-run marginal effects coincide. If the data generating process is dynamic, the within estimates in static models approximate short-run effects

[38]The standard error of the estimate for $\rho_1 = -\lambda\rho_0$ is computed with the Delta method.

[39]The marginal effect of age is not identified, and a separate effect for age squared is not of interest.

and the between estimates approximate long-run effects.[40] We should then disregard the reported long-run estimates for the static model, and similarly for the pure space dynamic model, because the first-difference transformation yields within estimates. The marginal effects from the static model simply equal the respective coefficient estimates. They are very close to the direct marginal effects from the space dynamic model. The difference between the total marginal effects is larger due to the positive indirect effects in the space dynamic model. The latter is highly statistically significant for union coverage.

[Table 5 about here.]

Stronger differences emerge when we include time dynamics. In the absence of additional spatial dynamics, the short-run direct and total effects are again equal to the respective coefficient estimates, while the long-run effects are more than twice as high due to the long-run multiplier, $1/(1-\lambda) = 2.19$. For marriage and the number of children, the insignificance of the coefficient estimates transmits into insignificant marginal effects. For union coverage, the results confirm the preceding argumentation that the within estimates of static models are approximations to short-run rather than long-run effects.

When we move on to time-space dynamic models that control both for wage persistence over time and interaction effects among cohabitants, the direct effects remain about the same as in the pure time dynamic model. The important difference lies again in the statistically significant indirect effects for job coverage under a union contract. In the short run, this increases the corresponding total effect by about 0.4 to 0.6 percentage points. In the long run, the difference is even more pronounced. The total effect is about 1.1 to 2.1 percentage points higher than under ignorance of within-household spillovers, depending on the restrictions on the spatial time lag. The effects are largest when we restrict $\rho_1 = 0$, but also under the nonlinear restriction $\rho_1 = -\lambda\rho_0$ the effects are slightly higher than with the unrestricted model. More importantly, due to the imprecise estimation of the spatial time lag coefficient in the latter model the long-run indirect effect turns statistically insignificant.

For single-earner households the exposition simplifies a lot because they do not benefit from spillover effects such that the direct and total marginal effects coincide, as reported in Table 6.

[40] See for example Egger and Pfaffermayr (2004).

The short-run marginal effects in all models are given by the corresponding coefficient estimates, and the long-run coefficients are appropriately scaled by the long-run multiplier. For the static and space dynamic model the qualification applies again that we should not put much emphasis on long-run effects when we acknowledge the autoregressive nature of the underlying data generating process. In all models that include time dynamics it is clearly visible that the single-earner marginal effects are virtually the same.

[Table 6 about here.]

We are now ready to compute the short-run and long-run cohabitation premiums that dual earners receive from cohabiting with another worker. This is done in Table 7. Clearly, there is no such premium in the static and the time dynamic model. Under the remaining specifications, the short-run cohabitation premium for union coverage varies between 0.4 and 0.6 percentage points. The corresponding long-run premium lies in the range of 1.1 and 2.1 percentage points, leaving aside the space dynamic model due to the known reasons. Besides the long-run premium in the unrestricted model, these premiums are statistically significant at the 5 percent level. Their magnitude is economically meaningful and nonnegligible given a total long-run effect of about 17 to 19 percent.

[Table 7 about here.]

Finally, we can use the coefficient estimates to make inference on the structural parameters in the bargaining model set out in Section 2. Due to the restriction on the spatial time lag, we can identify only two of the three parameters θ, ϕ, and δ. For a given bargaining weight θ, we can obtain estimates of the semi-elasticities on unemployment benefits, ϕ, and on intra-household transfers, δ, as follows:

$$\hat{\phi} = \frac{1}{1-\theta}\hat{\lambda}, \tag{53}$$

$$\hat{\delta} = \frac{\theta}{1-\theta}\hat{\rho}_0. \tag{54}$$

For a meaningful range of the bargaining weight, Figure 1 plots the corresponding parameter estimates with 95 percent confidence bands based on the coefficient estimates from the restricted time-space dynamic wage equation. For $\theta \in (0, 0.5)$, such that the employer has the higher bargaining weight, the unemployment benefits semi-elasticity tends to be smaller than unity, although not statistically significantly different from unity for values of θ between 0.4 and 0.5. As expected, it is much higher than the semi-elasticity of intra-household transfers that remains close to zero, although statistically significantly positive as a direct consequence of the statistical significance of the spatial lag coefficient.

[Figure 1 about here.]

Summarizing the estimation results, we find evidence for wage persistence over time and within-household spillover effects. On statistical grounds, we have to reject all model specifications that ignore either or both of the two dynamic model components. On the other side, the completely unrestricted model seems to be overparameterized as we can neither reject the absence of the spatial time lag nor a nonlinear restriction on its coefficient. With both specifications we find a significant cohabitation premium for union coverage both in the short and the long run.

8 Conclusion

The formulation of an unconditional transformed likelihood function for short-T dynamic panel data models requires a specification of the marginal distribution for the initial observations. The presence of spatial lags complicates the respective derivations under the assumption that the initial observations are generated from the same process as the remaining observations. The resulting likelihood function involves higher-order spatial lag polynomials. In general, their order can increase with the sample size which requires a truncation for consistent estimation and admissible finite sample performance. When appropriate measures are taken to rein the parameter proliferation, the unconditional transformed likelihood estimator is shown to perform well in finite samples with a smaller root mean square error of the marginal effects estimates than the competing bias-corrected conditional likelihood estimator of Yu et al. (2008).[41]

The derivations simplify considerably by imposing a particular nonlinear restriction on the

[41] It must be noted that using transformed likelihood-based estimators for short-T dynamic panel data models, with or without spatial effects, is not without downsides. For example, recent research by Dhaene and Jochmans (2016), Bun et al. (2017), and Juodis (2018) has revealed potential identification failures due to multimodality of transformed likelihood estimators when T is small.

spatial time lag as discussed earlier by Parent and LeSage (2010, 2011, 2012). The same restriction is also implied by a theoretical bargaining model derived in this paper to rationalize a time-space dynamic wage equation. The empirical results based on PSID data are in favor of a model specification that allows for both autoregressive dynamics and within-household spillovers. They also support the aforementioned nonlinear restriction on the spatial time lag. The rich dynamic model specification allows to separate long-run from short-run effects, and to split total marginal effects into direct and indirect marginal effects. Furthermore, the difference between the marginal effects for single-earner and dual-earner households gives rise to a regressor-specific cohabitation premium.

Simplifications also occur for particular spatial weights matrices with a finite order of their minimal polynomial. Such situations may occur for example in socio-economic settings where individuals interact with others in groups of manageable size while there are no interactions between groups. The resulting spatial weights matrix has a block-diagonal structure with many repeated eigenvalues. The finite order of the corresponding spatial lag polynomials also allows to obtain analytical expressions for the marginal effects as explicit functions of the model parameters.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*(1), 5–27.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

Andini, C. (2007). Returns to education and wage equations: a dynamic approach. *Applied Economics Letters 14*(8), 577–579.

Andini, C. (2013). How well does a dynamic Mincer equation fit NLSY data? Evidence based on a simple wage-bargaining model. *Empirical Economics 44*(3), 1519–1543.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers.

Anselin, L. (2001). Spatial econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Chapter 14, pp. 310–330. Malden: Blackwell Publishing.

Anselin, L. and A. K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and D. E. A. Gilles (Eds.), *Handbook of Applied Economic Statistics*, Chapter 7, pp. 237–289. New York: Marcel Dekker.

Arellano, M. and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Baltagi, B. H., P. Egger, and M. Pfaffermayr (2013). A generalized spatial panel data model with random effects. *Econometric Reviews 32*(5–6), 650–685.

Baltagi, B. H., B. Fingleton, and A. Pirotte (2014). Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics 76*(1), 112–138.

Bao, Y. (2013). Finite-sample bias of the QMLE in spatial autoregressive models. *Econometric Theory 29*(1), 68–88.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics 2*(3), 489–502.

Bhargava, A. and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51*(6), 1635–1659.

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory 21*(4), 795–837.

Blundell, R. and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Bun, M. J. G. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23*(2), 200–210.

Bun, M. J. G., M. A. Carree, and A. Juodis (2017). On maximum likelihood estimation of dynamic panel data models. *Oxford Bulletin of Economics and Statistics 79*(4), 463–494.

Bun, M. J. G. and J. F. Kiviet (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics 132*(2), 409–444.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica 59*(4), 953–965.

Choi, H.-J., J. M. Joesch, and S. Lundberg (2008). Sons, daughters, wives, and the labour market outcomes of West German men. *Labour Economics 15*(5), 795–811.

Chudik, A. and M. H. Pesaran (2013). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews 32*(5–6), 592–649.

Dhaene, G. and K. Jochmans (2016). Likelihood inference in an autoregression with fixed effects. *Econometric Theory 32*(5), 1178–1215.

Egger, P. and M. Pfaffermayr (2004). Estimating long and short run effects in static panel models. *Econometric Reviews 23*(3), 199–214.

Elhorst, J. P. (2001). Dynamic models in space and time. *Geographical Analysis 33*(2), 119–140.

Elhorst, J. P. (2005). Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis 37*(1), 85–106.

Elhorst, J. P. (2010). Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics 40*(5), 272–282.

Elhorst, J. P., E. Zandberg, and J. De Haan (2013). The impact of interaction effects among neighbouring countries on financial liberalization and reform: A dynamic spatial panel data approach. *Spatial Economic Analysis 8*(3), 293–313.

Evans, P. and J. U. Kim (2014). The spatial dynamics of growth and convergence in Korean regional incomes. *Applied Economics Letters 21*(16), 1139–1143.

Fischer, M. M. and J. P. LeSage (2015). A Bayesian space-time approach to identifying and interpreting regional convergence clubs in Europe. *Papers in Regional Science 94*(4), 677–702.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica 70*(4), 1639–1657.

Hayakawa, K. and M. H. Pesaran (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models. *Journal of Econometrics 188*(1), 111–134.

Henz, U. and M. Sundström (2001). Partner choice and women's paid work in Sweden. The role of earnings. *European Sociological Review 17*(3), 295–316.

Ho, C.-Y., W. Wang, and J. Yu (2013). Growth spillover through trade: A spatial dynamic panel data approach. *Economics Letters 120*(3), 450–453.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*(1), 107–150.

Juodis, A. (2018). First difference transformation in panel VAR models: Robustness, estimation, and inference. *Econometric Reviews 37*(6), 650–693.

Kapoor, M., H. H. Kelejian, and I. R. Prucha (2007). Panel data models with spatially correlated error components. *Journal of Econometrics 140*(1), 97–130.

Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*(2), 509–533.

Keller, W. and C. H. Shiue (2007). The origin of spatial interaction. *Journal of Econometrics 140*(1), 304–332.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68*(1), 53–78.

Kripfganz, S. (2015). Matrix powers and marginal effects in spatial autoregressive models. Manuscript, Goethe University Frankfurt. https://ssrn.com/abstract=2393639.

Kripfganz, S. and C. Schwarz (2019). Estimation of linear dynamic panel data models with time-invariant regressors. *Journal of Applied Econometrics 34*(4), 526–546.

Lee, L.-F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory 18*(2), 252–277.

Lee, L.-F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews 22*(4), 307–335.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*(6), 1899–1925.

Lee, L.-F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics 137*(2), 489–514.

Lee, L.-F. and J. Yu (2010a). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory 26*(2), 564–597.

Lee, L.-F. and J. Yu (2010b). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*(2), 165–185.

Lee, L.-F. and J. Yu (2010c). Some recent developments in spatial panel data models. *Regional Science and Urban Economics 40*(5), 255–271.

Lee, L.-F. and J. Yu (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics 180*(2), 174–197.

LeSage, J. P. and R. K. Pace (2009). *Introduction to Spatial Econometrics*. London: Chapman.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics 31*(4), 451–472.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*(1), 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–1426.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Associaton 70*(349), 120–126.

Parent, O. and J. P. LeSage (2010). A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B: Methodological 44*(5), 633–645.

Parent, O. and J. P. LeSage (2011). A space-time filter for panel data models containing random effects. *Computational Statistics and Data Analysis 55*(1), 475–490.

Parent, O. and J. P. LeSage (2012). Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics 42*(4), 727–738.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*(3), 599–607.

Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews 31*(5), 483–531.

Semykina, A. and J. M. Wooldridge (2013). Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics 28*(1), 47–61.

Stratton, L. S. (2002). Examining the wage differential for married and cohabiting men. *Economic Inquiry 40*(2), 199–212.

Tao, J. and J. Yu (2012). The spatial time lag in panel data models. *Economics Letters 117*(3), 544–547.

Verhelst, B. and D. Van den Poel (2014). Deep habits in consumption: a spatial panel analysis using scanner data. *Empirical Economics 47*(3), 959–976.

Yu, J., R. de Jong, and L.-F. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics 146*(1), 118–134.

Yu, J. and L.-F. Lee (2012). Convergence: A spatial dynamic panel data approach. *Global Journal of Economics 1*(1), 1250006.

A Derivatives of the log-likelihood function

A.1 Unrestricted time-space dynamic model

In the following, I consider the approximation $\tilde{\Omega} \approx \Omega \otimes \mathbf{I}_N$ from Section 3.4, with $\Omega = \sigma_u^2 \Omega^*$ given in equation (26). Hsiao et al. (2002) decompose ${\Omega^*}^{-1} = \mathbf{C}'\mathbf{D}^{-1}\mathbf{C}$, where

$$\mathbf{C} = \begin{pmatrix} c_0 & 0 & \dots & 0 \\ c_0 & c_1 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ c_0 & c_1 & \dots & c_{T-1} \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} c_0 c_1 & 0 & \dots & 0 \\ 0 & c_1 c_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & c_{T-1} c_T \end{pmatrix},$$

with $c_s = 1 + s(\tau - 1), s = 0, 1, \dots, T$. Define $\mathbf{P} = \mathbf{D}^{-1/2}\mathbf{C}$, and let $\mathbf{p}_1$ and $\mathbf{P}_{2..T}$ denote the first column and the columns 2 to T of $\mathbf{P}$, respectively. Furthermore, $\tilde{\mathbf{G}} = (\mathbf{p}_1 \otimes \mathbf{R}_N)\phi(\mathbf{W}_N)\Delta\tilde{\mathbf{Z}}$ and $\mathbf{G} = (\mathbf{P}_{2..T} \otimes \mathbf{R}_N)\Delta\mathbf{Z}$. Then, $\Delta\mathbf{u}'(\Omega^{-1} \otimes \mathbf{I}_N)\Delta\mathbf{u} = \sigma_u^{-2}\Delta\tilde{\mathbf{u}}'\Delta\tilde{\mathbf{u}}$ with

$$\Delta\tilde{\mathbf{u}} = (\mathbf{P} \otimes \mathbf{I}_N)\Delta\mathbf{u} = [\mathbf{P} \otimes (\mathbf{R}_N\mathbf{S}_N)]\mathbf{\Phi}_{NT}\Delta\mathbf{y} - \tilde{\mathbf{G}}\tilde{\boldsymbol{\theta}} - \mathbf{G}\boldsymbol{\theta}. \tag{55}$$

The first-order derivatives with respect to $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}$, and σ_u^2 are given by

$$\frac{\partial \ln \mathcal{L}}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma_u^2}\mathbf{G}'\Delta\tilde{\mathbf{u}}, \tag{56}$$
$$\frac{\partial \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}}} = \frac{1}{\sigma_u^2}\tilde{\mathbf{G}}'\Delta\tilde{\mathbf{u}}, \tag{57}$$
$$\frac{\partial \ln \mathcal{L}}{\partial \sigma_u^2} = -\frac{NT}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\Delta\tilde{\mathbf{u}}'\Delta\tilde{\mathbf{u}}. \tag{58}$$

Setting them equal to zero yields the following closed-form solutions for given values of ρ_0, ρ_2, τ, and $\phi_1, \phi_2, \dots, \phi_{K_N}$:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{G}'\tilde{\mathbf{M}}\mathbf{G}\right)^{-1}\mathbf{G}'\tilde{\mathbf{M}}[\mathbf{P} \otimes (\mathbf{R}_N\mathbf{S}_N)]\mathbf{\Phi}_{NT}\Delta\mathbf{y}, \tag{59}$$
$$\hat{\tilde{\boldsymbol{\theta}}} = \left(\tilde{\mathbf{G}}'\mathbf{M}\tilde{\mathbf{G}}\right)^{-1}\tilde{\mathbf{G}}'\mathbf{M}[\mathbf{P} \otimes (\mathbf{R}_N\mathbf{S}_N)]\mathbf{\Phi}_{NT}\Delta\mathbf{y}, \tag{60}$$

where $\mathbf{M} = \mathbf{I}_{NT} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$, $\tilde{\mathbf{M}} = \mathbf{I}_{NT} - \tilde{\mathbf{G}}(\tilde{\mathbf{G}}'\tilde{\mathbf{G}})^{-1}\tilde{\mathbf{G}}'$, and

$$\hat{\sigma}_u^2 = \frac{1}{NT}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}^*\hat{\boldsymbol{\theta}}^*\right]' \boldsymbol{\Phi}_{NT}' \\ (\mathbf{I}_T \otimes \mathbf{R}_N)' \left(\Omega^{*-1} \otimes \mathbf{I}_N\right) (\mathbf{I}_T \otimes \mathbf{R}_N)\boldsymbol{\Phi}_{NT}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}^*\hat{\boldsymbol{\theta}}^*\right]. \quad (61)$$

Inserting these expressions back into the log-likelihood function yields a concentrated log-likelihood function that can be maximized with respect to the remaining parameters. The first-order derivatives for these parameters are given by

$$\frac{\partial \ln \mathcal{L}}{\partial \rho_0} = \frac{1}{\sigma_u^2}(\boldsymbol{\Phi}_{NT}\Delta\mathbf{y})'(\mathbf{I}_T \otimes \mathbf{R}_N)'(\mathbf{P} \otimes \mathbf{W}_N)'\Delta\tilde{\mathbf{u}} - T\sum_{l=1}^{N}\frac{\omega_l}{1-\rho_0\omega_l}, \quad (62)$$

$$\frac{\partial \ln \mathcal{L}}{\partial \rho_2} = \frac{1}{\sigma_u^2}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]' \boldsymbol{\Phi}_{NT}'(\mathbf{P} \otimes \mathbf{W}_N)'\Delta\tilde{\mathbf{u}} - T\sum_{l=1}^{N}\frac{\omega_l}{1-\rho_2\omega_l}, \quad (63)$$

$$\frac{\partial \ln \mathcal{L}}{\partial \phi_k} = -\frac{1}{\sigma_u^2}(\mathbf{S}_N\Delta\mathbf{y}_1 - \Delta\tilde{\mathbf{Z}}\tilde{\boldsymbol{\theta}})'\mathbf{R}_N'(\mathbf{p}_1 \otimes \mathbf{W}_N^k)'\Delta\tilde{\mathbf{u}} + \sum_{l=1}^{N}\frac{\omega_l^k}{1+\sum_{h=1}^{K_N}\phi_h\omega_l^h}, \quad k = 1, 2, \ldots, K_N, \quad (64)$$

$$\frac{\partial \ln \mathcal{L}}{\partial \tau} = -\frac{NT}{2[1+T(\tau-1)]} + \frac{1}{2\sigma_u^2[1+T(\tau-1)]^2}\Delta\mathbf{u}'\left[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N\right]\Delta\mathbf{u}, \quad (65)$$

where $\boldsymbol{\vartheta} = (T, T-1, \ldots, 1)'$. Setting the latter equal to zero yields

$$\hat{\tau} = \frac{T-1}{T} + \frac{1}{\hat{\sigma}_u^2 NT^2}\left[(\mathbf{I}_T \otimes \hat{\mathbf{S}}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}^*\hat{\boldsymbol{\theta}}^*\right]' \hat{\boldsymbol{\Phi}}_{NT}' \\ (\mathbf{I}_T \otimes \hat{\mathbf{R}}_N)' \left[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N\right] (\mathbf{I}_T \otimes \hat{\mathbf{R}}_N)\hat{\boldsymbol{\Phi}}_{NT}\left[(\mathbf{I}_T \otimes \hat{\mathbf{S}}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}^*\hat{\boldsymbol{\theta}}^*\right], \quad (66)$$

which implies a lower bound for τ at $(T-1)/T$.

The second-order derivatives are:

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = -\frac{1}{\sigma_u^2}\mathbf{G}'\mathbf{G}, \quad \frac{\partial^2 \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}}\partial \tilde{\boldsymbol{\theta}}'} = -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}'\tilde{\mathbf{G}}, \quad \frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}\partial \tilde{\boldsymbol{\theta}}'} = -\frac{1}{\sigma_u^2}\mathbf{G}'\tilde{\mathbf{G}},$$
$$\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}\partial \sigma_u^2} = -\frac{1}{\sigma_u^4}\mathbf{G}'\Delta\tilde{\mathbf{u}}, \quad \frac{\partial^2 \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}}\partial \sigma_u^2} = -\frac{1}{\sigma_u^4}\tilde{\mathbf{G}}'\Delta\tilde{\mathbf{u}}, \quad \frac{\partial^2 \ln \mathcal{L}}{\partial (\sigma_u^2)^2} = \frac{NT}{2\sigma_u^4} - \frac{1}{\sigma_u^6}\Delta\tilde{\mathbf{u}}'\Delta\tilde{\mathbf{u}},$$

and

$$
\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_0^2} =& -\frac{1}{\sigma_u^2}(\mathbf{\Phi}_{NT}\Delta\mathbf{y})'(\mathbf{I}_T\otimes\mathbf{R}_N)'\left[\Omega^{*-1}\otimes(\mathbf{W}_N'\mathbf{W}_N)\right](\mathbf{I}_T\otimes\mathbf{R}_N)\mathbf{\Phi}_{NT}\Delta\mathbf{y}\\
&-T\sum_{l=1}^{N}\left(\frac{\omega_l}{1-\rho_0\omega_l}\right)^2,\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_2^2} =& -\frac{1}{\sigma_u^2}\left[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y}-\Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]'\mathbf{\Phi}_{NT}'\left[\Omega^{*-1}\otimes(\mathbf{W}_N'\mathbf{W}_N)\right]\\
&\mathbf{\Phi}_{NT}\left[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y}-\Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]-T\sum_{l=1}^{N}\left(\frac{\omega_l}{1-\rho_2\omega_l}\right)^2,\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \phi_{k_1}\partial\phi_{k_2}} =& -\frac{1}{\sigma_u^2}(\mathbf{S}_N\Delta\mathbf{y}_1-\Delta\tilde{\mathbf{Z}}\tilde{\boldsymbol{\theta}})'\mathbf{R}_N'\left(\mathbf{p}_1\otimes\mathbf{W}_N^{k_1}\right)'\left(\mathbf{p}_1\otimes\mathbf{W}_N^{k_2}\right)\mathbf{R}_N(\mathbf{S}_N\Delta\mathbf{y}_1-\Delta\tilde{\mathbf{Z}}\tilde{\boldsymbol{\theta}})\\
&-\sum_{l=1}^{N}\frac{\omega_l^{k_1+k_2}}{(1+\sum_{h=1}^{K_N}\phi_h\omega_l^h)^2},\quad k_1,k_2=1,2,\ldots,K_N,\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \tau^2} =& \frac{NT^2}{2[1+T(\tau-1)]^2}-\frac{T}{\sigma_u^2[1+T(\tau-1)]^3}\Delta\mathbf{u}'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}')\otimes\mathbf{I}_N]\Delta\mathbf{u},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}\partial\rho_0} =& -\frac{1}{\sigma_u^2}\mathbf{G}'(\mathbf{P}\otimes\mathbf{W}_N)(\mathbf{I}_T\otimes\mathbf{R}_N)\mathbf{\Phi}_{NT}\Delta\mathbf{y},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}}\partial\rho_0} =& -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}'(\mathbf{P}\otimes\mathbf{W}_N)(\mathbf{I}_T\otimes\mathbf{R}_N)\mathbf{\Phi}_{NT}\Delta\mathbf{y},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}\partial\rho_2} =& -\frac{1}{\sigma_u^2}\mathbf{G}'(\mathbf{P}\otimes\mathbf{W}_N)\mathbf{\Phi}_{NT}\left[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y}-\Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]-\frac{1}{\sigma_u^2}\Delta\mathbf{Z}'(\mathbf{P}_{2..T}\otimes\mathbf{W}_N)'\Delta\tilde{\mathbf{u}},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}}\partial\rho_2} =& -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}'(\mathbf{P}\otimes\mathbf{W}_N)\mathbf{\Phi}_{NT}\left[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y}-\Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]-\frac{1}{\sigma_u^2}[\phi(\mathbf{W}_N)\Delta\tilde{\mathbf{Z}}]'(\mathbf{p}_1\otimes\mathbf{W}_N)'\Delta\tilde{\mathbf{u}},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \sigma_u^2\partial\rho_0} =& -\frac{1}{\sigma_u^4}(\mathbf{\Phi}_{NT}\Delta\mathbf{y})'(\mathbf{I}_T\otimes\mathbf{R}_N)'(\mathbf{P}\otimes\mathbf{W}_N)'\Delta\tilde{\mathbf{u}},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \sigma_u^2\partial\rho_2} =& -\frac{1}{\sigma_u^4}\left[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y}-\Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]'\mathbf{\Phi}_{NT}'(\mathbf{P}\otimes\mathbf{W}_N)'\Delta\tilde{\mathbf{u}},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}\partial\tau} =& -\frac{1}{\sigma_u^2[1+T(\tau-1)]^2}\Delta\mathbf{Z}'\left[(\tilde{\boldsymbol{\vartheta}}\boldsymbol{\vartheta}')\otimes\mathbf{I}_N\right]\Delta\mathbf{u},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}}\partial\tau} =& -\frac{T}{\sigma_u^2[1+T(\tau-1)]^2}\Delta\tilde{\mathbf{Z}}'(\boldsymbol{\vartheta}'\otimes\mathbf{I}_N)\Delta\mathbf{u},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \sigma_u^2\partial\tau} =& -\frac{1}{2\sigma_u^4[1+T(\tau-1)]^2}\Delta\mathbf{u}'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}')\otimes\mathbf{I}_N]\Delta\mathbf{u},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_0\partial\tau} =& -\frac{1}{\sigma_u^2[1+T(\tau-1)]^2}(\mathbf{\Phi}_{NT}\Delta\mathbf{y})'(\mathbf{I}_T\otimes\mathbf{R}_N)'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}')\otimes\mathbf{W}_N']\Delta\mathbf{u},\\
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_2\partial\tau} =& -\frac{1}{\sigma_u^2[1+T(\tau-1)]^2}\left[(\mathbf{I}_T\otimes\mathbf{S}_N)\Delta\mathbf{y}-\Delta\mathbf{Z}^*\boldsymbol{\theta}^*\right]'\mathbf{\Phi}_{NT}'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}')\otimes\mathbf{W}_N']\Delta\mathbf{u},
\end{aligned}
$$

where $\tilde{\boldsymbol{\vartheta}} = (T-1, T-2, \dots, 1)'$. Finally,

$$
\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta} \partial \phi_k} =& \frac{1}{\sigma_u^2} \mathbf{G}'(\mathbf{p}_1 \otimes \mathbf{W}_N^k)\mathbf{R}_N(\mathbf{S}_N \Delta \mathbf{y}_1 - \Delta \tilde{\mathbf{Z}} \tilde{\boldsymbol{\theta}}), \\
\frac{\partial^2 \ln \mathcal{L}}{\partial \tilde{\boldsymbol{\theta}} \partial \phi_k} =& \frac{1}{\sigma_u^2} \tilde{\mathbf{G}}'(\mathbf{p}_1 \otimes \mathbf{W}_N^k)\mathbf{R}_N(\mathbf{S}_N \Delta \mathbf{y}_1 - \Delta \tilde{\mathbf{Z}} \tilde{\boldsymbol{\theta}}) + \frac{1}{\sigma_u^2}(\mathbf{R}_N \Delta \tilde{\mathbf{Z}})'(\mathbf{p}_1 \otimes \mathbf{W}_N^k)' \Delta \tilde{\mathbf{u}}, \\
\frac{\partial^2 \ln \mathcal{L}}{\partial \sigma_u^2 \partial \phi_k} =& \frac{1}{\sigma_u^4} (\mathbf{S}_N \Delta \mathbf{y}_1 - \Delta \tilde{\mathbf{Z}} \tilde{\boldsymbol{\theta}})' \mathbf{R}_N' (\mathbf{p}_1 \otimes \mathbf{W}_N^k)' \Delta \tilde{\mathbf{u}}, \\
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_0 \partial \phi_k} =& \frac{1}{\sigma_u^2} (\boldsymbol{\Phi}_{NT} \Delta \mathbf{y})'(\mathbf{I}_T \otimes \mathbf{R}_N)'(\mathbf{P} \otimes \mathbf{W}_N)'(\mathbf{p}_1 \otimes \mathbf{W}_N^k)\mathbf{R}_N(\mathbf{S}_N \Delta \mathbf{y}_1 - \Delta \tilde{\mathbf{Z}} \tilde{\boldsymbol{\theta}}) \\
&+ \frac{1}{\sigma_u^2}(\mathbf{R}_N \Delta \mathbf{y}_1)'(\mathbf{p}_1 \otimes \mathbf{W}_N^{k+1})' \Delta \tilde{\mathbf{u}}, \\
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_2 \partial \phi_k} =& \frac{1}{\sigma_u^2} \left[(\mathbf{I}_T \otimes \mathbf{S}_N) \Delta \mathbf{y} - \Delta \mathbf{Z}^* \boldsymbol{\theta}^* \right]' \boldsymbol{\Phi}_{NT}' (\mathbf{P} \otimes \mathbf{W}_N)'(\mathbf{p}_1 \otimes \mathbf{W}_N^k)\mathbf{R}_N(\mathbf{S}_N \Delta \mathbf{y}_1 - \Delta \tilde{\mathbf{Z}} \tilde{\boldsymbol{\theta}}) \\
&+ \frac{1}{\sigma_u^2}(\mathbf{S}_N \Delta \mathbf{y}_1 - \Delta \tilde{\mathbf{Z}} \tilde{\boldsymbol{\theta}})'(\mathbf{p}_1 \otimes \mathbf{W}_N^{k+1})' \Delta \tilde{\mathbf{u}}, \\
\frac{\partial^2 \ln \mathcal{L}}{\partial \phi_k \partial \tau} =& \frac{T}{\sigma_u^2 [1 + T(\tau - 1)]^2} (\mathbf{R}_N \mathbf{S}_N \Delta \mathbf{y}_1)'(\boldsymbol{\vartheta} \otimes \mathbf{W}_N^k)' \Delta \mathbf{u},
\end{aligned}
$$

for all $k = 1, 2, \dots, K_N$, and

$$
\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_0 \partial \rho_2} =& - \frac{1}{\sigma_u^2} (\boldsymbol{\Phi}_{NT} \Delta \mathbf{y})'(\mathbf{I}_T \otimes \mathbf{R}_N)' \left[\Omega^{*-1} \otimes (\mathbf{W}_N' \mathbf{W}_N) \right] \boldsymbol{\Phi}_{NT} \left[(\mathbf{I}_T \otimes \mathbf{S}_N) \Delta \mathbf{y} - \Delta \mathbf{Z}^* \boldsymbol{\theta}^* \right] \\
&- \frac{1}{\sigma_u^2} (\boldsymbol{\Phi}_{NT} \Delta \mathbf{y})'(\mathbf{P} \otimes \mathbf{W}_N^2)' \Delta \tilde{\mathbf{u}}.
\end{aligned}
$$

A.2 Restricted time-space dynamic model

Again, I make use of the decomposition for Ω^{*-1} to obtain

$$
\Delta \tilde{\mathbf{u}}_r = (\mathbf{P} \otimes \mathbf{I}_N) \Delta \mathbf{u}_r = [\mathbf{P} \otimes (\mathbf{R}_N \mathbf{S}_N)] \Delta \mathbf{y} - \tilde{\mathbf{G}}_r \tilde{\boldsymbol{\theta}}_r - \mathbf{G} \mathbf{H} \boldsymbol{\theta}_r - \psi (\mathbf{p}_1 \otimes \boldsymbol{\iota}_N), \tag{67}
$$

where $\tilde{\mathbf{G}}_r = (\mathbf{p}_1 \otimes \mathbf{R}_N) \Delta \tilde{\mathbf{Z}}_r$, and $\mathbf{G}$ is defined as before. The first-order derivatives become

$$
\frac{\partial \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r} = \frac{1}{\sigma_u^2} \mathbf{H}' \mathbf{G}' \Delta \tilde{\mathbf{u}}_r, \tag{68}
$$

$$
\frac{\partial \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r} = \frac{1}{\sigma_u^2} \tilde{\mathbf{G}}_r' \Delta \tilde{\mathbf{u}}_r, \tag{69}
$$

$$
\frac{\partial \ln \mathcal{L}_r}{\partial \sigma_u^2} = - \frac{NT}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \Delta \tilde{\mathbf{u}}_r' \Delta \tilde{\mathbf{u}}_r, \tag{70}
$$

and

$$\frac{\partial \ln \mathcal{L}_r}{\partial \psi} = \frac{1}{\sigma_u^2}(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N)'\Delta\tilde{\mathbf{u}}_r. \tag{71}$$

Similar to the unrestricted model, we can obtain closed-form solutions for these parameters for given values of ρ_0, κ, and τ that can be used to set up a concentrated log-likelihood function. The first-order derivatives for the remaining parameters are given by

$$\frac{\partial \ln \mathcal{L}_r}{\partial \rho_0} = \frac{1}{\sigma_u^2}\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho \boldsymbol{\theta}_r\right]' \Delta\tilde{\mathbf{u}}_r - T\sum_{l=1}^{N} \frac{\omega_l}{1-\rho_0\omega_l}, \tag{72}$$

$$\frac{\partial \ln \mathcal{L}_r}{\partial \rho_2} = \frac{1}{\sigma_u^2}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right]'(\mathbf{P} \otimes \mathbf{W}_N)'\Delta\tilde{\mathbf{u}}_r - T\sum_{l=1}^{N} \frac{\omega_l}{1-\rho_2\omega_l}, \tag{73}$$

$$\frac{\partial \ln \mathcal{L}_r}{\partial \kappa} = \frac{1}{\sigma_u^2}\boldsymbol{\theta}_r'\dot{\mathbf{H}}_\kappa'\mathbf{G}'\Delta\tilde{\mathbf{u}}_r, \tag{74}$$

where

$$\dot{\mathbf{H}}_\rho = \begin{pmatrix} 0 & \mathbf{0} \\ -1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \dot{\mathbf{H}}_\kappa = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix},$$

and

$$\frac{\partial \ln \mathcal{L}_r}{\partial \tau} = -\frac{NT}{2[1+T(\tau-1)]} + \frac{1}{2\sigma_u^2[1+T(\tau-1)]^2}\Delta\mathbf{u}_r'\left[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N\right]\Delta\mathbf{u}_r. \tag{75}$$

The second-order derivatives become

$$\frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_r'} = -\frac{1}{\sigma_u^2}\mathbf{H}'\mathbf{G}'\mathbf{G}\mathbf{H}, \quad \frac{\partial^2 \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r \partial \tilde{\boldsymbol{\theta}}_r'} = -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}_r'\tilde{\mathbf{G}}_r, \quad \frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \tilde{\boldsymbol{\theta}}_r'} = -\frac{1}{\sigma_u^2}\mathbf{H}'\mathbf{G}'\tilde{\mathbf{G}}_r,$$

$$\frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \sigma_u^2} = -\frac{1}{\sigma_u^4}\mathbf{H}'\mathbf{G}'\Delta\tilde{\mathbf{u}}_r, \quad \frac{\partial^2 \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r \partial \sigma_u^2} = -\frac{1}{\sigma_u^4}\tilde{\mathbf{G}}_r'\Delta\tilde{\mathbf{u}}_r, \quad \frac{\partial^2 \ln \mathcal{L}_r}{\partial (\sigma_u^2)^2} = \frac{NT}{2\sigma_u^4} - \frac{1}{\sigma_u^6}\Delta\tilde{\mathbf{u}}_r'\Delta\tilde{\mathbf{u}}_r,$$

$$\frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \psi} = -\frac{1}{\sigma_u^2}\mathbf{H}'\mathbf{G}'(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N), \quad \frac{\partial^2 \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r \partial \psi} = -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}_r'(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N),$$

$$\frac{\partial^2 \ln \mathcal{L}_r}{\partial \sigma_u^2 \partial \psi} = -\frac{1}{\sigma_u^4}\Delta\tilde{\mathbf{u}}_r'(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N), \quad \frac{\partial^2 \ln \mathcal{L}_r}{\partial \psi^2} = -\frac{1}{\sigma_u^2}(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N)'(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N),$$

$$\frac{\partial^2 \ln \mathcal{L}_r}{\partial \rho_0^2} = -\frac{1}{\sigma_u^2}\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right]'\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right]$$
$$- T\sum_{l=1}^{N}\left(\frac{\omega_l}{1-\rho_0\omega_l}\right)^2,$$

and

$$
\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \rho_2^2} =& -\frac{1}{\sigma_u^2}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right]'\left[\Omega^{*-1} \otimes (\mathbf{W}_N'\mathbf{W}_N)\right]\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right] \\
& - T\sum_{l=1}^{N}\left(\frac{\omega_l}{1-\rho_2\omega_l}\right)^2, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \tau^2} =& \frac{NT^2}{2[1+T(\tau-1)]^2} - \frac{T}{\sigma_u^2[1+T(\tau-1)]^3}\Delta\mathbf{u}_r'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N]\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \rho_0} =& -\frac{1}{\sigma_u^2}\mathbf{H}'\mathbf{G}'\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right] - \frac{1}{\sigma_u^2}\dot{\mathbf{H}}_\rho'\mathbf{G}'\Delta\tilde{\mathbf{u}}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r \partial \rho_0} =& -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}_r'\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right], \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \rho_2} =& -\frac{1}{\sigma_u^2}\mathbf{H}'\mathbf{G}'(\mathbf{P} \otimes \mathbf{W}_N)\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right] - \frac{1}{\sigma_u^2}\Delta\mathbf{Z}'(\mathbf{P}_{2..T} \otimes \mathbf{W}_N)'\Delta\tilde{\mathbf{u}}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r \partial \rho_2} =& -\frac{1}{\sigma_u^2}\tilde{\mathbf{G}}_r'(\mathbf{P} \otimes \mathbf{W}_N)\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right] - \frac{1}{\sigma_u^2}\Delta\tilde{\mathbf{Z}}_r'(\mathbf{p}_1 \otimes \mathbf{W}_N)'\Delta\tilde{\mathbf{u}}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \sigma_u^2 \partial \rho_0} =& -\frac{1}{\sigma_u^4}\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right]'\Delta\tilde{\mathbf{u}}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \sigma_u^2 \partial \rho_2} =& -\frac{1}{\sigma_u^4}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right]'(\mathbf{P} \otimes \mathbf{W}_N)'\Delta\tilde{\mathbf{u}}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \psi \partial \rho_0} =& -\frac{1}{\sigma_u^2}(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N)'\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right], \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \psi \partial \rho_2} =& -\frac{1}{\sigma_u^2}(\mathbf{p}_1 \otimes \boldsymbol{\iota}_N)'(\mathbf{P} \otimes \mathbf{W}_N)\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right], \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \tau} =& -\frac{1}{\sigma_u^2[1+T(\tau-1)]^2}\mathbf{H}'\Delta\mathbf{Z}'(\mathbf{I}_{T-1} \otimes \mathbf{R}_N)'\left[(\tilde{\boldsymbol{\vartheta}}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N\right]\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \tilde{\boldsymbol{\theta}}_r \partial \tau} =& -\frac{T}{\sigma_u^2[1+T(\tau-1)]^2}\Delta\tilde{\mathbf{Z}}_r'\mathbf{R}_N'(\boldsymbol{\vartheta}' \otimes \mathbf{I}_N)\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \sigma_u^2 \partial \tau} =& -\frac{1}{2\sigma_u^4[1+T(\tau-1)]^2}\Delta\mathbf{u}_r^{*\prime}[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N]\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \psi \partial \tau} =& -\frac{T}{\sigma_u^2[1+T(\tau-1)]^2}\boldsymbol{\iota}_N'(\boldsymbol{\vartheta}' \otimes \mathbf{I}_N)\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \rho_0 \partial \tau} =& -\frac{1}{\sigma_u^2[1+T(\tau-1)]^2}\Delta\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{R}_N)'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{W}_N']\Delta\mathbf{u}_r \\
& - \frac{1}{\sigma_u^2[1+T(\tau-1)]^2}(\Delta\mathbf{Z}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r)'(\mathbf{I}_{T-1} \otimes \mathbf{R}_N)'\left[(\tilde{\boldsymbol{\vartheta}}\boldsymbol{\vartheta}') \otimes \mathbf{I}_N\right]\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}_r}{\partial \rho_2 \partial \tau} =& -\frac{1}{\sigma_u^2[1+T(\tau-1)]^2}\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right]'[(\boldsymbol{\vartheta}\boldsymbol{\vartheta}') \otimes \mathbf{W}_N']\Delta\mathbf{u}_r, \\
\frac{\partial^2 \ln \mathcal{L}}{\partial \rho_0 \partial \rho_2} =& -\frac{1}{\sigma_u^2}\left[(\mathbf{P} \otimes \mathbf{W}_N)(\mathbf{I}_T \otimes \mathbf{R}_N)\Delta\mathbf{y} + \mathbf{G}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right]'(\mathbf{P} \otimes \mathbf{W}_N)\left[(\mathbf{I}_T \otimes \mathbf{S}_N)\Delta\mathbf{y} - \Delta\mathbf{Z}_r^*\boldsymbol{\theta}_r^*\right] \\
& - \frac{1}{\sigma_u^2}\left[(\mathbf{P} \otimes \mathbf{W}_N^2)\Delta\mathbf{y} + (\mathbf{P}_{2..T} \otimes \mathbf{W}_N)\Delta\mathbf{Z}\dot{\mathbf{H}}_\rho\boldsymbol{\theta}_r\right]'\Delta\tilde{\mathbf{u}}_r.
\end{aligned}
$$

B Conditional transformed likelihood estimation

Modeling the initial observations aims at obtaining unbiased estimators in dynamic panel data models. As an alternative, Yu et al. (2008) derive a bias reduction for the transformed likelihood estimator of time-space dynamic panel data models conditional on $\Delta \mathbf{y}_1$. The conditional log-likelihood function for the first-differenced model, for simplicity assuming $\rho_2 = 0$, is given by

$$\ln \mathcal{L}_c = -\frac{N(T-1)}{2} \ln(2\pi\sigma_u^2) - \frac{N}{2}\ln(T) - \frac{1}{2}\Delta \mathbf{u}_c'(\Omega_c^{-1} \otimes \mathbf{I}_N)\Delta \mathbf{u}_c + (T-1)\sum_{l=1}^{N} \ln(1-\rho_0\omega_l), \quad (76)$$

where $\Delta \mathbf{u}_c = (\Delta \mathbf{u}_2, \Delta \mathbf{u}_3, \ldots, \Delta \mathbf{u}_T)$ and

$$\Omega_c = \sigma_u^2 \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix}. \quad (77)$$

Let $\hat{\boldsymbol{\theta}}_c = (\hat{\rho}_0, \hat{\lambda}, \hat{\rho}_1, \hat{\beta}_0, \hat{\gamma}_0, \hat{\beta}_1, \hat{\gamma}_1, \sigma_u^2)'$ be the corresponding conditional quasi-maximum likelihood estimator. The bias-corrected estimator of Yu et al. (2008) can then be obtained as

$$\hat{\boldsymbol{\theta}}_{bc} = \hat{\boldsymbol{\theta}}_c + \frac{1}{T}\hat{\Sigma}_\theta^{-1}\xi(\hat{\boldsymbol{\theta}}_c), \quad (78)$$

where $\hat{\Sigma}_\theta$ is the inverse negative Hessian matrix evaluated at $\hat{\boldsymbol{\theta}}_c$, and

$$\xi(\hat{\boldsymbol{\theta}}_c) = \begin{pmatrix} \frac{1}{N}\operatorname{tr}\left(\mathbf{W}_N \mathbf{S}_N^{-1}(\hat{\lambda}\tilde{\mathbf{S}}_N^{-1} + \hat{\rho}_1 \mathbf{W}_N \tilde{\mathbf{S}}_N^{-1} + \mathbf{I}_N)\right) \\ \frac{1}{N}\operatorname{tr}\left(\tilde{\mathbf{S}}_N^{-1}\right) \\ \frac{1}{N}\operatorname{tr}\left(\mathbf{W}_N\tilde{\mathbf{S}}_N^{-1}\right) \\ \mathbf{0} \\ \frac{1}{2\hat{\sigma}_u^2} \end{pmatrix}, \quad (79)$$

where $\tilde{\mathbf{S}}_N = [(1-\lambda)\mathbf{I}_N - (\rho_0+\rho_1)\mathbf{W}_N]$. In the case of the spatial weights matrix $\mathbf{W}_N = \mathbf{I}_R \otimes \mathbf{B}_M$ introduced in Section 3.2, the inverses of $\mathbf{S}_N$ and $\tilde{\mathbf{S}}_N$ can be easily obtained with formula (11).

C Detailed data description

Table 8 lists the variables extracted from the PSID for the years 2001 to 2011. The interview number is used to link cohabiting workers to each other, and the relation to head allows to assign labor market outcomes from the family data set to the respective household head, who is always male in the PSID, or "wife" (legal wife or female cohabitant). Because the estimation methodology relies on constant spatial weights over time, I have to dismiss all individuals whose partner in the same family unit changed during the analyzed time horizon.

[Table 8 about here.]

The dependent variable in the regression analysis is the natural logarithm of the hourly wage rate for the current main job. Only individuals that continuously earned at least the federal minimum wage in each year are kept in the data set. The respective minimum wage was 5.15 USD in the years 2001 to 2009, and 7.25 USD in 2011.[42]

The age of the individuals is constructed as the current year minus the year of birth. Some apparently miscoded data points have been corrected to ensure that age is linearly growing with time. The main industry for the current job based on the 3-digit industry code from the 2000 census of population and housing is used to create three industry dummy variables: primary sector (codes 17–77), manufacturing (codes 107-399), and public administration (codes 937–987). The reference group are thus the remaining services industries. Regional dummy variables are created for the northeast, north central, and southern part of the United States. The western part serves as the reference group.

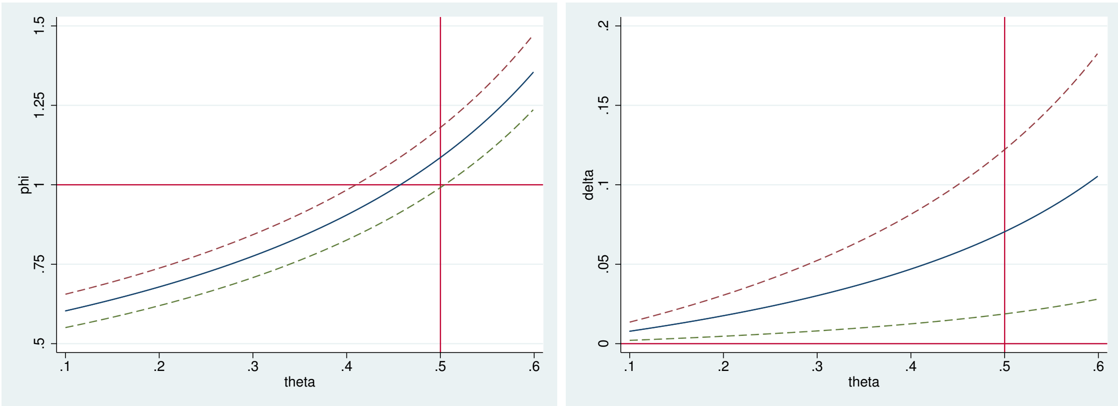
The set of independent variables further includes a dummy variable for union coverage of the current job, the marital status, and the number of children under the age of 18 living in the same family unit. Summary statistics for the variables used in the estimation of the time-space dynamic wage equation are presented in Table 9. The number of data points is 4,428 that form a balanced panel with 738 individuals observed at 6 time points.

[Table 9 about here.]

[42] The wage data for some individuals showed extreme jumps potentially due to miscoding. To avoid distortions due to such outliers, I dropped all individuals whose standard deviation of the hourly wage rate exceeded 30.

Figures

Figure 1: Implied parameter estimates for the wage bargaining model



Tables

Table 1: Simulation results: coefficients ρ_{y0} and ρ_{y1}

	ρ_{y0}	R	10		20		50	
M	Estimator	K_N	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0106	0.0533	-0.0036	0.0346	-0.0012	0.0213
	unrestricted	(0, 0)	-0.0132	0.0524	-0.0056	0.0338	-0.0019	0.0211
	unrestricted	(1, 1)			-0.0028	0.0357	-0.0004	0.0216
	bias-corrected		-0.0007	0.0524	-0.0006	0.0351	0.0002	0.0223
5	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0071	0.0547	-0.0045	0.0373	-0.0009	0.0237
	unrestricted	(0, 0)	-0.0121	0.0539	-0.0061	0.0368	-0.0016	0.0234
	unrestricted	(1, 1)	-0.0678	0.0907	-0.0131	0.0403	-0.0033	0.0244
	bias-corrected		-0.0063	0.0559	-0.0042	0.0386	-0.0000	0.0245
10	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0085	0.0578	-0.0053	0.0390	-0.0016	0.0245
	unrestricted	(0, 0)	-0.0131	0.0572	-0.0069	0.0388	-0.0021	0.0239
	unrestricted	(1, 1)	-0.0637	0.0893	-0.0146	0.0422	-0.0052	0.0251
	bias-corrected		-0.0101	0.0598	-0.0045	0.0404	-0.0013	0.0250
20	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0095	0.0579	-0.0047	0.0389	-0.0012	0.0255
	unrestricted	(0, 0)	-0.0130	0.0584	-0.0063	0.0377	-0.0013	0.0248
	unrestricted	(1, 1)	-0.0600	0.0867	-0.0149	0.0430	-0.0050	0.0265
	bias-corrected		-0.0093	0.0598	-0.0049	0.0414	-0.0009	0.0266
50	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0073	0.0566	-0.0034	0.0392	-0.0009	0.0261
	unrestricted	(0, 0)	-0.0105	0.0561	-0.0051	0.0382	-0.0013	0.0254
	unrestricted	(1, 1)	-0.0564	0.0843	-0.0136	0.0434	-0.0048	0.0272
	bias-corrected		-0.0077	0.0595	-0.0040	0.0413	-0.0007	0.0273

	ρ_{y1}	R	10		20		50	
M	Estimator	K_N	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0042	0.0292	0.0012	0.0180	0.0007	0.0108
	unrestricted	(0, 0)	-0.0204	0.0824	-0.0063	0.0493	-0.0015	0.0313
	unrestricted	(1, 1)			-0.0026	0.0522	0.0012	0.0320
	bias-corrected		0.0054	0.0778	0.0038	0.0533	0.0028	0.0348
5	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0033	0.0240	0.0020	0.0164	0.0004	0.0104
	unrestricted	(0, 0)	-0.0116	0.0733	-0.0040	0.0485	-0.0007	0.0305
	unrestricted	(1, 1)	-0.1286	0.1548	-0.0254	0.0571	-0.0065	0.0323
	bias-corrected		0.0007	0.0794	0.0005	0.0540	0.0031	0.0354
10	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0036	0.0242	0.0021	0.0161	0.0007	0.0103
	unrestricted	(0, 0)	-0.0077	0.0752	-0.0016	0.0509	-0.0001	0.0313
	unrestricted	(1, 1)	-0.1090	0.1360	-0.0264	0.0592	-0.0084	0.0334
	bias-corrected		0.0001	0.0819	0.0038	0.0581	0.0024	0.0351
20	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0038	0.0236	0.0019	0.0160	0.0005	0.0105
	unrestricted	(0, 0)	-0.0064	0.0736	-0.0005	0.0515	-0.0011	0.0312
	unrestricted	(1, 1)	-0.0983	0.1272	-0.0281	0.0605	-0.0106	0.0339
	bias-corrected		0.0021	0.0834	0.0036	0.0575	0.0006	0.0352
50	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0030	0.0228	0.0014	0.0159	0.0004	0.0105
	unrestricted	(0, 0)	-0.0085	0.0722	-0.0036	0.0493	0.0009	0.0330
	unrestricted	(1, 1)	-0.0933	0.1216	-0.0318	0.0600	-0.0087	0.0346
	bias-corrected		-0.0008	0.0817	0.0000	0.0564	0.0031	0.0363

Note: The first three estimators are described in Sections 4 and 5. The first number of the truncation order K_N refers to the polynomials $\pi_l(\mathbf{W}_N)$ and the second number to $\phi(\mathbf{W}_N)$. The fourth estimator is the bias-corrected conditional transformed likelihood estimator of Yu et al. (2008).

Table 2: Simulation results: coefficients λ_y and β

	λ_y	R	10		20		50	
M	Estimator	K_N	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0000	0.0907	0.0014	0.0594	-0.0011	0.0336
	unrestricted	(0, 0)	0.0080	0.0930	0.0024	0.0616	-0.0012	0.0354
	unrestricted	(1, 1)			0.0025	0.0610	-0.0019	0.0353
	bias-corrected		-0.0099	0.0784	-0.0038	0.0564	-0.0035	0.0351
5	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0021	0.0491	-0.0009	0.0340	-0.0003	0.0211
	unrestricted	(0, 0)	-0.0017	0.0495	-0.0013	0.0345	-0.0006	0.0214
	unrestricted	(1, 1)	0.0508	0.0793	0.0031	0.0349	0.0004	0.0214
	bias-corrected		-0.0060	0.0487	-0.0030	0.0340	-0.0019	0.0212
10	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0008	0.0342	0.0002	0.0241	-0.0001	0.0152
	unrestricted	(0, 0)	-0.0014	0.0345	-0.0002	0.0242	-0.0003	0.0153
	unrestricted	(1, 1)	0.0322	0.0485	0.0024	0.0245	0.0004	0.0154
	bias-corrected		-0.0031	0.0340	-0.0015	0.0240	-0.0014	0.0152
20	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0002	0.0242	-0.0003	0.0173	0.0000	0.0112
	unrestricted	(0, 0)	-0.0001	0.0242	-0.0005	0.0172	-0.0000	0.0112
	unrestricted	(1, 1)	0.0237	0.0343	0.0009	0.0173	0.0003	0.0112
	bias-corrected		-0.0015	0.0239	-0.0017	0.0169	-0.0010	0.0111
50	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0000	0.0153	0.0000	0.0112	-0.0000	0.0071
	unrestricted	(0, 0)	-0.0002	0.0153	-0.0000	0.0112	-0.0001	0.0071
	unrestricted	(1, 1)	0.0158	0.0221	0.0006	0.0112	0.0000	0.0071
	bias-corrected		-0.0013	0.0151	-0.0010	0.0111	-0.0011	0.0071

	β	R	10		20		50	
M	Estimator	K_N	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0045	0.0795	0.0015	0.0550	0.0000	0.0331
	unrestricted	(0, 0)	0.0063	0.0773	0.0025	0.0553	0.0005	0.0329
	unrestricted	(1, 1)			0.0016	0.0550	-0.0001	0.0329
	bias-corrected		0.0025	0.0767	0.0010	0.0547	-0.0002	0.0329
5	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0026	0.0496	-0.0000	0.0339	0.0003	0.0218
	unrestricted	(0, 0)	-0.0017	0.0492	0.0004	0.0338	0.0005	0.0219
	unrestricted	(1, 1)	0.0015	0.0498	0.0011	0.0338	0.0007	0.0219
	bias-corrected		-0.0023	0.0491	0.0002	0.0337	0.0004	0.0218
10	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0002	0.0340	0.0009	0.0250	-0.0005	0.0160
	unrestricted	(0, 0)	0.0005	0.0339	0.0012	0.0251	-0.0004	0.0160
	unrestricted	(1, 1)	0.0009	0.0342	0.0016	0.0251	-0.0002	0.0160
	bias-corrected		0.0004	0.0339	0.0011	0.0251	-0.0004	0.0159
20	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0010	0.0251	0.0002	0.0166	-0.0001	0.0112
	unrestricted	(0, 0)	0.0012	0.0250	0.0003	0.0166	-0.0000	0.0112
	unrestricted	(1, 1)	0.0005	0.0251	0.0004	0.0167	0.0001	0.0112
	bias-corrected		0.0012	0.0250	0.0003	0.0166	0.0000	0.0111
50	$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	-0.0005	0.0161	-0.0001	0.0112	-0.0002	0.0070
	unrestricted	(0, 0)	-0.0004	0.0161	-0.0000	0.0112	-0.0002	0.0070
	unrestricted	(1, 1)	-0.0012	0.0161	0.0001	0.0112	-0.0002	0.0070
	bias-corrected		-0.0003	0.0160	0.0001	0.0112	-0.0001	0.0070

Note: The first three estimators are described in Sections 4 and 5. The first number of the truncation order K_N refers to the polynomials $\pi_l(\mathbf{W}_N)$ and the second number to $\phi(\mathbf{W}_N)$. The fourth estimator is the bias-corrected conditional transformed likelihood estimator of Yu et al. (2008).

Table 3: Simulation results: marginal effects

Estimator	K_N	short-run direct Bias	short-run direct RMSE	short-run indirect Bias	short-run indirect RMSE	short-run total Bias	short-run total RMSE
$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0000	0.0339	-0.0006	0.0123	-0.0006	0.0415
unrestricted	(0, 0)	0.0003	0.0338	-0.0009	0.0123	-0.0006	0.0413
unrestricted	(1, 1)	0.0001	0.0337	-0.0002	0.0124	-0.0002	0.0413
bias-corrected		0.0000	0.0337	0.0000	0.0127	0.0001	0.0414

Estimator	K_N	long-run direct Bias	long-run direct RMSE	long-run indirect Bias	long-run indirect RMSE	long-run total Bias	long-run total RMSE
$\rho_{y1} = -\lambda_y \rho_{y0}$	(0, 1)	0.0008	0.0702	-0.0009	0.0221	-0.0000	0.0852
unrestricted	(0, 0)	0.0016	0.0705	-0.0053	0.0455	-0.0037	0.0953
unrestricted	(1, 1)	0.0023	0.0705	0.0004	0.0478	0.0027	0.0975
bias-corrected		0.0017	0.0705	0.0029	0.0544	0.0045	0.1017

Note: The first three estimators are described in Sections 4 and 5. The first number of the truncation order K_N refers to the polynomials $\pi_l(\mathbf{W}_N)$ and the second number to $\phi(\mathbf{W}_N)$. The fourth estimator is the bias-corrected conditional transformed likelihood estimator of Yu et al. (2008). The sample size is $R = 50$ and $M = 2$.

Table 4: Estimation results: coefficient estimates

$\ln(\text{wage})_{i,t}$	static	space dynamic	time dynamic	time-space dynamic		
	$\lambda = \rho_0 = \rho_1 = 0$	$\lambda = \rho_1 = 0$	$\rho_0 = \rho_1 = 0$	$\rho_1 = 0$	$\rho_1 = -\lambda\rho_0$	unrestricted
$\ln(\text{wage})_{i,t-1}$			0.5432	0.5413	0.5429	0.5429
			(0.0241)***	(0.0240)***	(0.0241)***	(0.0241)***
$\ln(\text{wage})_{j,t}$		0.0732		0.0503	0.0704	0.0717
		(0.0240)***		(0.0211)**	(0.0264)***	(0.0276)***
$\ln(\text{wage})_{j,t-1}$					-0.0382	-0.0435
					(0.0144)***	(0.0361)
$\text{age}^2_{i,t}$	-0.0002	-0.0002	0.0000	0.0000	0.0000	0.0000
	(0.0000)***	(0.0000)***	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\text{union}_{i,t}$	0.0651	0.0652	0.0778	0.0780	0.0779	0.0778
	(0.0122)***	(0.0122)***	(0.0131)***	(0.0131)***	(0.0131)***	(0.0131)***
$\text{marriage}_{i,t}$	0.0411	0.0418	0.0119	0.0123	0.0124	0.0124
	(0.0222)*	(0.0222)*	(0.0247)	(0.0247)	(0.0247)	(0.0247)
$\text{children}_{i,t}$	0.0070	0.0070	0.0027	0.0027	0.0027	0.0027
	(0.0037)*	(0.0037)*	(0.0044)	(0.0044)	(0.0044)	(0.0044)
log likelihood	1466.07	1470.69	1958.87	1961.70	1962.42	1962.43
likelihood ratio	$\chi^2_{39} = 992.72$***	$\chi^2_{38} = 983.48$***	$\chi^2_{2} = 7.12$**	$\chi^2_{1} = 1.45$	$\chi^2_{1} = 0.03$	

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: Variables are indexed with unit and time period, where j refers to the cohabitant. All regressions include industry, region, and time dummies. The models are estimated with the unconditional transformed QML estimator proposed in this paper. Standard errors are reported in parentheses. The likelihood ratio test statistics are computed as $2(\ln \mathcal{L} - \ln \mathcal{L}_0)$, where $\ln \mathcal{L}$ refers to the log-likelihood value of the unrestricted time-space dynamic model and $\ln \mathcal{L}_0$ to that of the respective restricted model.

Table 5: Estimation results: marginal effects for dual-earners

ln(wage)	static	space dynamic	time dynamic	time-space dynamic		
	$\lambda = \rho_0 = \rho_1 = 0$	$\lambda = \rho_1 = 0$	$\rho_0 = \rho_1 = 0$	$\rho_1 = 0$	$\rho_1 = -\lambda\rho_0$	unrestricted
short-run direct effects						
union	0.0651	0.0656	0.0778	0.0782	0.0782	0.0782
	(0.0122)***	(0.0123)***	(0.0131)***	(0.0132)***	(0.0132)***	(0.0132)***
marriage	0.0411	0.0420	0.0119	0.0124	0.0124	0.0124
	(0.0222)*	(0.0223)*	(0.0247)	(0.0248)	(0.0248)	(0.0248)
children	0.0070	0.0071	0.0027	0.0027	0.0027	0.0027
	(0.0037)*	(0.0038)*	(0.0044)	(0.0044)	(0.0044)	(0.0044)
short-run indirect effects						
union		0.0048		0.0039	0.0055	0.0056
		(0.0018)***		(0.0018)**	(0.0023)**	(0.0024)**
marriage		0.0031		0.0006	0.0009	0.0009
		(0.0019)		(0.0013)	(0.0018)	(0.0018)
children		0.0005		0.0001	0.0002	0.0002
		(0.0003)		(0.0002)	(0.0003)	(0.0003)
short-run total effects						
union	0.0651	0.0704	0.0778	0.0821	0.0838	0.0838
	(0.0122)***	(0.0133)***	(0.0131)***	(0.0140)***	(0.0143)***	(0.0143)***
marriage	0.0411	0.0451	0.0119	0.0130	0.0133	0.0133
	(0.0222)*	(0.0240)*	(0.0247)	(0.0260)	(0.0266)	(0.0266)
children	0.0070	0.0076	0.0027	0.0029	0.0029	0.0029
	(0.0037)*	(0.0040)*	(0.0044)	(0.0046)	(0.0047)	(0.0047)
long-run direct effects						
union	0.0651	0.0656	0.1702	0.1721	0.1712	0.1709
	(0.0122)***	(0.0123)***	(0.0302)***	(0.0305)***	(0.0303)***	(0.0303)***
marriage	0.0411	0.0420	0.0260	0.0272	0.0272	0.0272
	(0.0222)*	(0.0223)*	(0.0542)	(0.0545)	(0.0543)	(0.0543)
children	0.0070	0.0071	0.0059	0.0060	0.0060	0.0060
	(0.0037)*	(0.0038)*	(0.0096)	(0.0097)	(0.0096)	(0.0096)
long-run indirect effects						
union		0.0048		0.0189	0.0121	0.0106
		(0.0018)***		(0.0089)**	(0.0050)**	(0.0107)
marriage		0.0031		0.0030	0.0019	0.0017
		(0.0019)		(0.0061)	(0.0039)	(0.0038)
children		0.0005		0.0007	0.0004	0.0004
		(0.0003)		(0.0011)	(0.0007)	(0.0007)
long-run total effects						
union	0.0651	0.0704	0.1702	0.1910	0.1832	0.1815
	(0.0122)***	(0.0133)***	(0.0302)***	(0.0354)***	(0.0328)***	(0.0343)***
marriage	0.0411	0.0451	0.0260	0.0302	0.0291	0.0289
	(0.0222)*	(0.0240)*	(0.0542)	(0.0605)	(0.0581)	(0.0576)
children	0.0070	0.0076	0.0059	0.0067	0.0064	0.0063
	(0.0037)*	(0.0040)*	(0.0096)	(0.0107)	(0.0103)	(0.0102)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All regressions include industry, region, and time dummies. The models are estimated with the unconditional transformed QML estimator proposed in this paper. Standard errors are computed with the Delta method and reported in parentheses. The formulas for the short-run marginal effects are equations (45) to (47), and for the long-run marginal effects equations (49) to (51).

Table 6: Estimation results: marginal effects for single-earners

ln(wage)	static	space dynamic	time dynamic	time-space dynamic		
	$\lambda = \rho_0 = \rho_1 = 0$	$\lambda = \rho_1 = 0$	$\rho_0 = \rho_1 = 0$	$\rho_1 = 0$	$\rho_1 = -\lambda\rho_0$	unrestricted
short-run effects						
union	0.0651	0.0652	0.0778	0.0780	0.0779	0.0778
	(0.0122)***	(0.0122)***	(0.0131)***	(0.0131)***	(0.0131)***	(0.0131)***
marriage	0.0411	0.0418	0.0119	0.0123	0.0124	0.0124
	(0.0222)*	(0.0222)*	(0.0247)	(0.0247)	(0.0247)	(0.0247)
children	0.0070	0.0070	0.0027	0.0027	0.0027	0.0027
	(0.0037)*	(0.0037)*	(0.0044)	(0.0044)	(0.0044)	(0.0044)
long-run effects						
union	0.0651	0.0652	0.1702	0.1701	0.1703	0.1703
	(0.0122)***	(0.0122)***	(0.0302)***	(0.0300)***	(0.0301)***	(0.0301)***
marriage	0.0411	0.0418	0.0260	0.0269	0.0271	0.0271
	(0.0222)*	(0.0222)*	(0.0542)	(0.0539)	(0.0540)	(0.0540)
children	0.0070	0.0070	0.0059	0.0059	0.0059	0.0059
	(0.0037)*	(0.0037)*	(0.0096)	(0.0095)	(0.0096)	(0.0096)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All regressions include industry, region, and time dummies. The models are estimated with the unconditional transformed QML estimator proposed in this paper. Standard errors are computed with the Delta method and reported in parentheses. The formulas for the marginal effects are equations (47) and (51).

Table 7: Estimation results: cohabitation premiums

ln(wage)	static	space dynamic	time dynamic	time-space dynamic		
	$\lambda = \rho_0 = \rho_1 = 0$	$\lambda = \rho_1 = 0$	$\rho_0 = \rho_1 = 0$	$\rho_1 = 0$	$\rho_1 = -\lambda\rho_0$	unrestricted
short-run premiums						
union		0.0052		0.0041	0.0059	0.0060
		(0.0021)**		(0.0020)**	(0.0026)**	(0.0027)**
marriage		0.0033		0.0007	0.0009	0.0010
		(0.0221)		(0.0013)	(0.0019)	(0.0020)
children		0.0006		0.0001	0.0002	0.0002
		(0.0004)		(0.0002)	(0.0003)	(0.0003)
long-run premiums						
union		0.0052		0.0210	0.0129	0.0112
		(0.0021)**		(0.0107)**	(0.0057)**	(0.0120)
marriage		0.0033		0.0033	0.0021	0.0018
		(0.0221)		(0.0068)	(0.0042)	(0.0040)
children		0.0006		0.0007	0.0005	0.0004
		(0.0004)		(0.0012)	(0.0007)	(0.0008)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All regressions include industry, region, and time dummies. The models are estimated with the unconditional transformed QML estimator proposed in this paper. Standard errors are computed with the Delta method and reported in parentheses. The formulas for the cohabitation premiums are equations (48) and (52).

Table 8: PSID variable index

		2001	2003	2005	2007	2009	2011
wage	head	ER17235	ER21159	ER25148	ER36153	ER42188	ER47501
	wife	ER17805	ER21409	ER25406	ER36411	ER42440	ER47758
year of birth		ER33606	ER33706	ER33806	ER33906	ER34006	ER34106
union	head	ER17224	ER21150	ER25138	ER36143	ER42178	ER47491
	wife	ER17794	ER21400	ER25396	ER36401	ER42430	ER47748
marriage		ER17024	ER21023	ER25023	ER36023	ER42023	ER47323
children		ER17016	ER21020	ER25020	ER36020	ER42020	ER47320
industry	head	ER17227	ER21146	ER25128	ER36133	ER42168	ER47480
	wife	ER17797	ER21396	ER25386	ER36391	ER42420	ER47737
region		ER20376	ER24143	ER28042	ER41032	ER46974	ER52398
interview number		ER33601	ER33701	ER33801	ER33901	ER34001	ER34101
relation to head		ER33603	ER33703	ER33803	ER33903	ER34003	ER34103

Table 9: Summary statistics

	mean	std. dev.	min.	max.
ln(wage)	2.7152	0.4444	1.6390	5.1240
age	46.3117	10.1165	20	85
union	0.2676	0.4428	0	1
marriage	0.7378	0.4399	0	1
children	0.9266	1.0988	0	8
primary sector	0.0788	0.2695	0	1
manufacturing	0.1696	0.3753	0	1
public administration	0.0357	0.1855	0	1
northeast	0.1222	0.3275	0	1
north central	0.2841	0.4510	0	1
south	0.3839	0.4864	0	1