Appendices

A Details on Aggregate Level Analysis

Impulse response functions are computed following the algorithm. Step 1 - solve dynamically forward the estimated equation putting all shocks to zero; step 2 - simulate the equation setting the fiscal adjustment plan to 1% of GDP; step 3 - compute impulse response as a difference between step 2 and step 1; step 4 - compute confidence intervals using the block bootstrap to take into account serial correlation.

In Section 2.2 we build impulse response functions using truncated moving average model. In particular we estimate the following specification:

$$\begin{aligned}\Delta y_t = {} & \alpha + B_1(L)\cdot f_t^u \cdot EB_t + B_2(L) f_t^u \cdot TB_t + ...\\ & ... + C_1(L)\cdot f_t^a \cdot EB_t + C_2(L)\cdot f_t^a \cdot TB_t + ...\\ & ... + \sum_{j=1}^{H} D_j \cdot f_{t,t+j}^a \cdot EB_t + \sum_{j=1}^{H} E_j \cdot f_{t,t+j}^a \cdot TB_t + \epsilon_t\end{aligned}$$

with:

$$f_{t,t+j}^a = \delta_j^{TB}\cdot f_t^u \cdot TB_t + \epsilon_{t+j}^1, \text{ for } j = \overline{1,H}$$
$$f_{t,t+j}^a = \delta_j^{EB}\cdot f_t^u \cdot EB_t + \epsilon_{t+j}^2, \text{ for } j = \overline{1,H}$$

where B(L) and C(L) are polynomials of the length six, H - is the anticipation horizon and also equal to six. We follow Mertens and Ravn (2012) on this and six is the median implementation lag.

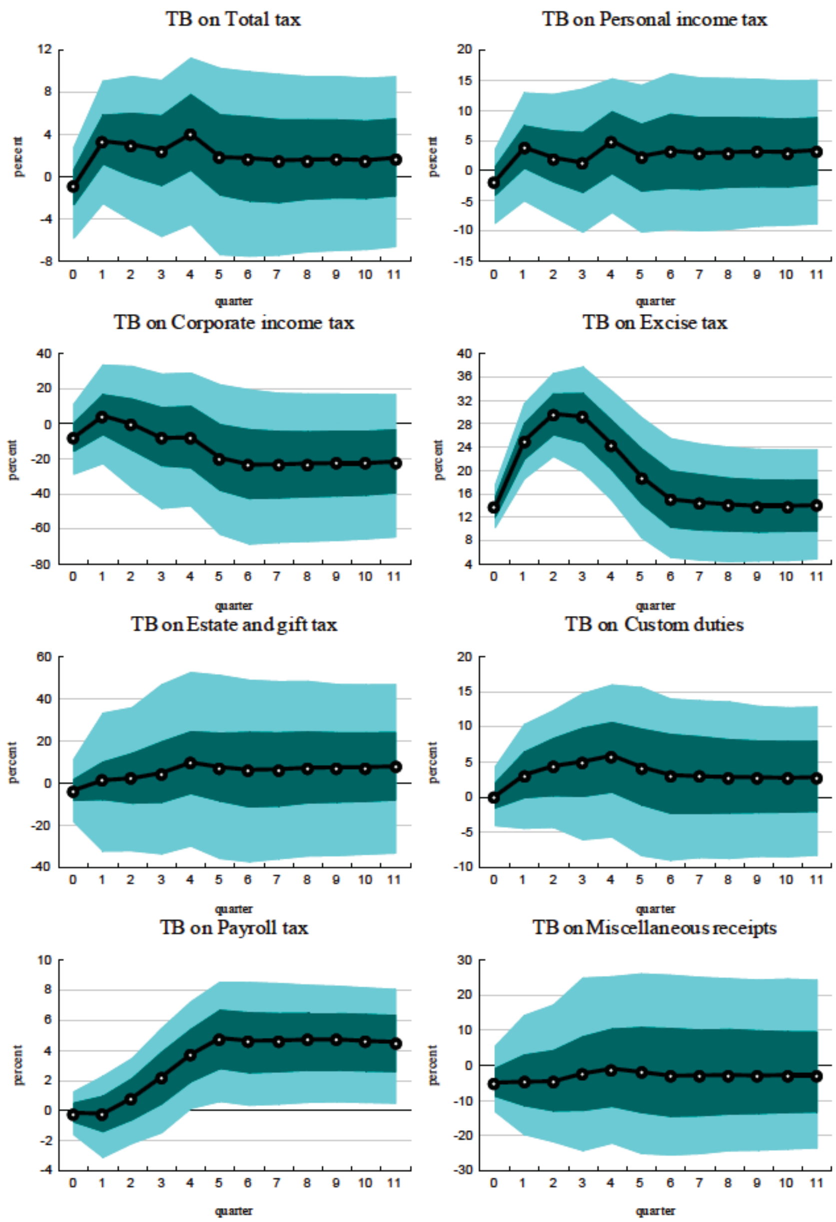
Figure 11 and 12 show the estimated impulse response functions of several other tax receipts shares of GDP to TB and EB fiscal adjustment plans. Those impulse responses are obtained using truncated moving average model.

B Industry Data

In this section we describe the data we use in our analysis.

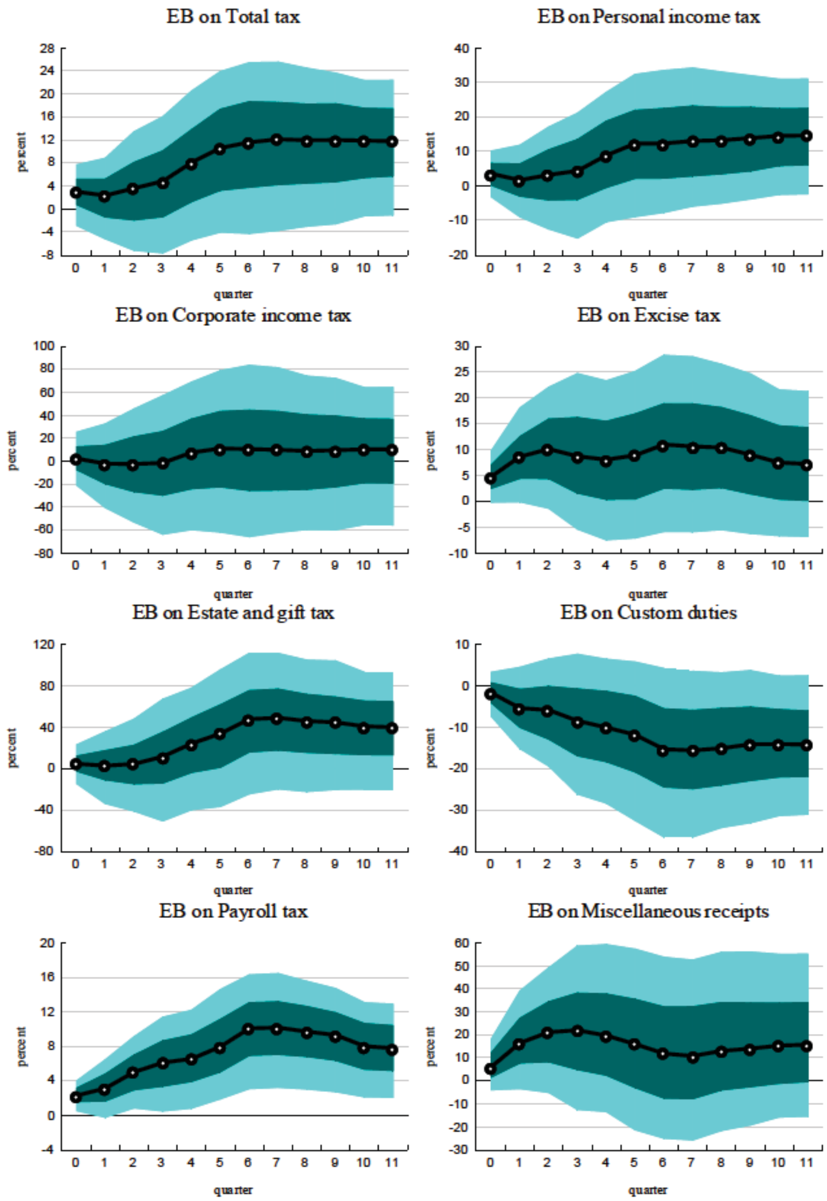
Firstly, the disaggregation level, $n = 62$, is determined by starting from the finest decomposition available on the Bureau of Economic Analysis (BEA) at a yearly frequency, namely 71 sectors, and then aggregating those sectors whose

Figure 11: Tax Receipts Response to TB Plans



data are not available for older years. We exclude the Government sector and consider only Government Enterprises as the only public, but politically independent, sector. The Government sector needs to be excluded since its outcome variable is G, government spending, which mechanically falls when a fiscal adjustment occurs.

Figure 12: Tax Receipts Response to EB Plans



Value Added

We use real industry value-added as the dependent variable, Δy_{it}. Value-added equals gross output minus intermediate inputs. It consists of compensation of employees, taxes on production and imports less subsidies (formerly indirect business taxes and non-tax payments), and gross operating surplus (formerly "other value added"). We prefer it over gross output to be consistent with Acemoglu, Akcigit, and Kerr (2016).[29]

Industry Specific Shares:

Following Acemoglu, Akcigit, and Kerr (2016), we construct the vector of industry-specific weights by exploiting information from the input-output tables, namely: $\omega_i^{EB} = \dfrac{Sales_{i\to G}}{Sales_i}$; where "G" stands for Government.[30] By doing so, we take into account the fact that the government purchases goods and services in different quantities from each sector.[31] Lastly, the vector of weights for the EB plan, denoted by $\boldsymbol{\omega}^{EB}$, is then normalized to one.
On the contrary, we assume that aggregate TB fiscal plans impact each sector in the same fashion, therefore, we set $\omega_i^{TB} = 1/n$ for all i and the $n \times 1$ vector will be: $\boldsymbol{\omega}^{TB} = 1/n \cdot \mathbf{1}_n$.

B.1 Input-Output Network

The BEA provides I-O tables that report the amount of commodity used (Use Table) and made (Make Table) by each industry. Horowitz, Planting, et al. (2006) outline the procedure to construct an industry-by-industry direct requirement matrix, with elements given by $SALES_{j\to i}/SALES_i$ for each sector. Let's denote this matrix by A and note that its elements coincide one to one with the weights of $\Delta y_{i,t}^{\text{down}}$ in Equation 6. Therefore, the downstream spatial variable can be written in vector notation as: $\Delta \boldsymbol{y}_t^{\text{down}} = A \cdot \Delta \log \boldsymbol{y}_t$ and matrix A can be constructed from the Make and Use Tables of the BEA.[32] Henceforth we will refer to matrix A as the "downstream matrix".
Finally, we construct a new matrix starting from A and using BEA's industry specific gross output, such that its $(ij)_{th}$ element is represented by $SALES_{i\to j}/SALES_i$, which coincides one to one with the weights of $\Delta y_{i,t}^{\text{up}}$ in Equation 6. We denote this new matrix by $\hat{A}^T$, and refer to it as the "upstream matrix". The upstream spatial variable can now be written in vector

[29] Their decision is justified by the fact that value-added is adjusted for energy costs, non-manufacturing input, and inventory changes which are all outside of the general equilibrium model which provides the theoretical underpinning to their empirical strategy.

[30] Our definition of Government encompasses both Federal and State&Local government spending. We therefore exclude here Government Enterprises, which instead are considered as part of the industrial network.

[31] We thank Roberto Perotti for this point.

[32] We use the Make and Use tables of year 1997, which is the closest to the occurrence of fiscal plans. Nevertheless, notice that I-O matrices are fairly stable over time.

notation as: $\Delta \boldsymbol{y}_t^{\text{up}} = \hat{A}^T \cdot \Delta \log \boldsymbol{y}_t$.

The construction of matrices A and $\hat{A}^T$ starts from the analysis of the Make and Use tables illustrated in chapter 12 of Horowitz, Planting, et al. (2006). We outline here the details of the construction and the precise mapping between the theory and the data.

The Use Table

The Use table is a commodity-by-industry table which illustrates the uses of commodities by intermediate and final users. The rows of the Use Table represent the commodities (or products) and the sum of the entries in a row is the total output of that commodity. On the contrary, the columns display the industries that employ them and the final users. Horowitz, Planting, et al. (2006) provides a useful numerical example with 3 industries:

Example of Use Table - 3 Industries

Commodity/Industry	1	2	3	Final demand	Total Commodity Output
1	50	120	120	40	330
2	180	30	60	130	400
3	50	150	50	20	270
Scrap	1	3	1	0	5
VA	47	109	34	/	190
Total Industry Output	328	412	265	190	/

What is of our interest is clearly the $n \times n$ commodity-by-industry part of the Table, whose values can be denoted with the following notation:

$$(Use)_{ij} = \text{INP}_{i \to j} := \text{Commodity } i \text{ used as input by Industry } j$$

Therefore, the $n \times n$ part of the Use Table we are going to use is:

$$U = \begin{bmatrix} \text{INP}_{1\to 1} & \text{INP}_{1\to 2} & \text{INP}_{1\to 3} \\ \text{INP}_{2\to 1} & \text{INP}_{2\to 2} & \text{INP}_{2\to 3} \\ \text{INP}_{3\to 1} & \text{INP}_{3\to 2} & \text{INP}_{3\to 3} \end{bmatrix} = \begin{bmatrix} 50 & 120 & 120 \\ 180 & 30 & 60 \\ 50 & 150 & 50 \end{bmatrix}$$

In practice, the above matrix U is a "symmetric" commodity-by-industry Use Table.

Next step boils down in constructing a *commodity-by-industry direct requirement table* by dividing each industry's input, $\text{INP}_{j\to i}$, by its corresponding total industry output, y_i. We denote such a matrix with letter B:

$$\text{B} = \begin{bmatrix} \frac{\text{INP}_{1\to1}}{y_1} & \frac{\text{INP}_{1\to2}}{y_2} & \frac{\text{INP}_{1\to3}}{y_3} \\ \frac{\text{INP}_{2\to1}}{y_1} & \frac{\text{INP}_{2\to2}}{y_2} & \frac{\text{INP}_{2\to3}}{y_3} \\ \frac{\text{INP}_{3\to1}}{y_1} & \frac{\text{INP}_{3\to1}}{y_2} & \frac{\text{INP}_{3\to3}}{y_3} \end{bmatrix} = \begin{bmatrix} \frac{50}{328} & \frac{120}{412} & \frac{120}{265} \\ \frac{180}{328} & \frac{30}{412} & \frac{60}{265} \\ \frac{50}{328} & \frac{150}{412} & \frac{50}{265} \end{bmatrix} = \begin{bmatrix} 0.152 & 0.291 & 0.453 \\ 0.549 & 0.073 & 0.226 \\ 0.152 & 0.364 & 0.189 \end{bmatrix}.$$

Notice one important thing: matrix B is different from matrix A, since $x_{i\to j} \neq \text{INP}_{i\to j}$: the former is an industry output flow, while the second measures a commodity flow to an industry.

The Make Table

The Make table is an industry-by-commodity table which shows the production of commodities by industries. Row i represents an industry and its summation delivers the total industry output, y_i. Column j represents a commodity and its summation delivers the total commodity output.
Borrowing again Horowitz, Planting, et al., 2006's 3 industries example, we have:

Example of Make Table - 3 Industries

Industry/Commodity	1	2	3	Scrap	Total Industry Output
1	300	25	0	3	328
2	30	360	20	2	412
3	0	15	250	0	265
Total Commodity Output	330	400	270	5	/

Similarly to what done for the Use Table, we are interested in the central $n \times n$ elements of the table, which we can denote by V. The generic element of the "heart" of the Make table is:

$$(Make)_{ij} = \text{OUT}_{i\to j} := \text{Commodity } j \text{ produced by Industry } i$$

Therefore, the $n \times n$ part of the Make Table we are going to employ is:

$$V = \begin{bmatrix} \mathrm{OUT}_{1\to1} & \mathrm{OUT}_{1\to2} & \mathrm{OUT}_{1\to3} \\ \\ \mathrm{OUT}_{2\to1} & \mathrm{OUT}_{2\to2} & \mathrm{OUT}_{2\to3} \\ \\ \mathrm{OUT}_{3\to1} & \mathrm{OUT}_{3\to2} & \mathrm{OUT}_{3\to3} \end{bmatrix} = \begin{bmatrix} 300 & 25 & 0 \\ 30 & 360 & 20 \\ 0 & 15 & 250 \end{bmatrix}$$

In practice, the above matrix V is a "symmetric" industry-by-commodity Make Table.

Analogously to what done before, we now take ratios; in particular, we divide each element of V by the total production of commodity j. The resulting matrix is denoted by D, and its generic element is:

$$(D)_{ij} = \frac{OUT_{i\to j}}{\sum_{k=1}^{n} OUT_{k\to j}} = \frac{OUT_{i\to j}}{C_j}$$

where $C_j := \sum_{k=1}^{n} OUT_{k\to j}$ is the total production of commodity j. D represents the share of industry i in the total production of commodity j; not surprisingly, Horowitz, Planting, et al. (2006) refer to this matrix as the "market share matrix". In the 3 industries/commodities example we have:

$$D = \begin{bmatrix} \frac{OUT_{1\to1}}{C_1} & \frac{OUT_{1\to2}}{C_2} & \frac{OUT_{1\to3}}{C_3} \\ \\ \frac{OUT_{2\to1}}{C_1} & \frac{OUT_{2\to2}}{C_2} & \frac{OUT_{1\to3}}{C_3} \\ \\ \frac{OUT_{3\to1}}{C_1} & \frac{OUT_{3\to2}}{C_2} & \frac{OUT_{3\to3}}{C_3} \end{bmatrix} = \begin{bmatrix} \frac{300}{330} & \frac{25}{400} & \frac{0}{270} \\ \\ \frac{30}{330} & \frac{360}{400} & \frac{20}{270} \\ \\ \frac{0}{330} & \frac{15}{400} & \frac{250}{270} \end{bmatrix} = \begin{bmatrix} 0.909 & 0.063 & 0 \\ 0.091 & 0.900 & 0.074 \\ 0 & 0.038 & 0.926 \end{bmatrix}$$

Adjustment for Scrap Products

The I-O accounts include a commodity for scrap, which is a byproduct of industry production. No industry produces scrap on demand; rather, it is the result of production to meet other demands. In order to make the I-O model work correctly, we have to eliminate scrap as a secondary product. At the same time, we must also keep industry output at the same level.

This adjustment is accomplished by calculating the ratio of non-scrap output to industry output for each industry and then applying these ratios to the market shares matrix in order to account for total industry output. More precisely, the non-scrap ratio, which I denote by θ_i, is defined as follows:

$$\theta_i = \frac{y_i - (\text{scrap})_i}{y_i}$$

and represents the share of total industry output i made of commodity different from "scrap". In the 3 industries example we have:

Industry	Tot.Ind.Out.	Scrap	Δ	θ_i
1	328	3	325	0.991
2	412	2	410	0.995
3	265	0	265	1

The market shares matrix, D, is adjusted for scrap by dividing each row by the non-scrap ratio for that industry. In the resulting transformation matrix, called W, the implicit commodity output of each industry has been increased. In other words, we are increasing each market share to take into account that to produce each unit of each commodity, industry i will produce $1/\theta_i$ units of output. In essence, we are spreading the production of commodity "scrap" over the production of all the other commodities:

$$W = \begin{bmatrix} \frac{OUT_{1\to1}}{C_1} \cdot \frac{1}{\theta_1} & \frac{OUT_{1\to2}}{C_2} \cdot \frac{1}{\theta_2} & \frac{OUT_{1\to3}}{C_3} \cdot \frac{1}{\theta_3} \\ \frac{OUT_{2\to1}}{C_1} \cdot \frac{1}{\theta_1} & \frac{OUT_{2\to2}}{C_2} \cdot \frac{1}{\theta_2} & \frac{OUT_{1\to3}}{C_3} \cdot \frac{1}{\theta_3} \\ \frac{OUT_{3\to1}}{C_1} \cdot \frac{1}{\theta_1} & \frac{OUT_{3\to2}}{C_2} \cdot \frac{1}{\theta_2} & \frac{OUT_{3\to3}}{C_3} \cdot \frac{1}{\theta_3} \end{bmatrix} = \begin{bmatrix} \frac{0.909}{0.991} & \frac{0.063}{0.991} & \frac{0}{0.991} \\ \frac{0.091}{0.995} & \frac{0.900}{0.995} & \frac{0.074}{0.995} \\ \frac{0}{1} & \frac{0.038}{1} & \frac{0.926}{1} \end{bmatrix} = \begin{bmatrix} 0.917 & 0.063 & \\ 0.091 & 0.904 & 0 \\ 0 & 0.038 & 0 \end{bmatrix}$$

The Direct Requirement Table

To summarize:

1. We constructed matrix B, a commodity-by-industry direct requirement table, whose columns tell us how much an industry j needs of commodity i relative to its own total industry production.

2. We constructed matrix W, an industry-by-commodity matrix which represent the market share - adjusted for scrap - of each industry i in the production of a commodity j.

By combining these two matrices we can obtain an industry-by-industry direct requirement matrix:

$$\underbrace{P}_{\text{industry}\times\text{industry}} := \underbrace{W}_{\text{industry}\times\text{commodity}} \cdot \underbrace{B}_{\text{commodity}\times\text{industry}}$$

In order to understand the meaning of each element of matrix P, it is important to derive it analytically:

$$P = \underbrace{\begin{bmatrix} \frac{OUT_{1\to1}}{C_1\cdot\theta_1} & \frac{OUT_{1\to2}}{C_2\cdot\theta_2} & \frac{OUT_{1\to3}}{C_3\cdot\theta_3} \\ \\ \frac{OUT_{2\to1}}{C_1\cdot\theta_1} & \frac{OUT_{2\to2}}{C_2\cdot\theta_2} & \frac{OUT_{1\to3}}{C_3\cdot\theta_3} \\ \\ \frac{OUT_{3\to1}}{C_1\cdot\theta_1} & \frac{OUT_{3\to2}}{C_2\cdot\theta_2} & \frac{OUT_{3\to3}}{C_3\cdot\theta_3} \end{bmatrix}}_{W} \cdot \underbrace{\begin{bmatrix} \frac{\mathrm{INP}_{1\to1}}{y_1} & \frac{\mathrm{INP}_{1\to2}}{y_2} & \frac{\mathrm{INP}_{1\to3}}{y_3} \\ \\ \frac{\mathrm{INP}_{2\to1}}{y_1} & \frac{\mathrm{INP}_{2\to2}}{y_2} & \frac{\mathrm{INP}_{2\to3}}{y_3} \\ \\ \frac{\mathrm{INP}_{3\to1}}{y_1} & \frac{\mathrm{INP}_{3\to2}}{y_2} & \frac{\mathrm{INP}_{3\to3}}{y_3} \end{bmatrix}}_{B}$$

Denoting by p_{ij} the generic element of P, we have:

$$p_{ij} = \frac{\frac{OUT_{i\to1}}{C_1\cdot\theta_1}\cdot\mathrm{INP}_{1\to j} + \frac{OUT_{i\to2}}{C_2\cdot\theta_2}\cdot\mathrm{INP}_{2\to j} + \frac{OUT_{i\to3}}{C_3\cdot\theta_3}\cdot\mathrm{INP}_{3\to j}}{y_j} \approx \frac{\mathrm{SALES}_{i\to j}}{SALES_j}$$

In other words, p_{ij} represents how much industry j depends on inputs form industry i relative to its own total industry output y_j.[33]

Notice that the transposed of matrix P is approximately equal to matrix A in the paper:

$$P \approx \begin{bmatrix} \frac{\mathrm{SALES}_{1\to1}}{\mathrm{SALES}_1} & \frac{\mathrm{SALES}_{1\to2}}{\mathrm{SALES}_2} & \frac{\mathrm{SALES}_{1\to3}}{\mathrm{SALES}_3} \\ \\ \frac{\mathrm{SALES}_{2\to1}}{\mathrm{SALES}_1} & \frac{\mathrm{SALES}_{2\to2}}{\mathrm{SALES}_2} & \frac{\mathrm{SALES}_{2\to3}}{\mathrm{SALES}_3} \\ \\ \frac{\mathrm{SALES}_{3\to1}}{\mathrm{SALES}_1} & \frac{\mathrm{SALES}_{3\to2}}{\mathrm{SALES}_2} & \frac{\mathrm{SALES}_{3\to3}}{\mathrm{SALES}_3} \end{bmatrix} \implies \boxed{A \approx P^T}$$

[33]Notice that a big assumption is made in the construction of this matrix: if industry i has adjusted market share of production of commodity K, $OUT_{i\to K}/(C_K\cdot\theta_K)$ equal to, say 10%, then it is assumed that if industry j purchases $z := \mathrm{INP}_{K\to j}$ dollars of commodity K, then 10% of z\$ come from industry i. This must be true on average but it might not be exactly true case by case.

Matrix P can be either constructed from the Make and Use table or downloaded from the BEA, as an industry-by-industry direct requirement table. Its transposed value identifies the matrix A in Equation (6).
The construction of matrix $\hat{A}^T$, in equation (7), is trivial once we have matrix A as well as a vector of average industry output.

C Spatial Econometric Estimation

We believe that our empirical methodology presents some results of independent interest. Although we do not want to divert attention from the macroeconomic focus of the paper, we believe certain econometric facts are worth mentioning here in the Appendix. We provide this discussion in the spirit of promoting the usage of these new techniques in macroeconomic analysis.

Firstly, the adoption of spatial econometric methods allows us to disentangle the direct and network effect of aggregate shocks. This is a novel and recent innovation in macroeconomics, as noted in Ozdagli and Weber (2017). Secondly, spatial models are traditionally estimated by row-normalizing and removing the main diagonal from the weighting matrix. Another common assumption is homoskedasticity of the error term. In a recent paper, Aquaro, Bailey, and Pesaran (2019) develop a new estimator which relaxes homoskedasticity and allow for different spatial coefficients, thus indirectly relaxing the row-normalization assumption. They refer to it as Heterogenous Spatial Autoregressive model (HSAR). They also point out that not assuming zero entries on the main diagonal of the weighting matrix is simply a re-parameterization of the model, which does not harm the statistical properties of the MLE, but does change the interpretation of the parameters.[34] Their econometric model, adopted by Ozdagli and Weber (2017), is very convenient for macroeconomic applications which use non-row-normalized, dense main diagonal weighting matrices and in a setting where units are subject to heteroskedastic idiosyncratic shocks.

However, we highlight that even the standard dynamic spatial panel autoregressive model of Yu, DeJong, and Lee (2008) can easily be relaxed to accommodate for non-zero entries on the main diagonal and non-row-normalized weighting matrix with heteroskedastic errors.[35] Our construction of a Bayesian MCMC, similar to the one in LeSage and Pace (2009), is thus an easy and natural extension to the more general version of the spatial panel autoregressive

[34] We are grateful to Hashem Pesaran for making us aware of this.
[35] We thank Lung-Fei Lee for pointing this out.

model of Yu, DeJong, and Lee (2008). Moreover, the Bayesian MCMC method provides an easy way to recover the posterior distributions of the aggregate effects of the shocks, as illustrated earlier.
We encourage macroeconomists to adopt spatial econometric tools to study the propagation of aggregate shocks into a network of sub-units (countries, industries, regions...) but in doing so we also recommend them to follow three good practices:

1. Firstly, always allow for heteroskedasticity, since sub-units in general have different volatilities.

2. Secondly, never remove the main diagonal from the empirically observed weighting matrices, in our case A and $\hat{A}^T$. In fact, zero-entries in the main diagonal imposes a lack of spillovers within the same observed unit ("intra-unit feedback"). This is a reasonable assumption when units are individuals - like in standard spatial econometric applications - but it is not sensible when units are aggregates, such as industries. Notice, that the empirically observed A and $\hat{A}^T$ weighting matrices from our analysis exhibit very dense main diagonals (see Figure 9).

3. Thirdly, never row-normalize the weighting matrices. Row-normalization flattens the differences in the degree of connection of each unit. For instance, in our application with the industrial network, A and $\hat{A}^T$ exhibit very different row-sums, indicative of different degrees of exposure to customer and supplying industries.

We recommend using either the Bayesian MCMC methodology developed here and detailed in Appendix C.3 or the HSAR model of Aquaro, Bailey, and Pesaran (2019), whenever the application requires heterogeneous spatial coefficients. The relationship between the two models is left for future research.

In what follows we outline the details of the spatial econometric estimator that we employ.

C.1 Log-likelihood

The standard way to estimate the parameters of Equations (6) and (7) is via maximum likelihood (see LeSage and Pace (2009) for an introduction to spatial econometrics). The asymptotic and small sample properties of the MLE have been studied in Lee (2004) for cross-sectional data, and in Yu, DeJong, and Lee (2008), for dynamic panel data models with fixed effects.

We provide here the derivation of the log-likelihood of the baseline model (6), necessary for the calculation of both the MLE and the conditional posterior distributions of the Bayesian MCMC.[36] Collecting fiscal adjustment plans, industry fixed effects and other controls into matrix X_t, from Equation (6):

$$
\begin{aligned}
&H_t^{-1} \cdot \underset{n\times 1}{\Delta y_t} = \underset{n\times k}{X_t} \cdot \beta + \varepsilon_t \\
&H_t = \left(I_n - \rho^{down} \cdot A \cdot TB_t - \rho^{up} \cdot \hat{A}^T \cdot EB_t\right)^{-1} \\
&\varepsilon_t \sim \mathscr{N}(0, \Omega), \forall t \in \{1, ..., T\} \\
&\Omega = diag(\sigma_1^2, ..., \sigma_n^2) \\
&\varepsilon_t \perp \varepsilon_{t+i}, \quad \forall t \in \{1, ..., T\}, \forall i \in \mathscr{Z}
\end{aligned}
$$

where k is the number of regressors.[37] We now make a convenient change in the notation: 1. we now use $'$ as a symbol for transposition instead of ^{T}; 2. we now set $\rho_1 = \rho^{down}$, $\rho_2 = \rho^{up}$, $A = W_1$ and $\hat{A}' = W_2$. We have:

$$
Z_t := H_t^{-1} \cdot \Delta y_t \sim \mathscr{N}(X_t\beta, \Omega) \implies \Delta y_t \sim \mathscr{N}(H_t X_t \beta, H_t \Omega H_t')
$$

The density function of the random vector Δy_t is:

$$
f(\underset{n\times 1}{\Delta y_t}|X_t, \rho, \beta, \Omega) = \frac{1}{\sqrt{(2\pi)^n \cdot |H_t \Omega H_t'|}} \exp\left\{-\frac{1}{2} \cdot (\Delta y_t - H_t X_t \beta)' \cdot (H_t \Omega H_t')^{-1} \cdot (\Delta y_t - H_t X_t \beta)\right\},
$$

with $\rho = \left[\rho^{down}, \rho^{up}\right]$.
Given that $(H_t \Omega H_t')^{-1} = (H_t')^{-1} \cdot \Omega^{-1} \cdot H_t^{-1}$ and $|H_t \Omega H_t'| = |H_t|^2 \cdot |\Omega|$, we have:

$$
\begin{aligned}
f(\Delta y_t|\cdot) &= (2\pi)^{-n/2} \cdot |H_t|^{-1} \cdot |\Omega|^{-1/2} \cdot \exp\Big\{ -\frac{1}{2}(Z_t - X_t\beta)' \cdot H_t' \cdot (H_t')^{-1} \cdot \Omega^{-1} \cdot H_t^{-1} \cdot H_t \cdot (Z_t - X_t \\
&= (2\pi)^{-n/2} \cdot |(I_n - \rho_1 W1 TB_t - \rho_2 W_2 EB_t)^{-1}|^{-1} \cdot |\Omega|^{-1/2} \exp\left\{ -\frac{1}{2}\varepsilon_t' \Omega^{-1} \varepsilon_t\right\} \\
&= (2\pi)^{-n/2} \cdot |I_n - \rho_1 \cdot W_1 \cdot TB_t - \rho_2 \cdot W_2 \cdot EB_t| \cdot |\Omega|^{-1/2} \exp\left\{ -\frac{1}{2}\varepsilon_t' \Omega^{-1} \varepsilon_t\right\},
\end{aligned}
$$

At this point we need to find the likelihood of the random vector $\Delta y = \begin{bmatrix}\Delta y_1' & \dots & \Delta y_T'\end{bmatrix}$. Since the model is static and we have assumed $cov(\varepsilon_t, \varepsilon_{t-k}) =$

[36] Results for the inverted model, Equation (7)) are symmetric to the baseline case.

[37] k in our baseline is n fixed effects plust 6 fiscal adjustment components (unexpected, announced and future for both TB and EB plans) plus 2 year dummies for 2008 and 2009.

$\underset{n\times n}{\rho^0}$, then Δy_t is *iid* over time. By consequence, the following holds:

$$f(\underset{nT\times 1}{\Delta y}|X_1,\ldots,X_T,\rho,\beta,\Omega) = \prod_{t=1}^{T} f(\underset{n\times 1}{\Delta y_t}|X_t,\rho,\beta,\Omega) = \big((2\pi)^n|\Omega|\big)^{-T/2}\cdot$$
$$\cdot\prod_{t=1}^{T}|I_n - \rho_1\cdot W_1\cdot TB_t - \rho_2\cdot W_2\cdot EB_t|\exp\Big\{-\frac{1}{2}\cdot\sum_{t=1}^{T}\varepsilon_t'\Omega^{-1}\varepsilon_t\Big\}.$$

Now we divide the time series of length T in three different sub-periods. In doing so, consider the following new parameters:

- t_1: set of years when a tax based fiscal adjustment occurs.
  Formally $t_1 := \{1,...,t,...,T_1|\ t \text{ such that } TB_t = 1\}$. We set: $H_t|\ t\in t_1 = (I_n-\rho_1\cdot W_1)^{-1} = H_\tau$.

- t_2: set of years when an expenditure tax based fiscal adjustment occurs.
  Formally: $t_2 := \{1,...,t,...,T_2|\ t \text{ such that } EB_t = 1\}$. We set $H_t|\ t\in t_2 = (I_n-\rho_2\cdot W_2)^{-1} = H_\gamma$.

- t_3: set of years when neither a tax based fiscal adjustment nor an expenditure based fiscal adjustment occurs.
  Formally $t_3 := \{1,...,t,...,T_3|\ t \text{ such that } TB_t = 0 \wedge EB_t = 0\}$. We set $H_t|\ t\in t_3 = (I_n)^{-1} = I_n$.

Therefore, we have that t_1, t_2 and t_3 account for a partition of the whole time series and $T = T_1+T_2+T_3$. By consequence we have:

$$\begin{aligned}
\prod_{t=1}^{T}|I_n-\rho_1W_1TB_t-\rho_2W_2EB_t| &= \prod_{t=1}^{T}|H_t^{-1}| \\
&= \prod_{t=1}^{T}\frac{1}{|H_t|} \\
&= \prod_{t\in t_1}^{T_1}\frac{1}{|H_t|}\cdot\prod_{t\in t_2}^{T_2}\frac{1}{|H_t|}\cdot\prod_{t\in t_3}^{T_3}\frac{1}{|H_t|} \\
&= |H_\tau|^{-T_1}\cdot|H_\gamma|^{-T_2}\cdot|I_n|^{-T_3} \\
&= |I_n-\rho_1\cdot W_1|^{T_1}\cdot|I_n-\rho_2W_2|^{T_2}
\end{aligned}$$

At this point, we rewrite the probability density function of our dependent variable as:

$$f(\Delta y_t|X_1,\ldots,X_T,\rho,\beta,\Omega)=(2\pi)^{-nT/2}\cdot|\Omega|^{-T/2}\cdot$$
$$\cdot|I_n-\rho_1\cdot W_1|^{T_1}\cdot|I_n-\rho_2 W_2|^{T_2}\cdot\exp\left\{-\frac{1}{2}\cdot\sum_{t=1}^{T}\varepsilon_t'\cdot\Omega^{-1}\cdot\varepsilon_t\right\}.$$

Finally, the log-likelihood of our dataset is:

$$\log\mathscr{L}(\rho,\beta,\Omega|\Delta y_1,\ldots,\Delta y_T,X_1,\ldots,X_T)=-\frac{nT}{2}\ln(2\pi)-\frac{T}{2}\cdot\ln(|\Omega|)+$$
$$+T_1\cdot\ln(|I_n-\rho_1\cdot W_1|)+T_2\cdot\ln(|I_n-\rho_2W_2|)-\frac{1}{2}\cdot\sum_{t=1}^{T}\varepsilon_t'\cdot\Omega^{-1}\cdot\varepsilon_t.$$

with:

$$\varepsilon_t=Z_t-X_t\cdot\beta=H_t^{-1}\cdot\Delta y_t-X_t\beta=(I_n-\rho_1W_1TB_t-\rho_2W_2EB_t)\cdot\Delta y_t-X_t\cdot\beta.$$

Furthermore, we impose the condition $\lambda_{\min}^{-1}<\hat{\rho}_1<\lambda_{\max}^{-1}$ and $\mu_{\min}^{-1}<\hat{\rho}_2<\mu_{\max}^{-1}$, where λ and μ are the eigenvalues of the spatial matrices W_1 and W_2 respectively. This condition guarantees that the estimated model will have positive definite covariance matrix (see Ord (1975)).
Notice that in the inverted model of Equation (7), it is enough to switch the definition of W_1 and W_2 by setting: $A=W_2$ and $\hat{A}'=W_1$.

C.2 The Analytical Fisher Information Matrix

In order to derive the Fisher Information Matrix we firstly need to obtain the total gradient of the log-likelihood function. Let's start with the spatial coefficient ρ_1:

$$\frac{\partial\log\mathscr{L}(\theta|\Delta y,X)}{\partial\rho_1}=T_1\frac{1}{|I_n-\rho_1W_1|}\frac{\partial|I_n-\rho_1W_1|}{\partial\rho_1}-\frac{1}{2}\sum_{t=1}^{T}\frac{\partial(Z_t'\Omega^{-1}Z_t)}{\partial\rho_1}-2\frac{\partial(Z_t'\Omega^{-1}X_t\beta)}{\partial\rho_1}.$$

By some matrix algebra, it is possible to show that:

$$\frac{\partial(Z_t'\Omega^{-1}Z_t)}{\partial\rho_1}=-TB_t\cdot\Delta y_t'\cdot\Omega^{-1}\cdot W_1\cdot\Delta y_t-TB_t\cdot\Delta y_t'\cdot W_1'\Omega^{-1}\cdot\Delta y_t$$
$$+2\rho_1\cdot TB_t^2\cdot\Delta y_t'\cdot W_1\cdot\Omega^{-1}\cdot W_1\cdot\Delta y_t'+2\rho_2\cdot TB_t\cdot EB_t\cdot\Delta y_t'\cdot W_1\cdot\Omega^{-1}\cdot W_2\cdot\Delta y_t'$$

Since our fiscal adjustment plans are mutually exclusive, we have that $TB_t \cdot EB_t = 0$ for all t. Moreover, by rearranging the above expression, we get:

$$\frac{\partial(Z_t'\Omega^{-1}Z_t)}{\partial\rho_1} = -2 \cdot TB_t \cdot \Delta y_t' \cdot (I_n - \rho_1 \cdot W_1') \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t$$

After other matrix algebra, we get:

$$-2 \cdot \frac{\partial(Z_t \cdot \Omega^{-1}X_t\beta)}{\partial\rho_1} = 2 \cdot TB_t \cdot \Delta y_t' \cdot W_1' \cdot \Omega^{-1} \cdot X_t \cdot \beta$$

Wrapping up all together, and employing the notation introduced earlier: $(I_n - \rho_1 W_1)^{-1} = H_\tau$, we have:

$$\begin{aligned}
\frac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial\rho_1} &= T_1 \frac{1}{|I_n - \rho_1 W_1|} \frac{\partial |I_n - \rho_1 W_1|}{\partial \rho_1} + \\
&+ \sum_{t\in t_1}^{T_1} \Big[\Delta y_t' \cdot (I_n - \rho_1 \cdot W_1') \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t - \Delta y_t' \cdot W_1' \cdot \Omega^{-1} \cdot X_t \cdot \beta\Big] = \\
&= T_1 \frac{1}{|I_n - \rho_1 W_1|} \cdot |I_n - \rho_1 W_1| \cdot Tr\Big((I_n - \rho_1 W_1)^{-1} \cdot (-W_1)\Big) + \\
&+ \sum_{t\in t_1}^{T_1} \Big[\big((I_n - \rho_1 \cdot W_1) \cdot \Delta y_t\big)' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t - \beta' \cdot X_t' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\Big] \\
&= -T_1 \cdot Tr\Big(H_\tau \cdot W_1\Big) + \sum_{t\in t_1}^{T_1} \Big[\big(Z_t - X_t\beta\big)' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\Big] \\
&= \sum_{t\in t_1}^{T_1} \big(\varepsilon_t' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big) - T_1 \cdot Tr\big(H_\tau \cdot W_1\big).
\end{aligned}$$

By simmetry we have that:

$$\frac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial\rho_2} = \sum_{t\in t_2}^{T_2} \big(\varepsilon_t' \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t\big) - T_2 \cdot Tr\big(H_\gamma \cdot W_2\big),$$

with $H_\gamma = (I_n - \rho_2 W_2)^{-1}$, from the previous notation.
As far as concern the derivative with respect to β, we have already seen when

concentrating the log-likelihood that:

$$\begin{aligned}\frac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \beta} &= X' \cdot \Sigma^{-1} \cdot Z - X' \cdot \Sigma^{-1} \cdot X \cdot \beta \\ &= X' \cdot \Sigma^{-1} \cdot (Z - X \cdot \beta) = \\ &= X' \cdot \Sigma^{-1} \cdot \varepsilon = \\ &= \sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot \varepsilon_t.\end{aligned}$$

Concerning the derivatives with respect to σ_i^2, we need firstly to acknowledge that:

$$\sum_{t=1}^{T} \varepsilon_t' \cdot \Omega^{-1} \cdot \varepsilon_t = \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\varepsilon_{i,t}^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{t=1}^{T} \varepsilon_{i,t}^2,$$

and that:

$$\ln(|\Omega|) = \ln(\prod_{i=1}^{n} \sigma_i^2) = \sum_{i=1}^{n} \ln(\sigma_i^2).$$

Therefore, we have that:

$$\begin{aligned}\frac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \sigma_i^2} &= -\frac{T}{2} \frac{\partial \ln(|\Omega|)}{\partial \sigma_i^2} - \frac{1}{2} \cdot \frac{\partial}{\partial \sigma_i^2} \sum_{t=1}^{T} \varepsilon_t' \cdot \Omega^{-1} \cdot \varepsilon_t \\ &= -\frac{T}{2 \cdot \sigma_i^2} + \frac{1}{2 \cdot \sigma_i^4} \cdot \sum_{t=1}^{T} \varepsilon_{i,t}^2.\end{aligned}$$

We now have all the elements to write down the gradient of the log-likelihood:

$$\nabla \log \mathscr{L}(\theta|\Delta y, X) = \begin{bmatrix} \dfrac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_1} \\ \\ \dfrac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_2} \\ \\ \dfrac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \beta} \\ \\ \dfrac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \sigma_1^2} \\ \vdots \\ \dfrac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \sigma_n^2} \end{bmatrix}_{38\times 1} = \begin{bmatrix} \sum_{t\in t_1}^{T_1} \big(\varepsilon_t' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big) - T_1 \cdot Tr\big(H_\tau \cdot W_1\big) \\ \\ \sum_{t\in t_2}^{T_2} \big(\varepsilon_t' \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t\big) - T_2 \cdot Tr\big(H_\gamma \cdot W_2\big) \\ \\ \sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot \varepsilon_t \\ \\ -\dfrac{T}{2\cdot\sigma_1^2} + \dfrac{1}{2\cdot\sigma_1^4} \cdot \sum_{t=1}^{T} \varepsilon_{1,t}^2 \\ \vdots \\ -\dfrac{T}{2\cdot\sigma_n^2} + \dfrac{1}{2\cdot\sigma_n^4} \cdot \sum_{t=1}^{T} \varepsilon_{n,t}^2 \end{bmatrix}$$

Another round of derivation is now needed. Let's start with the first row of the matrix: all the derivatives of $\dfrac{\partial \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_1}$ with respect to all the parameters. To simplify notation we will refer with $\mathscr{H}_{ij}$ to the element of row i and column j of the Hessian matrix.

$$\begin{aligned}
\mathscr{H}_{1,1} &= \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_1^2} = \sum_{t\in t_1}^{T_1} \big(\frac{\partial \varepsilon_t'}{\partial \rho_1} \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big) - T_1 \cdot \frac{\partial Tr\big(H_\tau \cdot W_1\big)}{\partial \rho_1} \\
&= \sum_{t\in t_1}^{T_1} \big((-\Delta y_t' \cdot W_1') \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big) - T_1 \cdot Tr\big(\frac{\partial H_\tau}{\partial \rho_1} \cdot W_1\big) = \\
&= -\sum_{t\in t_1}^{T_1} \big(\Delta y_t' \cdot W_1' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big) - T_1 \cdot Tr\big((-H_\tau \cdot (-W_1) \cdot H_\tau) \cdot W_1\big) = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \sum_{t\in t_1}^{T_1} \big(\Delta y_t' \cdot W_1' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big)
\end{aligned}$$

Symmetrically we have:

$$\mathscr{H}_{2,2} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_2^2} =$$
$$= -T_2 \cdot Tr\big(W_2 \cdot H_\gamma \cdot W_2 \cdot H_\gamma\big) - \sum_{t \in t_2}^{T_2} \big(\Delta y_t' \cdot W_2' \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t\big)$$

Going back to the first row, we now calculate the cross derivative with respect to $\rho 2$. Before doing so, recall that, being the log-likelihood a continuously diffirentiable function, the Schwarz's theorem applies and the Hessian matrix is symmetric.

$$\mathscr{H}_{1,2} = \mathscr{H}_{2,1} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_1 \partial \rho_2} = 0.$$

Going on with the calculation we have:

$$\begin{aligned}
\mathscr{H}_{1,3:1,23} &= \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_1 \partial \beta} = \sum_{t \in t_1}^{T_1} \big(\frac{\partial \varepsilon_t'}{\partial \beta} \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big) \\
&= -\sum_{t \in t_1}^{T_1} X_t' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t \\
&= -X_\tau' \cdot \underbrace{(I_{T_1} \otimes \Omega^{-1})}_{\Sigma_\tau^{-1}} \cdot (I_{T_1} \otimes W_1) \cdot \Delta y_\tau
\end{aligned}$$

where $\mathscr{H}_{1,3:1,23}$ means all the elements of the first row, from column 3 up to column 23. X_τ and Δy_τ represent X and Δy but for the only years when a tax based fiscal adjustment occur:

$$X_\tau = \underset{T_1 n \times k}{\begin{bmatrix} X_1 \\ \vdots \\ X_t \\ \vdots \\ X_{T_1} \end{bmatrix}} \quad \text{and} \quad \Delta y_\tau = \underset{T_1 n \times k}{\begin{bmatrix} \Delta y_1 \\ \vdots \\ \Delta y_t \\ \vdots \\ \Delta y_{T_1} \end{bmatrix}} \quad \text{with } t \in t_1,$$

Symmetrically:

$$\begin{aligned}
\mathscr{H}_{2,3:2,23} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \rho_2 \partial \beta} &= \sum_{t \in t_2}^{T_2} \big(\frac{\partial \varepsilon_t'}{\partial \beta} \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t \big) \\
&= -\sum_{t \in t_2}^{T_2} X_t' \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t \\
&= -X_\gamma' \cdot \underbrace{(I_{T_2} \otimes \Omega^{-1})}_{\Sigma_\gamma^{-1}} \cdot (I_{T_2} \otimes W_2) \cdot \Delta y_\gamma,
\end{aligned}$$

with:

$$X_\gamma = \underset{T_2 n \times k}{\begin{bmatrix} X_1 \\ \vdots \\ X_t \\ \vdots \\ X_{T_2} \end{bmatrix}} \quad \text{and} \quad \Delta y_\gamma = \underset{T_2 n \times k}{\begin{bmatrix} \Delta y_1 \\ \vdots \\ \Delta y_t \\ \vdots \\ \Delta y_{T_2} \end{bmatrix}} \quad \text{with } t \in t_2,$$

$$\begin{aligned}
\mathscr{H}_{3,3:23,23} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \beta^2} &= \frac{\partial}{\partial \beta^2} \left(\sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot \varepsilon_t \right) \\
&= \sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot \frac{\partial (Z_t - X_t \cdot \beta)}{\partial \beta^2} \\
&= \sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot X_t \\
&= -X' \cdot \Sigma^{-1} \cdot X.
\end{aligned}$$

$$\mathscr{H}_{3,24:23,38} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial \beta \partial \sigma^2} = \sum_{t=1}^{T} X_t' \cdot \frac{\partial \Omega^{-1}}{\partial \sigma^2} \cdot \varepsilon_t$$

The generic element of the above matrix is a $k \times 1$ vector:

$$-\sigma_1^{-4} \cdot \sum_{t=1}^{T} X_{1,t}' \cdot \varepsilon_{i,t}.$$

Going on with the calculation:

$$\mathscr{H}_{i,i|i\in[24,38]} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial(\sigma_i^2)^2} = \frac{T}{2}\cdot\frac{1}{\sigma_i^4}\cdot\left(1-\frac{2}{T\cdot\sigma_i^2}\cdot\sum_{t=1}^{T}\varepsilon_{i,t}^2\right).$$

$$\mathscr{H}_{23+i,23+j|i,j\in[1,n]} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial\sigma_i^2\partial\sigma_j^2} = 0 \quad \forall i \neq j.$$

$$\begin{aligned}
\mathscr{H}_{1,24:1,38} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial\rho_1\partial\sigma_i^2} &= \frac{\partial}{\partial\sigma_i^2}\Big(\sum_{t\in t_1}^{T_1}\varepsilon_t'\cdot\Omega^{-1}\cdot W_1\cdot\Delta y_t\Big) \\
&= \frac{\partial}{\partial\sigma_i^2}\Big(\sum_{t\in t_1}^{T_1} Tr\big(\varepsilon_t'\cdot\Omega^{-1}\cdot W_1\cdot\Delta y_t\big)\Big) \\
&= \frac{\partial}{\partial\sigma_i^2}\Big(Tr\big((\sum_{t\in t_1}^{T_1}\Delta y_t\cdot\varepsilon_t')\cdot\Omega^{-1}\cdot W_1\big)\Big) \\
&= Tr\Big((\sum_{t\in t_1}^{T_1}\Delta y_t\cdot\varepsilon_t')\cdot\frac{\partial\Omega^{-1}}{\partial\sigma_i^2}\cdot W_1\Big)
\end{aligned}$$

Note that

$$\frac{\partial\Omega^{-1}}{\partial\sigma_i^2} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -\sigma_i^{-4} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} = diag(0,\cdots,0,-\sigma_i^{-4},0,\cdots,0)$$

Symmetrically:

$$\mathscr{H}_{2,24:2,38} = \frac{\partial^2 \log \mathscr{L}(\theta|\Delta y, X)}{\partial\rho_2\partial\sigma_i^2} = Tr\Big((\sum_{t\in t_2}^{T_2}\Delta y_t\cdot\varepsilon_t')\cdot\frac{\partial\Omega^{-1}}{\partial\sigma_i^2}\cdot W_2\Big)$$

At this point we have all the elements to construct the Hessian matrix of the log-likelihood.

To sum up, first row:

- $\mathscr{H}_{1,1} = -T_1\cdot Tr\big(W_1\cdot H_\tau\cdot W_1\cdot H_\tau\big) - \sum_{t\in t_1}^{T_1}\big(\Delta y_t'\cdot W_1'\cdot\Omega^{-1}\cdot W_1\cdot\Delta y_t\big)$

- $\mathscr{H}_{1,2} = 0$

- $\mathscr{H}_{1,3:1,23} = -\sum_{t\in t_1}^{T_1} X_t' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t$

- $\mathscr{H}_{1,24:1,38} = Tr\Big(\big(\sum_{t\in t_1}^{T_1} \Delta y_t \cdot \varepsilon_t'\big) \cdot \dfrac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_1\Big).$

Second row:

- $\mathscr{H}_{2,1} = 0$

- $\mathscr{H}_{2,2} = -T_2 \cdot Tr\big(W_2 \cdot H_\gamma \cdot W_2 \cdot H_\gamma\big) - \sum_{t\in t_2}^{T_2} \big(\Delta y_t' \cdot W_2' \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t\big)$

- $\mathscr{H}_{2,3:2,23} = -\sum_{t\in t_2}^{T_2} X_t' \cdot \Omega^{-1} \cdot W_2 \cdot \Delta y_t$

- $\mathscr{H}_{2,24:2,38} = Tr\Big(\big(\sum_{t\in t_2}^{T_2} \Delta y_t \cdot \varepsilon_t'\big) \cdot \dfrac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_2\Big).$

From row 3 to row 23:

- $\mathscr{H}_{3,1:23,1} = \mathscr{H}'_{1,3:1,23}$

- $\mathscr{H}_{3,2:23,2} = \mathscr{H}'_{2,3:2,23}$

- $\mathscr{H}_{3,3:23,23} = \sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot X_t$

- $\mathscr{H}_{3,24:23,38} = \sum_{t=1}^{T} X_t' \cdot \dfrac{\partial\Omega^{-1}}{\partial\sigma^2} \cdot \varepsilon_t$

From row 24 to the last row (number 38):

- $\mathscr{H}_{24,1:38,1} = \mathscr{H}'_{1,24:1,38}$

- $\mathscr{H}_{24,2:38,2} = \mathscr{H}'_{2,24:2,38}$

- $\mathscr{H}_{24,3:38,23} = \mathscr{H}'_{3,24:23,38}$

- $\mathscr{H}_{23+i,23+j|i,j\in[1,n]} = \begin{cases} \dfrac{T}{2} \cdot \dfrac{1}{\sigma_i^4} \cdot \left(1 - \dfrac{2}{T \cdot \sigma_i^2} \cdot \sum\limits_{t=1}^{T} \varepsilon_{i,t}^2\right) & \forall i = j \in [1, n] \\ \\ 0 \quad \forall i \neq j \end{cases}$

The last step we have to make to finally obtain the Fisher Information Matrix is taking expectations of every element.

$$\begin{aligned}
E[\mathscr{H}_{1,1}] &= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \sum_{t\in t_1}^{T_1} E\big[\Delta y_t' \cdot W_1' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t\big] = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \sum_{t\in t_1}^{T_1} E\big[Tr\big(W_1 \cdot \Delta y_t \cdot \Delta y_t' \cdot W_1' \cdot \Omega^{-1}\big)\big] = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \sum_{t\in t_1}^{T_1} Tr\big(W_1 \cdot E\big[\Delta y_t \cdot \Delta y_t'\big] \cdot W_1' \cdot \Omega^{-1}\big) = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \sum_{t\in t_1}^{T_1} Tr\big(W_1 \cdot E\big[H_\tau \cdot X_t \cdot \beta \cdot \varepsilon_t' \cdot H_\tau' + \\
&+ H_\tau \cdot X_t \cdot \beta \cdot \beta' \cdot X_t' \cdot H_\tau' + H_\tau \cdot \varepsilon_t \cdot \varepsilon_t' \cdot H_\tau' \cdot \varepsilon_t \cdot \beta' \cdot X_t' \cdot H_\tau'\big] \cdot W_1' \cdot \Omega^{-1}\big) = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \sum_{t\in t_1}^{T_1} Tr\Big(W_1 \cdot \Big[H_\tau \cdot X_t \cdot \beta \cdot E[\varepsilon_t'] \cdot H_\tau' + \\
&+ H_\tau \cdot X_t \cdot \beta \cdot \beta' \cdot X_t' \cdot H_\tau' + H_\tau \cdot E[\varepsilon_t \cdot \varepsilon_t'] \cdot H_\tau' + E[\varepsilon_t] \cdot \beta' \cdot X_t' \cdot H_\tau'\Big] \cdot W_1' \cdot \Omega^{-1}\Big) = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \\
&- \sum_{t\in t_1}^{T_1} Tr\Big(W_1 \cdot \Big[H_\tau \cdot X_t \cdot \beta \cdot \beta' \cdot X_t' \cdot H_\tau' + H_\tau \cdot \Omega \cdot H_\tau'\Big] \cdot W_1' \cdot \Omega^{-1}\Big) = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau\big) - \\
&- \sum_{t\in t_1}^{T_1} Tr\big(W_1 \cdot H_\tau \cdot X_t \cdot \beta \cdot \beta' \cdot X_t' \cdot H_\tau' \cdot W_1' \cdot \Omega^{-1} + W_1 \cdot H_\tau \cdot \Omega \cdot H_\tau' \cdot W_1' \cdot \Omega^{-1}\big) = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau + H_\tau' \cdot W_1' \cdot \Omega^{-1} \cdot W_1 \cdot H_\tau \cdot \Omega\big) - \\
&- \sum_{t\in t_1}^{T_1} Tr\big(\beta' \cdot X_t' \cdot H_\tau' \cdot W_1' \cdot \Omega^{-1} \cdot W_1 \cdot H_\tau \cdot X_t \cdot \beta\big) =
\end{aligned}$$

Setting $M_1^\tau = H_\tau' \cdot W_1' \cdot \Omega^{-1} \cdot W_1 \cdot H_\tau$ we can rewrite the above identity as:

$$
\begin{aligned}
E[\mathscr{H}_{1,1}] &= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau + M_1^\tau \cdot \Omega\big) - \sum_{t \in t_1}^{T_1} \beta' \cdot X_t' \cdot M_1^\tau \cdot X_t \cdot \beta = \\
&= -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau + M_1^\tau \cdot \Omega\big) - \beta' \cdot X_\tau' \cdot \big(I_{T_1} \otimes M_1^\tau\big) \cdot X_\tau \cdot \beta.
\end{aligned}
$$

Simmetrically:

$$
E[\mathscr{H}_{2,2}] = -T_2 \cdot Tr\big(W_2 \cdot H_\gamma \cdot W_2 \cdot H_\gamma + M_1^\gamma \cdot \Omega\big) - \beta' \cdot X_\gamma' \cdot \big(I_{T_2} \otimes M_1^\gamma\big) \cdot X_\gamma \cdot \beta.
$$

with $M_1^\gamma = H_\gamma' \cdot W_2' \cdot \Omega^{-1} \cdot W_2 \cdot H_\gamma$.

Going on with the calculation:

$$
\begin{aligned}
E[\mathscr{H}_{1,3:1,23}] &= E\Big[- \sum_{t \in t_1}^{T_1} X_t' \cdot \Omega^{-1} \cdot W_1 \cdot \Delta y_t \Big] = \\
&= - \sum_{t \in t_1}^{T_1} X_t' \cdot \Omega^{-1} \cdot W_1 \cdot E\Big[H_\tau \cdot X_t \cdot \beta + H_\tau \cdot \varepsilon_t\Big] = \\
&= - \sum_{t \in t_1}^{T_1} X_t' \cdot \Omega^{-1} \cdot W_1 \cdot H_\tau \cdot X_t \cdot \beta \\
&= X_\tau' \cdot (I_{T_1} \otimes M_2^\tau) \cdot X_\tau \cdot \beta
\end{aligned}
$$

with $M_2^\tau = \Omega^{-1} \cdot W_1 \cdot H_\tau$.

Simmetrically:

$$
E[\mathscr{H}_{2,3:2,23}] = X_\gamma' \cdot (I_{T_2} \otimes M_2^\gamma) \cdot X_\gamma \cdot \beta
$$

with $M_2^\gamma = \Omega^{-1} \cdot W_2 \cdot H_\gamma$.

Next step:

$$
\begin{aligned}
E[\mathscr{H}_{1,24:1,38}] &= Tr\Big(\big(\sum_{t\in t_1}^{T_1} E\big[\Delta y_t \cdot \varepsilon_t'\big]\big) \cdot \frac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_1\Big) = \\
&= Tr\Big(\big(\sum_{t\in t_1}^{T_1} E\big[\Delta y_t \cdot \varepsilon_t'\big]\big) \cdot \frac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_1\Big) = \\
&= Tr\Big(\big(\sum_{t\in t_1}^{T_1} H_\tau \cdot E\big[\varepsilon_t \cdot \varepsilon_t'\big]\big) \cdot \frac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_1\Big) = \\
&= T_1 \cdot Tr\Big(H_\tau \cdot \Omega \cdot \frac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_1\Big) = \\
&= T_1 \cdot Tr\Big(\Omega \cdot \frac{\partial\Omega^{-1}}{\partial\sigma_i^2} \cdot W_1 \cdot H_\tau\Big),
\end{aligned}
$$

Notice that

$$
\Omega \cdot \frac{\partial\Omega^{-1}}{\partial\sigma_i^{-2}} = -\sigma_i^2 \cdot I_{ii}
$$

where the generic element of matrix I_{ii} is given by

$$
\omega_{s,t} = \begin{cases} 1 & s = i, j = i \\ 0 & \text{otherwise} \end{cases}
$$

Therefore

$$
\begin{aligned}
E[\mathscr{H}_{1,23+i}] &= T_1 \cdot \sigma_i^{-2} \cdot Tr\Big(I_{ii} \cdot W_1 \cdot H_\tau\Big) = \\
&= T_1 \cdot \sigma_i^{-2} \cdot \Big(W_1 \cdot H_\tau\Big)_{ii}
\end{aligned}
$$

Finally we have that:

$$
E[\mathscr{H}_{1,24:1:38}] = T_1 \cdot diag\Big(\Omega^{-1} \cdot W_1 \cdot H_\tau\Big) = T_1 \cdot diag(M_2^\tau).
$$

Simmetrically:

$$
E[\mathscr{H}_{2,24:2:38}] = T_2 \cdot diag\Big(\Omega^{-1} \cdot W_2 \cdot H_\gamma\Big) = T_2 \cdot diag(M_2^\gamma).
$$

Going on:

$$E[\mathscr{H}_{3,3:23,23}] = E[\sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot X_t] = \sum_{t=1}^{T} X_t' \cdot \Omega^{-1} \cdot X_t = X' \cdot \Sigma^{-1} \cdot X$$

$$\begin{aligned} E[\mathscr{H}_{3,24:23,38}] &= E[\sum_{t=1}^{T} X_t' \cdot \frac{\partial \Omega^{-1}}{\partial \sigma^2} \cdot \varepsilon_t] \\ &= \sum_{t=1}^{T} X_t' \cdot \frac{\partial \Omega^{-1}}{\partial \sigma^2} \cdot E[\varepsilon_t] \\ &= \underset{k \times n}{\mathbf{0}} \end{aligned}$$

$$E[\mathscr{H}_{23+i,23+j|i,j\in[1,n]}] = \begin{cases} \frac{T}{2} \cdot \frac{1}{\sigma_i^4} \cdot \left(1 - \frac{2}{T \cdot \sigma_i^2} \cdot \sum_{t=1}^{T} E[\varepsilon_{i,t}^2]\right) & \forall i = j \in [1,n] \\ \\ 0 \quad \forall i \neq j \end{cases}$$

$$= \begin{cases} -\frac{T}{2} \cdot \frac{1}{\sigma_i^4} & \forall i = j \in [1,n] \\ 0 \quad \forall i \neq j \end{cases}$$

$$= -\frac{T}{2} \cdot \begin{bmatrix} \sigma_1^{-4} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-4} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ o & 0 & \cdots & \sigma_n^{-4} \end{bmatrix} = -\frac{T}{2} \cdot V$$

We finally have all the elements of the Fisher Information Matrix for our panel (with dummy variables) spatial model:

$$\mathscr{I} =$$

$$\begin{bmatrix} \begin{matrix} -T_1 \cdot Tr\big(W_1 \cdot H_\tau \cdot W_1 \cdot H_\tau + M_1^\tau \cdot \Omega\big) - \\ -\beta' \cdot X_\tau' \cdot \big(I_{T_1} \otimes M_1^\tau\big) \cdot X_\tau \cdot \beta \end{matrix} & 0 & \big(X_\tau' \cdot (I_{T_1} \otimes M_2^\tau) \cdot X_\tau \cdot \beta\big)' & T_1 \cdot diag(M_2^\tau)' \\ 0 & \begin{matrix} -T_2 \cdot Tr\big(W_2 \cdot H_\gamma \cdot W_2 \cdot H_\gamma + M_1^\gamma \cdot \Omega\big) - \\ -\beta' \cdot X_\gamma' \cdot \big(I_{T_2} \otimes M_1^\gamma\big) \cdot X_\gamma \cdot \beta \end{matrix} & \big(X_\gamma' \cdot (I_{T_2} \otimes M_2^\gamma) \cdot X_\gamma \cdot \beta\big)' & T_2 \cdot diag(M_2^\gamma)' \\ X_\tau' \cdot (I_{T_1} \otimes M_2^\tau) \cdot X_\tau \cdot \beta & X_\gamma' \cdot (I_{T_2} \otimes M_2^\gamma) \cdot X_\gamma \cdot \beta & X' \cdot \Sigma^{-1} \cdot X & \underset{k\times n}{\mathbf{0}} \\ T_1 \cdot diag(M_2^\tau) & T_2 \cdot diag(M_2^\gamma) & \underset{n\times k}{\mathbf{0}} & -\frac{T}{2} \cdot V \end{bmatrix}$$

C.3 Bayesian MCMC - Technical Details

Even if the MLE is a common standard method in spatial econometric applications, we have two valid reasons for not adopting it: 1. non-stationary estimates of aggregate total effects; 2. prior information on the values of the parameters. Let's explore both the issues.

1. Non-Stationary Solutions

We can estimate the parameters by maximizing the concentrated log-likelihood over the compact set which guarantees a positive definite matrix (see Ord (1975)): $C^{down} = \left(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right)$ and $C^{up} = \left(\mu_{\min}^{-1}, \mu_{\max}^{-1}\right)$. The standard errors are constructed using the analytical Fisher Information of the model, centered on the point estimates, $\hat{\rho}^{down}$ and $\hat{\rho}^{up}$. The asymptotic results of Yu, De-Jong, and Lee (2008) guarantees the asymptotic normality of the parameters of equation (6) and (7) (See Theorem 3 case $n/T \to 0$). For instance, for the estimator of ρ^{down} we have:

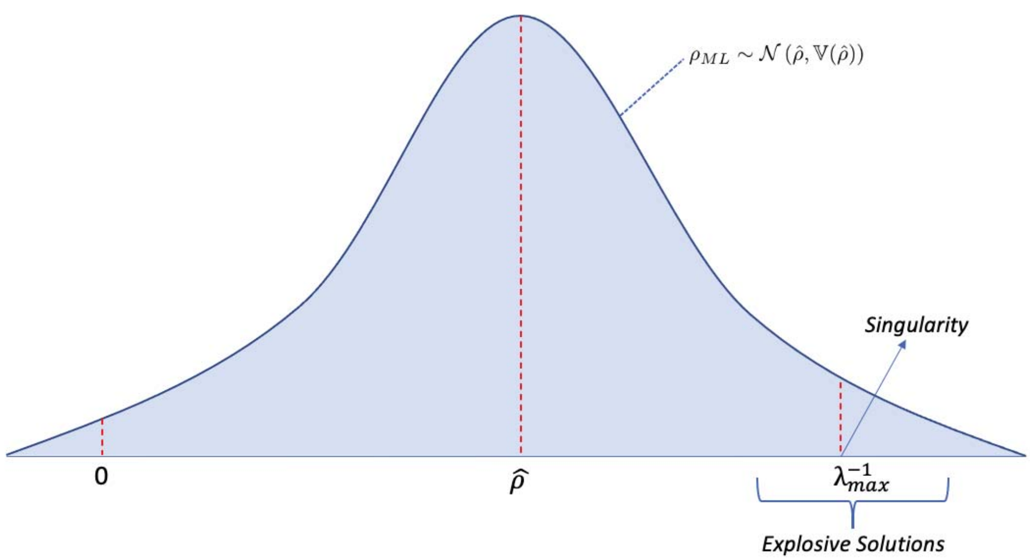
$$\sqrt{T \cdot n}\left(\hat{\rho}_{nT}^{down} - \rho^{down}\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

where σ^2 is the asymptotic variance of the MLE, obtained by the calculating the analytical Fisher Information matrix of our model. However, we are interested in estimating the aggregate total effect of fiscal consolidations, not the parameters of the model themselves. At page 70, LeSage and Pace (2009) suggest to construct the asymptotic distribution of the average total effect (our aggregate total effect) by following these steps: 1. estimate the parameters of the model via MLE; 2. Draw values of the parameters by their approximate asymptotic distribution ($\tilde{\rho}^{down} \approx \mathcal{N}\left(\hat{\rho}_{nT}^{down}, \frac{\hat{\sigma}^2(\hat{\rho}_{nT}^{down})}{nT}\right)$; 3. Calculate at each step the aggregate total effect. After doing so we calculated the standard errors of the ATE_{TB} by calculating the standard deviation of the asymptotic distribution so constructed. We obtained explosive solutions. This is a surprising result, in fact, the asymptotic normality of the average effect is guaranteed by the Δ-method:

$$\sqrt{T \cdot n}\left(ATE_{TB}(\hat{\rho}_{nT}^{down}) - ATE_{TB}(\rho^{down})\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \cdot \left(\frac{\partial ATE_{TB}(\rho^{down})}{\partial \rho^{down}}\right)^2\right)$$

where $ATE_{TB} : C^{down} \to \mathbb{R}$ and $ATE_{TB}(x) = v' \cdot (I_n - x \cdot A)^{-1} \cdot \omega_{TB}$ and v is a vector of industry output shares of total industrial production (the weights we use to calculate the aggregate effect of fiscal consolidations). What goes wrong in this procedure? The Δ-method is an asymptotic result, which might provide a terrible approximation of a finite sample distribution. It all boils down in finding a distribution which approximates well the small sample one. If $\hat{\rho}_{nT}^{down}$ is very closed to the boundary and its asymptotically normal standard errors are large, that is, they approach the boundary of C^{down} then we end up drawing values of ρ^{down} which deliver unrealistically large values of ATE_{TB}, because matrix $(I_n - \rho^{down} \cdot A)^{-1}$ becomes singular (the boundary is one eigenvalue of A). This situation is described in Figure 13.

Figure 13: Explosive Solutions of ATE_{TB}



2. Prior Information

We have two extra "prior" pieces of information on the value of the spatial parameters, ρ^{down} and ρ^{up}:

i. Values of ρ^{down} and ρ^{up} close to the boundaries will deliver unrealistically high values of ATE, ADE and ANE, since the determinant of matrices $(I_n - \rho^{down} \cdot A)$ and $(I_n - \rho^{up} \cdot \hat{A}^T)$ will approach zero by definition of eigenvalue. In turn, the elements of their inverse matrices will explode, as illustrated above. Therefore, we should assign less weight to values of ρ^{down} and ρ^{up} close to the boundaries.

ii. We know that industries that are close to each other in the production network will co-move. For instance, if industry X faces increasing prices for its input, it will shrink production and increase prices; in turn, customers of X will also face the same problem and will react similarly, by reducing production and increasing prices. Therefore, the direction of the spatial correlation among industries' output is positive: $\rho^{down} > 0$ and $\rho^{up} > 0$.

Model Estimation

We can integrate such prior information into our estimation and avoid non-stationarity aggregate effects, by adopting a Bayesian MCMC similar to the one introduced by LeSage and Parent (2007). We illustrate here how we implement the Bayesian MCMC to estimate the parameters of Equation (6) (baseline). The log-likelihood of that model is the one outlined above. The priors we employ on the parameters are:

$$
\begin{aligned}
&\pi(\beta) \propto constant \\
&\Omega = \sigma^2 \cdot V \quad \text{with } V = diag(v_1, ..., v_n) \\
&\pi(\sigma^2) \propto \frac{1}{\sigma^2} \\
&\pi(v_i) \overset{iid}{\sim} \Gamma^{-1}\left(\frac{r}{2}, \frac{r}{2}\right), \quad i = 1, ..., n \\
&\rho^{down} \sim Gen.Beta(d, d) \\
&\rho^{up} \sim Gen.Beta(d, d).
\end{aligned}
$$

We adopt non-informative priors for σ^2 and β to reflect our lack of information around the values of these parameters. Concerning r, a lower value generates more diffusion in the distributions of v_i, thus regulating our confidence towards heteroskedasticity. Unlike LeSage and Pace (2009), who suggest a value of 4, we set r equal to 3 to reflect a strong belief towards heteroskedasticity. For instance, industries in the Agriculture (NAICS 11) as well as Mining (NAICS 21) macro sectors, exhibit much higher volatilities than the rest of the industries.

We impose a "generalized (or non-standardized) $Beta(d, d)$ prior", with support from 0 to λ_{max}^{-1} for ρ^{down} and from 0 to $\hat{\lambda}_{max}^{-1}$ for ρ^{up}. We follow LeSage and Pace (2009) and set d equal to 1.1; which has the benefit of letting the generalized Beta prior to resemble a Uniform distribution (diffuse prior), but with low density at the boundaries, as illustrated in Figure 14. The choice of such a prior allows us to be agnostic about the specific value of the spatial parameters but at the same time it allows to embed the prior information we

Figure 14: Generalized Beta prior

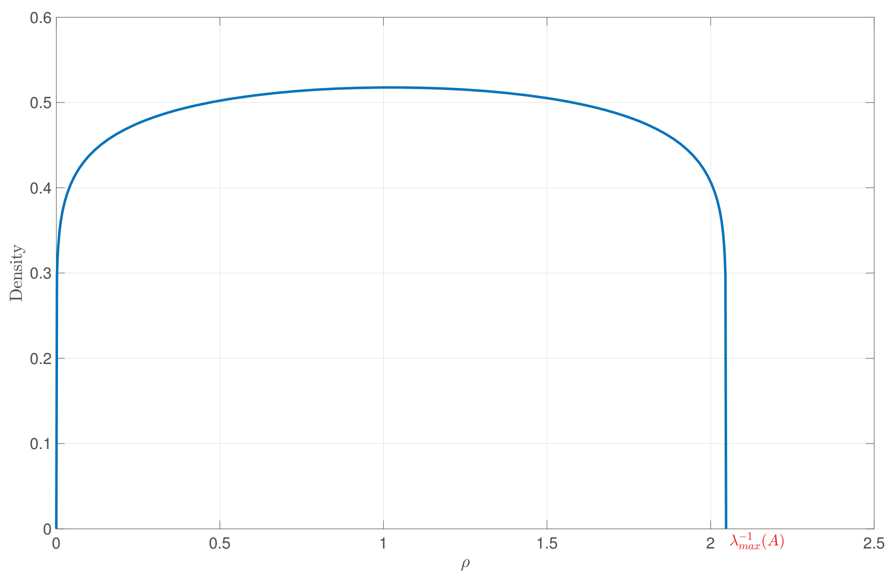


Figure 14: line-plot of a non-standardized $Beta(1.1, 1.1)$ *density function, with support from* $(0, \lambda^{-1}_{max}(A) = 2.047)$ *which we employ as a prior for the spatial parameter* ρ^{down}.

have into their estimates.
Furthermore, we assume that all the prior distributions are independent from each other. We use the standard "Metropolis within Gibbs" algorithm, and we obtain an approximation of the posterior densities for each parameter of the model.
We now outline the precise steps of the procedure:

1. **Initialization:** Set up initial values for the parameters: $\beta_{(0)}, \sigma^2_{(0)}, V_{(0)}, \rho^{down}_{(0)}, \rho^{up}_{(0)}$, where $V_{(0)} = diag\left(v^2_{1,(0)}, ..., v^2_{n,(0)}\right)$.

2. **Gibbs Sampling:**

   a) Draw $\beta_{(1)}$ from the conditional posterior distribution, which is obtained by mixing the likelihood with a normal prior with mean c (a vector of zeros in our simulation) and covariance matrix L. In order to not add any information, we simply set L to be equal to a diagonal matrix whose entries are infinite (1e12 in our simulation):

$$P(\beta_{(0)}|\mathscr{D}, \sigma^2_{(0)}, V_{(0)}, \rho^{down}_{(0)}, \rho^{up}_{(0)}) = \mathscr{N}(c^*, L^*) \propto \mathscr{L}(\theta|\mathscr{D}) \cdot \mathscr{N}(c, L)$$

$$c^* = \frac{1}{T} \cdot (\sum_{t=1}^{T} X'_t \cdot V^{-1}_{(0)} \cdot X_t + \frac{\sigma^2_{(0)}}{T} \cdot L^{-1})^{-1} \cdot (\frac{1}{T} \cdot \sum_{t=1}^{T} X'_t \cdot V^{-1}_{(0)} \cdot H_t \cdot \Delta y_t + \frac{\sigma^2_{(0)}}{T} \cdot L^{-1} \cdot c)$$

$$L^* = \frac{\sigma^2_{(0)}}{T} \cdot (\sum_{t=1}^{T} X'_t \cdot V^{-1}_{(0)} \cdot X_t + \frac{\sigma^2_{(0)}}{T} \cdot L^{-1})^{-1}$$

b) Draw $\sigma^2_{(1)}$ from the conditional posterior distribution, which is proportional to likelihood times an inverse gamma distribution as a prior:

$$P(\sigma^2_{(1)}|\mathscr{D}, \beta_{(1)}, V_{(0)}, \rho^{down}_{(0)}, \rho^{up}_{(0)}) = \Gamma^{-1}(\frac{\theta_1}{2}, \frac{\theta_2}{2}) \propto \mathscr{L}(\theta|\mathscr{D}) \cdot \Gamma^{-1}(a, b)$$

$$\theta_1 = nT + 2a \quad \theta_2 = \sum_{t=1}^{T} \varepsilon'_t \cdot V^{-1}_{(0)} \cdot \varepsilon_t + 2b$$

In practice we draw $\sigma^2_{(1)}$ from θ_2/χ_{θ_1}.
Notice that, setting a and b (the prior's parameters) equal to 0, is like putting a Jefferey's prior on σ^2. This is exactly what we do.

c) Draw $v_{i,(1)}$ from the following conditional posterior distribution, proportional to an inverse gamma prior:

$$P(v_{i,(1)}|\mathscr{D}, \sigma^2_{(1)}, \rho^{down}_{(0)}, \rho^{up}_{(0)}) = \Gamma^{-1}(\frac{q_1}{2}, \frac{q_2}{2}) \propto \mathscr{L}(\theta|\mathscr{D}) \cdot \Gamma^{-1}(\frac{r}{2}, \frac{r}{2})$$

$$q_1 = r + T \quad q_2 = \frac{1}{\sigma^2_{(1)}} \cdot \sum_{t=1}^{T} \varepsilon^2_{i,t} + r$$

In practice we draw $v_{i,(1)}$ from q_2/χ_{q_1}.
As anticipated above in the paper, since we are confident on the heteroskedastic behavior of industry value added, we set our prior hyperparameter r to be equal to 3 rather than 4, as done in LeSage and Pace (2009).
Replicating this procedure n times, we get a first simulation of matrix $V_{(1)}$.

3. **Metropolis-Hastings**: We now need to draw the spatial coefficients. However we cannot apply a simple Gibbs Sampling, since the conditional posterior distribution is not defined for them. LeSage and Pace (2009)

suggest the adoption of the Metropolis-Hastings algorithm to overcome this problem. To ease notation we set $\rho_1 := \rho^{down}$ and $\rho_2 := \rho^{up}$. The algorithm is the following:

(a) Draw ρ_1^c (where the c superscript stands for "candidate") from the (random walk) proposal distribution:

$$\rho_1^c = \rho_{1,(0)} + c_1 \cdot \mathcal{N}(0,1)$$

(b) Run a bernoulli experiment to determine the updated value of ρ_1:

$$\rho_{1,(1)} = \begin{cases} \rho_1^c & \pi \quad \text{(accept)} \\ \rho_{1,(0)} & 1-\pi \quad \text{(reject)} \end{cases}$$

Where π is equal to $\pi = \min\{1, \psi_{MH_1}\}$ and, setting: $A_\tau(\rho_1) = I_n - \rho_1 \cdot W_1$, we have:

$$\begin{aligned} \psi_{MH_1} = \frac{|A_\tau(\rho_1^c)|}{|A_\tau(\rho_{1,(0)})|} \cdot \exp\Big\{ -\frac{1}{2\sigma^2_{(1)}} \cdot \sum_{t\in t_1}^{T_1} \Big[\Delta y_t' \cdot \big(A_\tau(\rho_1^c)' \cdot V_{(1)}^{-1} \cdot A_\tau(\rho_1^c) - \\ - A_\tau(\rho_{1,(0)})' \cdot V_{(1)}^{-1} \cdot A_\tau(\rho_{1,(0)})\big) \cdot \Delta y_t - \\ - 2\beta' \cdot X_t' \cdot V_{(1)}^{-1}\big(A_\tau(\rho_1^c) - A_\tau(\rho_{1,(0)})\big) \cdot \Delta y_t \Big] \Big\} \cdot \\ \cdot \left[\frac{(\rho_1^c - 0)\cdot(\lambda_{max}^{-1} - \rho_1^c)}{(\rho_{1,(0)} - 0)\cdot(\lambda_{max}^{-1} - \rho_{1,(0)})} \right]^{d-1} \cdot \mathbf{1}\big(0 \le \rho_1^c \le \lambda_{max}^{-1}\big) \end{aligned}$$

Basically, we compute the probability to accept the candidate value from the proposal distribution, and then we update the value of ρ_1 by running the bernoulli experiment with such a probability of success. Notice that if we draw a value of ρ_1 outside the support of the beta prior, $\psi_{MH_1} = 0$ and then $\pi = 0$ and we clearly reject the candidate value.
We set d equal to 1.1, on both ρ_1 and ρ_2; this is done to resemble a Uniform (0,1) but with less density on its boundary values.

(c) Once updated ρ_1, we replicate the procedure for ρ_2. Setting $A_\gamma(\rho_2) =$

$I_n - \rho_2 \cdot W_2$ we have:

$$\psi_{MH_2} = \frac{|A_\gamma(\rho_2^c)|}{|A_\gamma(\rho_{2,(0)})|} \cdot \exp\Big\{ - \frac{1}{2\sigma^2_{(1)}} \cdot \sum_{t \in t_2}^{T_2} \Big[\Delta y_t' \cdot \big(A_\gamma(\rho_2^c)' \cdot V_{(1)}^{-1} \cdot A_\gamma(\rho_2^c) -$$
$$- A_\gamma(\rho_{2,(0)})' \cdot V_{(1)}^{-1} \cdot A_\gamma(\rho_{2,(0)}) \big) \cdot \Delta y_t -$$
$$- 2\beta' \cdot X_t' \cdot V_{(1)}^{-1} \big(A_\gamma(\rho_2^c) - A_\gamma(\rho_{2,(0)}) \big) \cdot \Delta y_t \Big] \Big\} \cdot$$
$$\cdot \left[\frac{(\rho_2^c - 0) \cdot (\hat{\lambda}_{max}^{-1} - \rho_2^c)}{(\rho_{2,(0)} - 0) \cdot (\hat{\lambda}_{max}^{-1} - \rho_{2,(0)})} \right]^{d-1} \cdot \mathbf{1}\big(0 \leq \rho_2^c \leq \hat{\lambda}_{max}^{-1} \big)$$

(d) At this point we need to update the variance of the proposal distributions: if the acceptance rate (number of acceptances over number of iterations of the Markov Chain) of the first parameter ρ_1 falls below 40% we need to reduce the value of c_1, the so called tuning parameter, which regulates the variance of the proposal distribution. The variance is reduced by rescaling it: $c_1' = \frac{c_1}{1.1}$. In this way, we are able to draw values closer to the current state of ρ_1, and therefore, we expect to increase the acceptance rate.
On the contrary, if the acceptance rate rises above 60%, we need to increase the tuning parameter, in order to draw values far from the current state, in this way we increase the chance to explore more the low-density parts of the distribution. We increase the variance of the candidate distribution by scaling upward its standar deviation: $c_1' = 1.1 \cdot c_1$.
Clearly we replicate this procedure also for ρ_2.

4. **Repeat**: Once updated all the values, we replicate steps 2 and 3, 45,000 times to make sure the acceptance rate has converged.

5. **Burn-in**: we drop the first 35,000 iterations of the Markov Chain, thus obtaining a vector of 10,000 observations for each of the parameters, which account for the simulated posterior distributions.

C.4 Simulating the ATE, ADE and ANE

We construct via Monte Carlo the distribution of the ATE, ADE and ANE. In particular we follow these steps:

1. **(Parameters)** Draw ρ^{down}, ρ^{up}, $\boldsymbol{\tau}$ and $\boldsymbol{\gamma}$ from their posterior distributions. To take into account the potential correlation among them, draw from the same iteration of the Bayesian MCMC.

2. **(Style of the plan)** Construct both a TB and an EB simulated fiscal plan, by drawing the style from a distribution which mimics the empirical one.

3. **(Average effects)** Construct ATE, ADE and ANE using the parameters drawn in step 1 and the style drawn in step 2.

4. Repeat 100,000 times steps from 1 though 3, to make sure all the possible combination of styles and parameters are simulated.

Step 2 allows us to claim that the baseline results reported in the paper are robust to different styles of fiscal plans.

Empirical distribution of style of fiscal plans

We are interested in simulating a 2 years fiscal consolidation made of an unexpected part, no announced part and a single year future part to be implemented in the second year of the simulation.
First of all, we want to simulate the unexpected part of the fiscal plan, therefore, we need to look at those years when an unanticipated shock occurs. Define the two sub-samples: $TB^u := \{t : 1, ..., T \mid tax_t^u > 0\}$ and $EB^u := \{t : 1, ..., T \mid exp_t^u > 0\}$. Then calculate the mean and the standard deviation of the unexpected component conditional on the occurrence of an unexpected shock:

$$\mu_\tau := \mathbb{E}(tax_t^u \mid t \in TB^u) \qquad \sigma_\tau := \sqrt{\mathbb{V}(tax_t^u \mid t \in TB^u)}$$
$$\mu_\gamma := \mathbb{E}(exp_t^u \mid t \in EB^u) \qquad \sigma_\gamma := \sqrt{\mathbb{V}(exp_t^u \mid t \in EB^u)}$$

In order to simulate a plausible unexpected component of the plan, we draw them from the following distributions:

$$\tilde{tax}^u \sim \mathscr{U}(\mu_\tau - \sigma_\tau, \mu_\tau + \sigma_\tau)$$
$$\tilde{exp}^u \sim \mathscr{U}(\mu_\gamma - \sigma_\gamma, \mu_\gamma + \sigma_\gamma)$$

where the $\tilde{}$ denotes a simulated component.
Concerning the future component, we need to predict what is the value of a one year ahead policy change, conditional on the occurrence of an unexpected policy change. Therefore, we run the following regressions:

$$tax_{t,1}^f = a_\tau + b_\tau \cdot tax_t^u \qquad \textit{with: } t \in TB^u$$
$$exp_{t,1}^f = a_\gamma + b_\gamma \cdot exp_t^u \qquad \textit{with: } t \in EB^u$$

The estimates of a_τ, b_τ, a_γ, b_γ will be stored and used to predict values of $tax^f_{t,1}$ and $exp^f_{t,1}$, conditional on the occurrence of an unexpected component. At this point we have all the ingredients to outline the steps we do in the construction of a simulated style of the plan:

1. Draw unexpected components from their candidate distributions: $\tilde{tax}^u \sim \mathscr{U}(\mu_\tau - \sigma_\tau, \mu_\tau + \sigma_\tau)$ and $\tilde{exp}^u \sim \mathscr{U}(\mu_\gamma - \sigma_\gamma, \mu_\gamma + \sigma_\gamma)$.
2. Predict the future component using the estimates of a_τ, b_τ, a_γ, b_γ. We have: $\tilde{tax}^f = \hat{a}_\tau + \hat{b}_\tau \cdot \tilde{tax}^u$ and $\tilde{exp}^f = \hat{a}_\gamma + \hat{b}_\gamma \cdot \tilde{exp}^u$.
3. Normalize the value to one: $\tilde{tax}^u + \tilde{tax}^f = 1$ and $\tilde{exp}^u + \tilde{exp}^f = 1$.

For each iteration of the MC simulation used to approximate the posterior distributions of the ATE, ADE and ANE, we repeat steps 1 through 3 to simulate the style of the plan.
In the first year of the simulation we calculate the effects of TB and EB plans with style given by: $\boldsymbol{s}_{TB} = [\tilde{tax}^u \; 0 \; \tilde{tax}^f]$ and $\boldsymbol{s}_{EB} = [\tilde{exp}^u \; 0 \; \tilde{exp}^f]$ respectively. In the second year of the simulation, the future component of the shock is rolled over and becomes an announced and implemented shock. Therefore we calculate the effects of TB and EB plans with style given by: $\boldsymbol{s}_{TB} = [0 \; \tilde{tax}^f \; 0]$ and $\boldsymbol{s}_{EB} = [0 \; \tilde{exp}^f \; 0]$ respectively.

D Estimates of Inverted Model

In this section we report the tables of estimates of the inverted model. Firstly, Table VII shows the estimates of Equation (7).

D.1 Model Selection - Vuong Test for Static Spatial Panel Data

We also provide results for a Vuong test of non-nested models, adapted to our spatial specification, as in Wooldridge (2010).

Firstly, the Vuong test (see Vuong (1989)) is meant to discriminate between two misspecified and non-nested models. Basically, we assume there is a hidden true model and we want to choose one of two competing non-nested models which fit the data equally well. The Vuong test calculates and compares the Kullback-Leibler distance between the two and the true model. In practice, is a t-test on the KL divergence. One problem we encounter is that it was developed for one-dimensional iid data, however, we deal with

Table VII: Estimation Results

Inverted Model - Equation (7)												
Parameters	MLE		Bayesian MCMC - Posterior Distributions:									
	$\hat{\theta}_i^{\text{ML}}$	MLE Std.	$\mathbb{E}(\theta_i)$	$\sqrt{\mathbb{V}(\theta_i)}$	$Pr(\theta_i < 0)$	5%	10%	16%	50%	84%	90%	95%
ρ^{up} (TB)	0.554	0.103	0.528	0.097	0.000	0.368	0.405	0.432	0.528	0.625	0.653	0.687
τ_u	0.684	1.283	0.815	1.193	0.247	-1.143	-0.712	-0.372	0.814	2.002	2.351	2.778
τ_a	-1.298	0.986	-1.290	0.919	0.920	-2.794	-2.463	-2.202	-1.293	-0.382	-0.112	0.225
τ_f	-0.080	0.426	-0.084	0.391	0.585	-0.726	-0.585	-0.474	-0.082	0.301	0.415	0.562
ρ^{down} (EB)	0.096	0.114	0.125	0.083	0.000	0.014	0.026	0.040	0.112	0.211	0.241	0.281
γ_u	0.073	1.126	0.050	1.034	0.480	-1.650	-1.272	-0.973	0.051	1.073	1.370	1.760
γ_a	1.286	0.617	1.296	0.567	0.011	0.361	0.572	0.732	1.295	1.861	2.023	2.226
γ_f	-0.502	0.282	-0.499	0.259	0.973	-0.923	-0.831	-0.757	-0.499	-0.241	-0.169	-0.075
D2008	-2.984	0.674	-2.934	0.633	1.000	-3.973	-3.744	-3.562	-2.936	-2.307	-2.120	-1.891
D2009	-5.710	0.674	-5.371	0.661	1.000	-6.469	-6.216	-6.025	-5.368	-4.717	-4.529	-4.290

Table VII: *θ_i denotes a generic parameter that we estimate. The columns report the following: $\hat{\theta}_i^{ML}$ is the ML point estimate; "MLE Std." is the standard deviation of the ML estimate, calculated using the analytical Fisher Information Matrix derived in Appendix C.2: $\sqrt{\mathscr{I}(\hat{\theta}^{ML})^{-1}_{ii}}$; $\mathbb{E}(\theta_i)$ is the expected value of the posterior distribution; $\sqrt{\mathbb{V}(\theta_i)}$ is the standard deviation of the posterior distribution; $Pr(\theta < 0)$ is the probability that a parameter is negative, calculated by integrating the posterior distribution; p% is the p-th percentile of the posterior distribution. For brevity we don't report here the Industry Fixed Effects and the Industry specific variances. In the first columns, the spatial parameters also report the type of fiscal plan they are interacted with (in blue).*

a panel whose observations are serially uncorrelated but spatially correlated. Wooldridge (2010) shows that the Vuong test can easily be extended to panel data models by accounting for serial correlation in the time series.[38] However, in our problem the $n \times 1$ vector of industry observations is iid over time and our asymptotic keeps the cross-sectional dimension, which is spatially correlated, fixed, and then let the time series to go to infinite $T \to \infty$. Economically speaking this makes sense: we observe those fixed 62 industries over time, however, the cross sectional dimension exceeds the times series one, 37 years. This means that our finite sample distribution will not be a very good approximation of the asymptotic one. However, this is the best we can do, given the data availability.

Let's derive now the Vuong Test. The quasi-log-likelihood of the baseline model, Equation (6), is:

$$\ell_{t,B}(\underbrace{\rho, \beta, \Omega}_{\theta_B}) = \log f_B(\Delta y_t | X_t; \theta_B) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2} \cdot \ln(|\Omega|) +$$
$$+ \ln(|I_n - \rho^{down} \cdot A \cdot TB_t - \rho^{up} \cdot \hat{A}' \cdot EB_t|) - \frac{1}{2} \cdot \varepsilon_t' \cdot \Omega^{-1} \cdot \varepsilon_t.$$

with:

$$\varepsilon_t = \Big(I_n - \rho^{down}\ A\ TB_t - \rho^{up}\ \hat{A}'\ EB_t\Big) \cdot \Delta y_t - X_t \cdot \beta.$$

[38]See Section 13.11.2 - Model Selection Tests.

The sum of the quasi-log-likelihood evaluated at the MLE, $\hat{\theta}_B$, for the baseline model is: $\mathscr{L}_B = \sum_{t=1}^{T} \ell_{t,B}(\hat{\theta}_B)$. Analogously, for the inverted model, Equation (7), the quasi-log-likelihood is:

$$\ell_{t,I}(\underbrace{\tilde{\rho}, \tilde{\beta}, \tilde{\Omega}}_{\theta_I}) = \log f_I(\Delta y_t | X_t; \theta_I) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \cdot \ln(|\tilde{\Omega}|) +$$

$$+ \ln(|I_n - \tilde{\rho}^{down} \cdot A \cdot EB_t - \tilde{\rho}^{up} \cdot \hat{A}' \cdot TB_t|) - \frac{1}{2} \cdot \varepsilon_t' \cdot \tilde{\Omega}^{-1} \cdot \varepsilon_t.$$

with:

$$\varepsilon_t = \Big(I_n - \tilde{\rho}^{down} \; A \; EB_t - \tilde{\rho}^{up} \; \hat{A}' \; TB_t \Big) \cdot \Delta y_t - X_t \cdot \tilde{\beta}.$$

The sum of the quasi-log-likelihood evaluated at the MLE, $\hat{\theta}_I$, for the inverted model is: $\mathscr{L}_I = \sum_{t=1}^{T} \ell_{t,I}(\hat{\theta}_I)$.
Following Wooldridge (2010), let's define the estimator for the variance of the KL divergence as:

$$\hat{\eta}^2 = \frac{1}{T} \cdot \sum_{t=1}^{T} \Big(\ell_{t,B}(\hat{\theta}_B) - \ell_{t,I}(\hat{\theta}_I) \Big)^2 .$$

Then, the Vuong Model Selection Statistic, VMS, is:

$$\begin{aligned} VMS &= T^{-1/2} \cdot \frac{(\mathscr{L}_B - \mathscr{L}_I)}{\hat{\eta}} \\ &= \frac{\frac{1}{T} \cdot \sum_{t=1}^{T} \Big(\ell_{t,B}(\hat{\theta}_B) - \ell_{t,I}(\hat{\theta}_I) \Big)}{\sqrt{\dfrac{\frac{1}{T} \cdot \sum_{t=1}^{T} \Big(\ell_{t,B}(\hat{\theta}_B) - \ell_{t,I}(\hat{\theta}_I) \Big)^2}{T}}} \xrightarrow{d} N(0,1) \end{aligned}$$

where the standard normal distribution holds under:

$$H_0 : \mathbb{E}[\ell_{t,B}(\theta_B^*)] = \mathbb{E}[\ell_{t,I}(\theta_I^*)]$$

where θ_B^* and θ_I^* are the pseudo-true values of the parameters. Basically, the null hypothesis is saying that the two potentially misspecified models fit the data equally well. Notice that the test is super easy to implement: 1) define the difference: $\hat{d}_t = \ell_{t,B}(\hat{\theta}_B) - \ell_{t,I}(\hat{\theta}_I)$; 2. Regress $\hat{d}_t$ on unity; 3. Run a t-test to verify that the average of the difference is statistically different from zero. We reject the null hypothesis in favor of a better fit to the data of the baseline model if $\hat{d}_t$ is statistically greater than zero. Notice that if this happens it

does not mean that the baseline model is correctly specified (although it could be), however, we can conclude that the baseline model fits better in terms of expected likelihood.
The value we obtain is VMS = 0.033 which is clearly not statistically different from zero. Even if positive sign of the statistics points at a better fit of the baseline model against the inverted one, there is not enough statistical evidence to claim that the baseline outperforms on average the inverted model.

D.2 Output Effect of Fiscal Plans in the Inverted Model

We report here the estimated posterior distributions of the ATE, ADE and ANE for fiscal adjustment plans obtained from the estimates of Equation (7) (inverted model).

Table VIII: Average Total, Direct and Network Effects of Fiscal Consolidations in the United States

	Inverted Model - Equation (7)										
	$\mathbb{E}(\theta)$	*%*	$\sqrt{\mathbb{V}(\theta)}$	$Pr(\theta < 0)$	*5%*	*10%*	*16%*	*50%*	*84%*	*90%*	*95%*
ATE_{TB}	-1.148	1	1.034	0.872	-2.909	-2.481	-2.162	-1.107	-0.131	0.140	0.480
ADE_{TB}	-0.848	0.74	0.756	0.872	-2.106	-1.819	-1.593	-0.835	-0.101	0.107	0.375
ANE_{TB}	-0.300	0.26	0.290	0.872	-0.828	-0.682	-0.572	-0.263	-0.029	0.030	0.102
ATE_{EB}	0.522	1	0.337	0.064	-0.048	0.096	0.203	0.536	0.847	0.936	1.046
ADE_{EB}	0.491	0.94	0.318	0.064	-0.044	0.089	0.188	0.501	0.799	0.886	0.990
ANE_{EB}	0.031	0.06	0.032	0.064	-0.002	0.002	0.005	0.024	0.059	0.073	0.091

Table VIII: *descriptive statistics of posterior distributions of Average Effects of a 2 years, 1% magnitude fiscal adjustment plan. 2 years means that results are calculated by cumulating the effect of the first year of the plan and then the second one. The style of the plan is simulated from a distribution which mimics the observed one; see Appendix C.3 for technical details. Columns:* $\mathbb{E}(\theta)$ *is the expected value of the posterior distribution; % is the share of ATE represented by ADE and ANE.* $\sqrt{\mathbb{V}(\theta)}$ *is the standard deviations of the posterior distribution;* $Pr(\theta < 0)$ *is the probability of negative values, calculated by integrating the posterior distribution; "p%" is the p-th percentile of the posterior distribution.*

The most important thing to notice is that the ANE of EB plans accounts for only 6% of their ATE, against the 12% of the baseline model. The relevance of ANE of TB plans is basically unaffected, diminishing only by 1% relative to the baseline (from 27% of the ATE to 26%). The statistical significance of the ANE of TB plans declines, since the posterior distribution shrinks towards zero.

E A Potential Theoretical Framework

We show here the theoretical framework which we have in mind when we refer to the theoretical transmission of demand and supply shocks. The model is a slight modification of Acemoglu, Akcigit, and Kerr (2016), which we adapted to allow for the propagation of a production tax.

The model considers a perfectly competitive economy with n sectors, where the market clearing condition for the generic industry i is:

$$y_i = c_i + \sum_{j=1}^{n} x_{ji} + G_i \tag{8}$$

where c_i is household's consumption of good produced by industry i; x_{ij}[39] is the quantity of goods produced in industry j used as inputs by industry i; G_i are government purchases.

$$\sum_{i=1}^{n} p_i G_i = T + \tau \sum_{i=1}^{n} p_i y_i \tag{9}$$

Each sector solves the following profit maximization problem:

$$\max_{l_i, \{x_{ij}\}_{j=1}^{n}} (1-\tau) \cdot p_i \cdot \left(\underbrace{l_i^{\alpha_i^l} \cdot \big(\prod_{j=1}^{n} x_{ij}^{\alpha_{ij}}\big)^{\rho}}_{y_i} \right) - w l_i - \sum_{j=1}^{n} p_j x_{ij}$$

where τ is a sales/production tax which mimics an excise tax.[40] Notice that the production function is similar to the one in Acemoglu, Carvalho, et al. (2012) and Carvalho (2014). All alpha's are non negative, and we assume constant return to scale: $\alpha_i^l + \rho \cdot \sum_{j=1}^{n} a_{ij} = 1$. Notice here, that thanks to the Cobb-Douglas specification, ρ can be interpreted as the share of intermediates in production.

The economy is populated by a representative agent, who maximizes utility subject to a budget constraint:

$$\max_{l, \{c_i\}_{i=1}^{n}} (1-l)^{\lambda} \cdot \prod_{i=1}^{n} c_i^{\beta_i} \quad \text{s.t.} \sum_{i=1}^{n} p_i c_i \leq wl$$

[39] In Equation (8) we actually have x_{ji}, that is, the amount of good i used as input by industry j; we then sum over the j-s to obtain the total demand of good i from all the industries.

[40] For example, an excise is a special type of sales tax, which is sector-specific. Excise tax might be of two types: ad valorem (percentage of values of a good) and specific (tax paid per unit). The excise tax may be paid by the producer, retailer, and consumer. Moreover, it might be taken on federal, state, and local levels.

with $\sum_{i=1}^{n} \beta_i = 1$.
Firms and households take all prices as given, and the market-clearing conditions are satisfied in the goods market and the labor market. Government actions are taken as given and the wage is chosen as a numeraire $(w = 1)$.

We do not explicitly model a government budget constraints, since during years of fiscal consolidations, spending cuts are not compensated by tax reductions and viceversa. For simplicity we also do not model government debt and deficit.

Households. The household problems returns the following equilibrium conditions:

$$\frac{p_i \cdot c_i}{\beta_i} = \frac{p_j \cdot c_j}{\beta_j} \quad \forall i, j$$
$$l = \frac{1}{1+\lambda}$$
$$c_i = \frac{\beta_i}{p_i} \cdot \frac{1}{1+\lambda} \quad \forall i$$
$$\sum_{i=1}^{n} p_i \cdot c_i = \frac{1}{1+\lambda}$$

Therefore, in equilibrium we have:

$$d \log c_i = -d \log p_i \quad \forall i$$

that is, percent changes in consumption of good i only depend on percent changes in the price of the same good (with Cobb-Douglas utility income and substitution effects cancel out).

Firms Firms maximize profits and in equilibrium the following FOCs hold true:

$$(1-\tau) \cdot p_i \cdot \rho \cdot a_{ij} \cdot \frac{y_i}{x_{ij}} = p_j$$
$$(1-\tau) \cdot p_i \cdot \rho \cdot a_i^l \cdot \frac{y_i}{l_i} = 1$$
$$y_i = l_i^{\alpha_i^l} \cdot \left(\prod_{j=1}^{n} x_{ij}^{\alpha_{ij}}\right)^{\rho}$$

Acemoglu, Akcigit, and Kerr (2016) notes that solving the dual problem (cost minimization) and obtaining the unit cost function is beneficial to the analysis.

The unit cost function is equal to:

$$C(p_1, ..., p_n) = \underbrace{\left(\frac{1}{a_i^l}\right)^{a_i^l} \cdot \left(\prod_{j=1}^{n} \left(\frac{1}{\rho \cdot a_{ij}}\right)^{a_{ij}}\right)^{\rho}}_{:=B_i} \left(\prod_{j=1}^{n} p_j^{a_{ij}}\right)^{\rho}.$$

Because of perfect competition, price equals marginal cost. Therefore:

$$(1-\tau) \cdot p_i = C(p_1, ..., p_n)$$

By log differentiating the above expression, we have:

$$d \log p_i = \rho \cdot \sum_{j=1}^{n} a_{ij} \cdot d \log p_j + \frac{\tau}{1-\tau} d \log \tau$$

The above expression implies that prices are affected only by changes in the production tax τ. Moreover, from profit maximiation we also have:

$$\rho \cdot a_{ij} = \frac{1}{1-\tau} \cdot \frac{p_j \cdot x_{ij}}{p_i \cdot y_i} \propto \frac{\text{SALES}_{j \to i}}{\text{SALES}_i}.$$

In other word, if sector i is affected by a tax shock, the effect is propagated downstream to the customers, via x_{ij}. This should be clear if we substitute the firm's FOC condition into the previous expression:

$$d \log p_i = \frac{1}{1-\tau} \cdot \sum_{j=1}^{n} \frac{p_j \cdot x_{ij}}{p_i \cdot y_i} \cdot d \log p_j + \frac{\tau}{1-\tau} d \log \tau.$$

E.1 Network effect of a tax shock

We want to know what is the output effect of a change in the production tax. In order to do so, we need to look at the resource constraint (assuming for simplicity that $G_i = 0$ for all sectors):

$$\begin{aligned}
y_i &= c_i + \sum_{j=1}^{n} x_{ji} \\
\frac{y_i}{c_i} &= 1 + \sum_{j=1}^{n} \frac{x_{ji}}{c_i} \quad \text{plug in: } x_{ji} = (1-\tau)\cdot p_j \cdot \rho \cdot a_{ji} \cdot \frac{y_j}{p_i} \text{ (Firm FOC)} \\
\frac{y_i}{c_i} &= 1 + (1-\tau)\cdot \rho \sum_{j=1}^{n} a_{ji} \cdot \frac{p_j \cdot y_j}{p_i \cdot c_i} \quad \text{plug in: } c_i = \frac{\beta_i}{\beta_j} \cdot \frac{p_j \cdot c_j}{p_i} \text{ (HH FOC)} \\
\frac{y_i}{c_i} &= 1 + (1-\tau)\cdot \rho \sum_{j=1}^{n} a_{ji} \cdot \frac{\beta_i}{\beta_j} \cdot \frac{y_j}{c_j} \quad \text{Denote by: } \theta_i := y_i / c_i \\
\theta_i &= 1 + (1-\tau)\cdot \rho \sum_{j=1}^{n} \underbrace{a_{ji} \cdot \frac{\beta_i}{\beta_j}}_{m_{ij}} \cdot \theta_j
\end{aligned}$$

Denote by $M := [m_{ij}]_{i,j=1,\ldots,n}$. Then, in matrix notation the above expression becomes:

$$\boldsymbol{\theta} = \mathbf{1}_n + (1-\tau)\cdot \rho \cdot M \cdot \boldsymbol{\theta} \implies \boldsymbol{\theta} = (I_n - (1-\tau)\cdot \rho \cdot M)^{-1} \cdot \mathbf{1}_n$$

Notice that the equilibrium level of the output-to-consumption ratio, θ_i, has a nice analytical form which, however, depends on τ. Therefore, when τ changes, also this ratio changes and we don't have $d\log y_i = d\log c_i$ as in Acemoglu, Akcigit, and Kerr (2016).

Differentiating the above expression yields:

$$\begin{aligned}
d\boldsymbol{\theta} &= \frac{\partial \left(I_n - (1-\tau)\cdot \rho \cdot M\right)^{-1}}{\partial \tau} \cdot \mathbf{1}_n d\tau \\
&= -\rho \cdot (I_n - (1-\tau)\cdot \rho \cdot M)^{-1} \cdot M \cdot (I_n - (1-\tau)\cdot \rho \cdot M)^{-1} \cdot \mathbf{1}_n d\tau
\end{aligned}$$

Using the $d\log$ notation:

$$d\log \boldsymbol{\theta} = -\underbrace{\tau \cdot \rho \cdot \Theta^{-1} \cdot (I_n - (1-\tau)\cdot \rho \cdot M)^{-1} \cdot M \cdot (I_n - (1-\tau)\cdot \rho \cdot M)^{-1}}_{:=F} \cdot \mathbf{1}_n d\log \tau$$

where $\Theta = diag(\theta_1, ..., \theta_n)$. Recalling the definition of θ_i, we have:

$$d\log \boldsymbol{y} = d\log \boldsymbol{c} - F \cdot \mathbf{1}_n \cdot d\log \tau \implies d\log y_i = d\log c_i - \phi_i \cdot d\log \tau$$

where ϕ_i is the i-th element of vector $F \cdot \mathbf{1}_n$. Notice that if τ were fixed (i.e. $d\log\tau = 0$), percent changes in consumption would be equal to the one of output, as in Acemoglu, Akcigit, and Kerr (2016).

At this point we can find the relationship between output changes and tax shocks. Consider the following three equations we derived earlier:

$$\begin{cases} d\log y_i &= d\log c_i - \phi_i \cdot d\log\tau \\ d\log c_i &= -d\log p_i \\ d\log p_i &= \rho \cdot \sum_{j=1}^{n} a_{ij} \cdot d\log p_j + \dfrac{\tau}{1-\tau} d\log\tau \end{cases}$$

Combining the three equations above yields the following expression:

$$\begin{aligned} d\log y_i &= \rho \cdot \sum_{j=1}^{n} a_{ij} \cdot d\log y_j - \underbrace{\left(\phi_i + \frac{\tau}{1-\tau} - \rho \cdot \sum_{j=1}^{n} \phi_j \cdot a_{ij}\right)}_{=\psi_i > 0} \cdot d\log\tau \\ &= \rho \cdot \sum_{j=1}^{n} a_{ij} \cdot d\log y_j - \psi_i \cdot d\log\tau \end{aligned}$$

which is Equation (4) in the paper.

E.2 Network effect of a spending shock

Suppose now that $\tau = 0$ and that the government reduces its purchases from all sectors (i.e. $d\log G_i < 0$). We want to find the relationship between the percent change in output, $d\log y_i$ and percent changes in government purchases $d\log G_i$.

Consider the resource constraint of the economy:

$$y_i = c_i + G_i + \sum_{j=1}^{n} x_{ji} \quad \text{Log-differentiate}$$

$$d\log y_i = \frac{c_i}{y_i} \underbrace{d\log c_i}_{=0 \ (d\log p_i = 0)} + \frac{G_i}{y_i} d\log G_i + \sum_{j=1}^{n} \frac{x_{ji}}{y_i} \cdot d\log x_{ji} \quad \text{(Firm FOC)} \ x_{ji} = p_j \rho a_{ji} \frac{y_j}{p_i}$$

$$d\log y_i = \frac{G_i}{y_i} d\log G_i + \rho \cdot \sum_{j=1}^{n} \underbrace{a_{ji} \cdot \frac{p_j \ y_j}{p_i \ y_i}}_{:=\hat{a}_{ji}} d\log x_{ji}$$

$$d\log y_i = \frac{G_i}{y_i} d\log G_i + \rho \cdot \sum_{j=1}^{n} \hat{a}_{ji} \cdot d\log x_{ji}$$

From the firm's FOC, we have:

$$d \log y_i = \underbrace{d \log p_j}_{0} + d \log x_{ji} - \underbrace{d \log p_i}_{=0}$$

therefore we can retrieve Equation (2):

$$d \log y_i = \rho \cdot \sum_{j=1}^{n} \hat{a}_{ji} \cdot d \log y_j + \frac{G_i}{y_i} d \log G_i.$$