Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Bsc2 Econometrics/Economics

Estimating R&D spillovers using Lasso and Elastic-Net methods

Marisha van der Arend (535761)

Supervisor:	S. Koobs
Second assessor:	D. van Dijk
Date final version:	2nd July 2023

Abstract

Assessing spillover effects can hold significant value. Accurately estimating spillovers can be valuable in policy design, as it allows for the recognition of individuals who have a positive impact on others. However, estimating spillover effects might turn out to be challenging in a panel data structure that involves numerous individuals and a relatively short time horizon. This paper aims to estimate spillover effects by utilizing a Double Pooled Lasso approach, which is based on the method proposed by Manresa (2016). The approach is further adapted to apply the Double Pooled estimator to Elastic-Net. This paper considers an adaptive Lasso and Elastic-Net approach with five different expressions for the weights and two different expressions for the penalty parameter. This paper performs a simulation in order to find the best method for estimating spillovers. Finally, the best models are applied to find the spillover effects of R&D investment on a company's sales. The data is collected from the United States and ranges from 1980 to 2001 and it includes 21 years of data and 263 companies. This paper finds that there is no single method which consistently outperforms all other methods. The best performing methods are the method which uses the penalty parameter and the weights of Manresa (2016), and the two methods with weights based on Lasso and Elastic-Net without weights.

1 Introduction

When the economy in one country is all of a sudden performing badly, in most cases this also affects other countries. This is an example of a negative spillover effect, where one country influences many other countries. A spillover effect can be positive or negative and is defined as a phenomenon where one event, which happens at a certain place, company, etcetera, has an effect somewhere else.

Finding a good model to determine spillover effects in business settings is of great importance. If it becomes possible to determine the effect of spillovers, policies can be designed to maximize the social benefits of these spillovers. This paper will focus on spillover effect of the R&D stock of a company on the sales of other companies. Earlier research confirmed the presence of spillovers in R&D investments. At international level, Coe and Helpman (1995) and Bayoumi, Coe and Helpman (1999) find that the R&D capital stock in one country can influence their own factor productivity together with the factor productivity of other countries. At company level, Bloom, Schankerman and Van Reenen (2013) found a positive spillover effect from R&D investments on the firm performance, which outperforms the negative business stealing effects from product rival firms. Furthermore, O'Mahony and Vecchi (2009) found a higher productivity in groups which are more skill intensive and have a higher R&D capital stock. Additionally, Jefferson, Huamao, Xiaojing and Xiaoyun (2006) find that, using data from Chinese companies, R&D expenditure leads to higher sales, and that the returns to investment from R&D are higher than those of fixed investment. Thus, there seems to be evidence that R&D investments do not only affect the performance of its own company, but also the performance of other companies.

Previous research also looked at where spillovers were mainly present. Audretsch and Feldman (1996) found that industries that rely heavily on knowledge spillovers tend to have a higher concentration of innovative activity. Furthermore, Audretsch and Vivarelli (1996) found that firm R&D expenditure mainly affects the innovation of all companies, while university R&D expenditure affects mostly the innovations of small firms. Bernstein and Nadiri (1988) estimated the spillover network between five high-tech industries and found that the private rate of return of R&D were actually smaller than the social rate of return. Most industries generated spillovers on two industries, except for the industry nonelectrical machinery. The scientific instruments industry generated the largest amount of spillovers.

In order to model the spillover effects, this paper uses a Cobb-Douglas production function. The Cobb-Douglas function models the relationship between the amount of certain inputs on the amount of output that is produced (Douglas, 1976). The most well-known version of the Cobb-Douglas production function has two inputs, capital and labor, to generate output (Zellner, Kmenta & Dreze, 1966). This function also includes the technological progress, but this influences all companies or countries in the same way. Other research papers have modified the conventional Cobb-Douglas function to better suit their research objectives. For example, Yuan, Liu and Wu (2009) added the input factor energy to the Cobb-Douglas function with capital, labor, and technological process to model the economic growth in China. Additionally, Vasylieva, Lieonov, Liulov and Kyrychenko (2018) use a Cobb Douglas function with as output GDP per capita, and as inputs technological progress, capital, labor, macroeconomic stability, openness of the economy and foreign direct investment. This paper uses the Cobb-Douglas function established in the paper of Manresa (2016), where the inputs are capital, labor, technological progress, knowledge capital, and knowledge spillovers.

In order to estimate the parameters in the Cobb-Douglas function, data is necessary. The data that is collected comes from Compustat and is company-level panel data from the United States. In the Cobb-Douglas function, the sales of a company is used as a measure for output, the capital stock of a company is used for the input capital and the number of employees of a company is used for the input labor. The input knowledge capital uses the company's R&D stock as a measure, and the input knowledge spillovers uses the R&D stock of all other companies in the dataset. This paper uses panel data, which implies that data is collected from different companies over multiple time periods. In total this paper uses 22 years of data and 263 companies, which implies that for each company in the dataset 22 years of data is available for all variables. However, in total only 21 years can be used for estimation, as the R&D stock is lagged in the Cobb-Douglas function.

Even though the presence of spillovers seems to be a widely accepted fact and a model for estimating spillovers is established, estimating the magnitude of the spillover effects seems to be empirically challenging. The main issue is that within the dataset the number of time periods, $T = 21$, is smaller than the number of companies, $N = 263$. Therefore, there is the issue of high-dimensionality, as the final spillovers that should be estimated is an $N \times N$ matrix, while only $N \times T$ observations are available for this estimation (Manresa, 2016). Therefore, a simple ordinary-least-squares approach to estimate the spillovers can no longer be used, because if there are more variables to be estimated compared to the number of observations there is no unique solution and thus the ordinary-least-squares is undefined. Thus, different approaches should be considered. Another issue could be that spillovers might take place through characteristics which are not included in the model or that the model chosen might not account for dynamic, time-invariant or firm-specific effects (Arcidiacono, Foster, Goodpaster & Kinsler, 2012; Görg & Strobl, 2001; Dimelis, 2005). This could lead to biased and inaccurate estimates for spillovers.

In order to estimate spillover effects in a dataset with high-dimensionality, a Lasso and an Elastic-Net approach is used. Lasso is an abbreviation for least absolute shrinkage and selection operator and it minimizes the squared residual sum of the model and is subjected to a constraint that the sum of the absolute values of the coefficients should be smaller than or equal to a certain constant (Tibshirani, 1996). Lasso essentially shrinks the coefficients, where it selects the most important variables and sets the coefficients of the other variables equal to zero. As the model with spillovers that is attempted to be estimated in this paper will likely have sparsity, that is the total companies that provide spillover effects small compared to all the companies available that could possibly provide spillovers, the Lasso estimator seems to be appropriate to reduce dimensionality and select only the variables which cause spillovers. However, Lasso also has some issues which can lead to a bad performance of this estimator. The first issue is that when the number of observations, T, is larger than the number of variables, N, Lasso will select at most T variables (Zou & Hastie, 2005). Furthermore, if a certain group of variables have a high pairwise correlation, then lasso mostly selects only one variable from the group and does not care which variable is selected. Lastly, if there are high correlations between predictors and $T > N$, the lasso estimator is dominated by a ridge estimator, which sets no coefficients equal to zero.

In order to overcome the issues of Lasso mentioned in the last paragraph, Elastic-Net was designed, which should eliminate the first two issues and should deliver a better prediction than Lasso when the variables are highly correlated and $T > N$. Elastic-Net is essentially a linear combination between a Lasso and a Ridge regression (Zou & Hastie, 2005). A Ridge regression minimizes the squared residual sum of the model and is subjected to a constraint that the sum of the squared values of the coefficients should be smaller than or equal to a certain constant (Hoerl & Kennard, 1970). Thus, Ridge shrinks the coefficients towards zero, but does not actually set any variables equal to zero, which is also the reason why using only Ridge is not appropriate in this case as scarcity will not be achieved with only Ridge. An Elastic-Net regression minimizes the squared residual sum of the model and is subjected to the constraint that the linear combination of the sum of squared values of the coefficients and the sum of absolute values of the coefficients should be smaller than a certain constant (Zou & Hastie, 2005). As it is a linear combination between Ridge and Lasso, it does set values equal to zero depending on the value of α, were a larger value of α resembles a Lasso regression more and thus sets more values equal to zero. Therefore, this model is also appropriate to reduce dimensionality of the model and only select the variables which are the most important. One would generally expect Elastic-Net to outperform Lasso when there are correlations and the number of spillovers are larger than the number of observations, T.

In order to estimate the spillovers in this paper the Pooled Lasso and the Double Pooled Lasso approach is used from the paper of Manresa (2016). This approach is made more general so it is also usable for the Elastic-Net estimator. The Pooled estimation technique can estimate the spillover effect in case no control variables are present and first uses the Lasso (Elastic-Net) estimator for the selection of the variables and then uses regular ordinary-least-squares for estimating the magnitude of the selected variables in order to eliminate shrinkage bias. The Double Pooled estimation technique is used in case there are control variables, thus in the final estimation of the model, and essentially applies the Pooled estimation multiple times. It should be noted though that not the regular Lasso and Elastic-Net models are used, but rather the adaptive Lasso and a variation of the adaptive Elastic-Net model, which include weights in the penalty term. The reason for using weights in the penalty term is to try to make the models consistent for model selection, as without weights this is only true when the estimated model satisfies a strong condition Zou and Zhang (2009).

This paper considers different expressions for the weights and the penalty parameters in the Lasso and Elastic-Net methods. There are two different expressions for the penalty parameter, one is based on the paper of Manresa (2016) and the other one is computed with cross-validation. There are five different expressions for the weights. Firstly, there is an equally-weighted model, which is essentially a Lasso or Elastic-Net estimator without weights. Additionally, weights are constructed using the sample variance of the independent variable x and different weights are constructed using the procedure described in Manresa (2016). Furthermore, weights are calculated using the obtained coefficients from a Lasso or Elastic-Net regression without weights, depending on whether the Lasso or Elastic-Net model is used to estimate the parameters. Finally, weights are calculated using the obtained coefficients from a Ridge regression without weights. In total there are ten Lasso models and ten Elastic-Net models, which will be compared against each other to determine which model has the best performance. The performance measures for comparison are the Frobenius norm, the mean squared error and the percentage accurate predicted non-zero and zero values. For the Double Pooled estimator also the performance measure of the accuracy of the control parameters is included.

This main goal of this paper is to find the best model to estimate the spillover effects of R&D investments between firms. In order to find the answer to this question it is important to research which expression for the weights is the most suitable. Furthermore, it should be determined what the optimal penalty term should be. It is also important to see whether the additional complexity of the Elastic-Net approach actually gives significantly better estimation results compared to the simpler Lasso approach. Furthermore, this paper uses the best performing models to interpret the spillover effects.

This paper finds that the method that is used in Manresa (2016) actually underperforms against some simpler methods that use the estimated coefficients of an Lasso (Elastic-Net) model without weights in order to set the weights. Only for larger sample sizes the method of Manresa (2016) performs well, but for smaller samples it performs quite bad. Generally, Lasso outperforms Elastic-Net even when correlations are introduced. Furthermore, using the penalty parameter of Manresa (2016) outperforms a penalty parameter found by cross-validation for a small time period, while the penalty parameter with cross-validation outperforms the penalty parameter of Manresa (2016) for larger time periods. Thus, the best method seems to be a method which uses Lasso weights computed with the coefficients estimated from a Lasso model without weights.

This paper also attempts to calculate spillover effects. Three models find that an increase in a company's own R&D stock leads to lower sales, while the other three models find the opposite effect. It is thus challenging to determine the actual impact of R&D investment on sales. Furthermore, most of the companies that provide spillover effects, give negative values for these effects. However, two companies that do provide positive spillovers are both larger than the average companies. The companies that receive positive spillovers are small companies.

This paper contributes to existing literature by comparing the performance of a Lasso and an Elastic-Net estimator on panel data to detect spillover effects. It applies the Pooled and Double Pooled techniques in the paper of Manresa (2016) and also gives some other variations to this paper in order to try to improve the technique that was developed in her paper. Furthermore, this paper also includes a simulation studies which compares many different models in order to find the best performing model.

The remainder of this paper is organized as follows: Section 2 presents the data that will be used in this paper and describes the data cleaning process. Section 3 introduces the main model, the Pooled estimator with the different expressions for the weights and the penalty parameter, and the Double Pooled estimator. This section also describes how the models will be compared. Section 4 will describe the results found in the simulation of both the Pooled as well as the Double pooled estimator. Section 5 will describe how the model will be applied to the data and gives the results. Finally, Section 6 will conclude.

2 Data

The dataset that is used in this paper is a combination between the NBER match of Compustat with the U.S. Patent Trade Office (USPTO) patent database. The period of interest in this paper is 1980-2001 and the data is obtained from the paper Bloom et al. (2013), more specifically on the site of Nicholas Bloom [1]. The dataset contains panel data on U.S. firms over multiple time periods. Compustat provides data on the real sales, the real market value, the capital stock, the number of employees, the R&D stock and the R&D expenditure. The USPTO provides data on the granted patents and all of the citations of these patents. However, this data is not used in this paper. In total the dataset of Bloom et al. (2013) contains 18,209 observations with 736 companies. However, not all companies have data in the years of interest, or there is missing data. Some observations may not be very useful because the R&D stock of companies remains constant over time, making it challenging to estimate parameters accurately. Therefore, some companies and observations have been removed. The data cleaning process can be found in the Appendix Section A. The summary statistics of the adapted dataset can be found in Table 1. Note that this adapted data set contains 22 years of data.

Table 1: Summary statistics of the adapted data set.

Variable	Obs.	Mean	Std. dev.	Min.	Max.
R&D Stock	5,789	1,126.60	3,910.28	0	47,343.38
Sales	5,789	3,718.23	11,387.07	5.01	140,609.60
Capital stock	5,786	1,482.15	4,580.44	0.91	72,825.98
Employment	5,786	24.07	61.48	0.11	876.80

[1] The dataset was found on the site https://nbloom.people.stanford.edu/research and the data was taken from the paper "Identifying Technology Spillovers and Product Market Rivalry".

3 Methodology

The main model that this paper estimates can be expressed as:

$$y_{it} = a_i + \beta_i x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + w_{it}^\top \theta + u_{it}, \tag{1}$$

where y_{it} is the outcome of company i at time t, x_{it} is the characteristic of company i at time t, w_{it} is a vector containing D control variables that affect all companies through θ, and u_{it} represents the idiosyncratic shocks which are uncorrelated with both $x_{1t}, ..., x_{Nt}$ and w_{it}. Furthermore, a_i is a company-specific intercept and β_i captures the effect of its own characteristic x_{it} on the outcome of a company y_{it}. Additionally, γ_{ij} captures the spillover effect from company j to company i.

Model (1) allows for heterogeneity of spillover effects between the individual companies, and does not specify the reference groups. This model can become problematic when the number of companies N surpasses the number of observations T, as in that case there may be an excess of parameters in comparison to the amount of data points, resulting in an unidentified model. Therefore this paper focuses on sparse structure of interactions, where the number of connections between companies, $\gamma_{ij} \neq 0$, is small but the identity and the magnitude of spillovers remains unrestricted.

The main goal of this paper is to estimate the spillover effect $\widehat{\gamma}_{ij}$ to see the effect of a certain characteristic x of one company on the outcome y of another company. To achieve this a Double Pooled estimation is performed. In order to explain this technique, first the the Pooled estimation technique is explained, as that technique is repeated several time within the Double Pooled estimation. Thus, understanding the Pooled estimation is essential for comprehending the Double Pooled estimation. Lastly, the performance measures used to compare the different models will be discussed.

3.1 The Pooled estimation

To gain a better understanding of the Pooled estimation technique, we examine the simplified model presented below (obtained by setting $\theta = 0$ in equation (1)):

$$y_{it} = a_i + \beta_i x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + u_{it}. \tag{2}$$

As the number of non-zero spillovers, $\gamma_{ij} \neq 0$, should be limited, it many values of γ_{ij} should be set to zero. This brings us to a Lasso estimator, as that method is known to select only a few parameters out of a large set of parameters in case an appropriate penalty term is chosen. More specifically, this paper uses a **Pooled Lasso estimator** (Manresa, 2016):

$$\widehat{\Gamma} = \underset{\Gamma}{\text{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{y}_{it} - \sum_{j=1}^{N} \gamma_{ij} \tilde{x}_{jt} \right)^2 + \frac{\lambda}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi_{ij} |\gamma_{ij}|, \tag{3}$$

where the parameter of interest $\widehat{\Gamma}$ is an $N \times N$ matrix which contains all of the spillover effects, γ_{ij}, where γ_{ii} corresponds to β_i. Furthermore, $\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and similarly $\tilde{x}_{jt} = x_{jt} - \frac{1}{T}\sum_{t=1}^{T} x_{jt}$, thus the data is demeaned which is done to remove the fixed effect a_i from the equation. Finally, ϕ_{ij} are the pair-specific weights and λ is the penalty parameter.

Even though the Pooled Lasso estimator is extremely useful for only selecting the most important variables, it still has some issues. One of these issues is that the Lasso estimator can perform quite poorly when the x-variables are highly correlated (Zou & Zhang, 2009). This could potentially be an issue here, as it is not unthinkable that the R&D capital stock between firms is correlated. Therefore, also the **Pooled Elastic-Net estimator** is used (based on (Bonaldi, Hortaçsu & Kastl, 2015) and (Khan & Shaw, 2016)):

$$\widehat{\Gamma} = \underset{\Gamma}{\operatorname{argmin}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \sum_{j=1}^{N}\gamma_{ij}\tilde{x}_{jt}\right)^2 + \frac{\lambda(1-\alpha)}{2NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ij}\gamma_{ij}^2 + \frac{\lambda\alpha}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ij}|\gamma_{ij}|, \quad (4)$$

with similar parameters as the Pooled Lasso estimator and where α essentially decides how closely the estimator resembles a Lasso or Ridge estimator. For $\alpha = 0$ one can obtain a Ridge estimator and for $\alpha = 1$ one obtains the Pooled Lasso estimator. For $0 \leq \alpha \leq 1$, Elastic-Net is thus a combination between Ridge and Lasso. This paper will focus mostly on an equal mix between the two methods with $\alpha = 0.5$.

The Pooled Lasso and the Pooled Elastic-Net estimator shown in equation (3) and (4) will be estimated for each row i in the matrix $\widehat{\Gamma}$ separately for simplicity. This is possible as minimizing the sum over each company i will give the same results as minimizing each company i separately. Therefore each row in $\widehat{\Gamma}$ in Elastic-Net will be estimated using the formula:

$$\widehat{\gamma}_i = \underset{(\gamma_{i1},\ldots,\gamma_{iN}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \sum_{j=1}^{N}\gamma_{ij}\tilde{x}_{jt}\right)^2 + \frac{\lambda(1-\alpha)}{2T}\sum_{j=1}^{N}\phi_{ij}\gamma_{ij}^2 + \frac{\lambda\alpha}{T}\sum_{j=1}^{N}\phi_{ij}|\gamma_{ij}|. \quad (5)$$

Note that each row $\widehat{\gamma}_i$ corresponding to Lasso can also be calculated using equation (5) by setting $\alpha = 1$.

Note that the estimator in equation (5) consists of two parts. The first part is the sum of the squared errors of the model we estimate, $\sum_{j=1}^{N}\gamma_{ij}\tilde{x}_{jt}$, compared to the actual observations, $\tilde{y}_{it}$. The second part is a penalization, which consists out two parts for Elastic-Net and only one part for Lasso. Generally, the sum of squared errors is decreasing in the number of spillover effects, while the penalization is increasing in the number of spillover effects. The number of spillover effects, $\gamma_{ij} \neq 0$, is then determined by the size of the penalty term λ, as a higher value of λ will results in more zeros in the matrix Γ.

Within equation (3) and (4) both the penalty term λ and the weights ϕ_{ij} must be specified. To that end different specifications for the weights ϕ_{ij} and the penalty term λ will be tested. There are two different specifications for the penalty term and five different specifications for the weights. Together with the Lasso and the Elastic-Net model there are in total 20 different methods that will be considered. Firstly, the different types of penalty parameters will be described, followed by an explanation about the different types of weights used in this paper. Finally, Table 2 will give an overview of the 20 methods with a short description of the types of weights and penalty term in each model.

3.1.1 The choice of penalty parameter

For the penalty term λ two different specifications are tested. For the first specification the the penalty term is chosen to be based on the paper of Belloni, Chen, Chernozhukov and Hansen (2012):

$$\lambda = c2\sqrt{T}\Phi^{-1}(1 - v/(2N)), \tag{6}$$

where Φ is the standardized Gaussian cumulative distribution function, c is a constant above 1, and v is the pre-specified level of error[2]. Note that this penalty term is the same for both Lasso as Elastic-Net.

For the second specification of λ cross-validation (CV) is used to recover λ. The reason for also looking at a penalty term which uses CV is to see whether this penalty term can outperform a penalty term based on theory. Note that since each row in the matrix $\widehat{\Gamma}$ is estimated separately, λ will also differ for each row. In this case thus λ will differ not only per row, but also per method (Elastic-Net versus Lasso). The penalty parameter λ will be chosen using 5-fold CV. The choice for 5-fold CV seems appropriate given that different values of T will be used, including a small value of 21 years and a larger value of 100 years in the simulation. Therefore, using 5-fold CV is suitable for a sample size of 21 years, as there is sufficient data to estimate the model in each of the five parts. Additionally, it is not too computationally heavy for larger sample sizes such as 100 years. Leave-one-out CV might be more appropriate for 21 years of data, but will be too computationally heavy for 100 years of data. Finally, with CV the penalty parameter is chosen such that the mean squared error (MSE) of the estimated model is minimised.

3.1.2 The choice of the weights

For the weights in equation (3) and (4) five different specifications are considered. Firstly, the weights $\phi_{ij} = 1,\ \forall(i,j) \in \{1,\dots,N\}$ are chosen. This is equal to a Lasso (Elastic-Net) regression without weights. It will be very interesting to see how well these weights perform compared to some more complicated weights. The performance of these weights will determine whether the more complicated weights discussed below are actually necessary or whether simplified weights already provide a satisfactory performance. The following two specifications choose the weights ϕ_{ij}^2 based on the estimator of $\mathbb{V}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{x}_{jt})$. Firstly, the weights are chosen according to the paper of Belloni and Chernozhukov (2013):

$$\phi_{ij}^2 = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2. \tag{7}$$

These weights are especially appropriate when the explanatory variable x_{it} is independent from u_{it}, which is independent and identically distributed (iid). However, this might not always be the case and therefore also another estimator of $\mathbb{V}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{x}_{jt})$ is necessary. Therefore this

[2] Within this paper all results are obtained using $c = 1.2$ and $v = 0.05$.

paper also looks at a variation of the HAC type estimator proposed by Newey and West (1986):

$$\phi_{ij}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2\widehat{\tilde{u}}_{it}^2 + \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}\widehat{\tilde{u}}_{it}\widehat{\tilde{u}}_{it-1}\right). \tag{8}$$

This is a variation of the HAC type estimator as it takes the average over the N companies to estimate the weights. It should be noted that this estimator is suitable when the error terms $(\tilde{u}_{i1}, ..., \tilde{u}_{iT})$ are iid. These weights are robust to heteroscedasticity as well as autocorrelation of unknown form. However, since the weights are necessary to calculate the error term $\widehat{\tilde{u}}_{it}$, the iterative strategy in Belloni et al. (2012) is used, which helps to set initial weights which are independent of $\widehat{\tilde{u}}_{it}$ [3].

The other two specifications of the weights are based on the paper of Zou and Zhang (2009). First, the weights are calculated by $\phi_{ij} = (|\widehat{\gamma}(lasso)_{ij}|)^{-\delta}$ for the Pooled Lasso estimator and $\phi_{ij} = (|\widehat{\gamma}(enet)_{ij}|)^{-\delta}$ for the Pooled Elastic-Net estimator, where δ is a positive constant. In this case $\widehat{\gamma}(lasso)_{ij}$ is calculated using equation (3) where the weights $\phi_{ij} = 1,\ \forall(i,j) \in \{1,\ldots,N\}$. Furthermore, $\widehat{\gamma}(enet)_{ij}$ is calculated using equation (4) where the weights $\phi_{ij} = 1,\ \forall(i,j) \in \{1,\ldots,N\}$. It should be noted that in case $\widehat{\gamma}(lasso)_{ij}$ or $\widehat{\gamma}(enet)_{ij}$ equals zero the weights ϕ_{ij} will be set close to infinity, as the formulas specified before would otherwise divide by zero. In this case, the Lasso or Elastic-Net estimator without weights, pre-selects the parameters that should be included in the model. Defining the weights according to these expressions might thus lead to including fewer parameters in the final model, as variables with a small magnitude of γ_{ij} will obtain a high weight ϕ_{ij} and might therefore not be selected in the final model. Therefore, these weights are appropriate when a model without weights can select the non-zero parameters quite accurately, but gives too many parameters a non-zero value. These weights can then be utilized to set inaccurately estimated non-zero parameters to zero.

Even though the formula for the weights as specified before can be useful to determine weights it should be noted that the weights largely depend on a Lasso (Elastic-Net) regression without weights. It is possible that these techniques select too few predictors compared to the actual model and subsequently assign extremely high weights to the predictors that were not chosen. This could result in the exclusion of these variables from Lasso (Elastic-Net) regression with weights as well. This could then lead to an underestimation of the effect and the number of spillovers. Therefore, this paper also considers a specification for the weights that are calculated using the coefficients which are obtained by a Ridge regression without weights (based on the paper of Chan and Chen (2011)). Chan and Chen (2011) describe in their paper that if the sample size is small or multicollinearity is an issue, these weights are more appropriate for giving estimates for the weights compared to the ordinary least squares approach. As specifications are considered in this paper where $T < N$, using a Ridge regression to determine the weights may be a more appropriate than using ordinary least squares. The weights are defined as $\phi_{ij} = (|\widehat{\gamma}(ridge)_{ij}|)^{-\delta}$, where $\widehat{\gamma}(ridge)_{ij}$ can be calculated using the following equation:

$$\widehat{\Gamma}(ridge) = \underset{\Gamma}{\operatorname{argmin}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \sum_{j=1}^{N}\gamma_{ij}\tilde{x}_{jt}\right)^2 + \frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}^2.$$

[3]More information about this iterative strategy can be found in the Appendix section B.

Using the weights from the coefficients of a Ridge regression has one big advantage. As Ridge shrinks all coefficients towards zero, but does not actually set any of the coefficients equal to zero that implies that none of the weights will be set to infinity. Thus all weights are defined and none of the the coefficients will be excluded a priori when applying these weights.

Combining Section 3.1.1 and section 3.1.2, there are in total 10 Lasso methods and 10 Elastic-Net methods. An overview of these methods together with an explanation can be found in Table 2.

Table 2: The different methods considered in the Pooled estimator

Method	**Explanation**
	Panel A: Lasso methods
EW static	This method uses weights $\phi_{ij} = 1, \forall (i,j) \in 1, ..., N$ and equation (6) for λ
Var(x) static	This method uses equation (7) for the weights and equation (6) for λ
HAC static	This method uses equation (8) for the weights and equation (6) for λ
Lasso static	This method uses weights $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-\delta}$ and equation (6) for λ
Ridge static	This method uses weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-\delta}$ and equation (6) for λ
EW CV	This method uses weights $\phi_{ij} = 1, \forall (i,j) \in 1, ..., N$ and cross-validation for λ
Var(x) CV	This method uses equation (7) for the weights and cross-validation for λ
HAC CV	This method uses equation (8) for the weights and cross-validation for λ
Lasso CV	This method uses weights $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-\delta}$ and cross-validation for λ
Ridge CV	This method uses weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-\delta}$ and cross-validation for λ
	Panel B: Elastic-Net methods
EW static	This method uses weights $\phi_{ij} = 1, \forall (i,j) \in 1, ..., N$ and equation (6) for λ
Var(x) static	This method uses equation (7) for the weights and equation (6) for λ
HAC static	This method uses equation (8) for the weights and equation (6) for λ
Enet static	This method uses weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-\delta}$ and equation (6) for λ
Ridge static	This method uses weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-\delta}$ and equation (6) for λ
EW CV	This method uses weights $\phi_{ij} = 1, \forall (i,j) \in 1, ..., N$ and cross-validation for λ
Var(x) CV	This method uses equation (7) for the weights and cross-validation for λ
HAC CV	This method uses equation (8) for the weights and cross-validation for λ
Enet CV	This method uses weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-\delta}$ and cross-validation for λ
Ridge CV	This method uses weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-\delta}$ and cross-validation for λ

3.1.3 The Post Pooled estimator

Once the Pooled Lasso (Elastic-Net) estimator has selected its final estimators, $\gamma_{ij} \neq 0$ for equation (2), the final spillover effects are estimated using a pooled panel regression on the selected spillover effects, together with other spillovers that a researcher might want to include in the model. The final estimator of $\widehat{\Gamma}$ is then called the **Post Pooled estimator** (Manresa, 2016):

$$\widehat{\Gamma}^P = \underset{(\gamma_{i1},\ldots,\gamma_{iN}):\gamma_{ij}=0 \text{ if } j \notin \widehat{T}_i}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{y}_{it} - \sum_{j=1}^{N} \gamma_{ij} \tilde{x}_{jt} \right)^2, \tag{9}$$

where $\widehat{T}_i$ are all the regressors which are selected in the Pooled Lasso (Elastic-Net) estimator together with the regressors that a researcher might want to include as well. This paper always includes the variable x_{it}, and thus the parameter γ_{ii}, in the Post Pooled estimator, as it is logical that a company's own characteristic influences the outcome of that company. The main reason for using this Post Pooled estimator is to eliminate shrinkage bias. The Lasso (Elastic-Net) estimator is likely to underestimate the effect of the coefficients, because of the penalty parameter trying to keep the coefficients low. The Lasso (Elastic-Net) estimator has typically a low variance, because fewer parameters are included and the parameters have a smaller magnitude. However, the use of penalty in parameter estimation may result in a higher bias, as it can lead to underestimation of the true value for some parameters. The Post Pooled estimator then introduces a higher variance, as more parameters might be included and the parameters typically have a larger magnitude. Nevertheless, the bias will be reduced, as the parameters are now estimated in such a way that the MSE, and thus the bias, is as small as possible. Therefore, the Post Pooled estimator is useful to find the true parameters without shrinkage bias, even if it comes at the cost of a higher variance.

3.2 The Double Pooled estimator

The Double Pooled estimator attempts to estimate model (1), where $\theta \neq 0$. Therefore, the Double Pooled method utilizes the Pooled method multiple times. However, not all 20 methods described in Table 2 are used in this case. Only the three best performing Elastic-Net and the three best performing Lasso models will be used for the Double Pooled estimator. These choices are based on a simulation study where the y-variable is generated without control variables. Thus, the 20 methods in Table 2 will be reduced to six for the Double Pooled estimator.

In order to provide estimates for the model in equation (1) one should first obtain a consistent estimate for θ. In order to estimate θ the double selection procedure from Belloni, Chernozhukov and Hansen (2014) will be used. This procedure will construct orthogonal projections of y_{it} and w_{it} on $x_{1t}, ..., x_{Nt}$ separately in order to minimize omitted variable bias because of selection mistakes. The final estimator of model (1) is called the **Double Pooled estimator** and is calculated using the following steps.

First, an orthogonal projection of w on $x_{1t}, ..., x_{NT}$ is constructed. For each $d \in \{1, \ldots, D\}$

$$w_{it}^{d} = \eta_{i}^{d} + \sum_{j=1}^{N} \lambda_{ij}^{d} x_{jt} + e_{it}^{d}, \tag{10}$$

where the number of $\lambda_{ij}^{d} \neq 0$ is limited and the mean of the error term e_{it}^{d} conditional on the x-variables is zero. Note that D represents the number of control variables and this regression will thus be performed for each control variable d. Then the values of λ_{ij}^{d} are calculated using the methods described in Section 3.1, where the variable y_{it} is replaced by w_{it}^{d} and we want to estimate $\widehat{\Lambda}^{d}$, which is a matrix containing all the values λ_{ij}^{d}, instead of $\widehat{\Gamma}$. The HAC weights are estimated using ${\phi_{ij}^{d}}^{2} = \widehat{\mathbb{V}}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{x}_{jt} \tilde{e}_{it}^{d}\right)$[4].

Secondly, an orthogonal projection of y on $x_{1t}, ..., x_{NT}$ is constructed:

$$y_{it} = \mu_{i} + \sum_{j=1}^{N} \nu_{ij} x_{jt} + \upsilon_{it}, \tag{11}$$

[4]The precise expressions for the weights and the initial weights of the HAC method can be found in the Appendix Section C.

where $\mu_i = \alpha_i + \eta_i^\top \theta$, $\nu_{ij} = \gamma_{ij} + \lambda_{ij}^\top \theta$, and $\upsilon_{it} = e_{it}^\top \theta + u_{it}$. Additionally, once again the number of variables $\nu_{ij} \neq 0$ is limited. Similar to before, the values of ν_{ij} are calculated using the methods described in section 3.1, where in this case we want to estimate $\widehat{V}$, which contains all the values ν_{ij}, instead of $\widehat{\Gamma}$. The HAC weights are estimated using $\phi_{ij}^2 = \widehat{\mathbb{V}}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{x}_{jt} \tilde{\upsilon}_{it}\right)$[5].

The next step is to use the obtained estimates of $\widehat{\lambda}_i^d$ and $\widehat{\nu}_i$ to estimate $\widehat{\theta}$. Therefore, $\widehat{\theta}$ will be estimated by the following pooled panel regression:

$$\tilde{y}_{it} - \widehat{\nu}_i \tilde{x}_t = \theta(\tilde{w}_{it} - \widehat{\lambda}_i \tilde{x}_t) + \epsilon_{it}, \tag{12}$$

where θ is a vector containing $(\theta^1, \ldots, \theta^D)$, $\tilde{w}_{it}$ is a vector containing $(\tilde{w}_{it}^1, \ldots, \tilde{w}_{it}^D)$, $\widehat{\lambda}_i$ is a matrix containing $(\widehat{\lambda}_i^1, \ldots, \widehat{\lambda}_i^D)$, and ϵ_{it} is the error term of the regression. Thus, $\widehat{\theta}$ will be a $1 \times D$ vector containing D values of θ for all of the D control variables.

Finally, to estimate the actual spillover effects the following formula is used:

$$\tilde{y}_{it} - \widehat{\theta}\tilde{w}_{it} = \sum_{j=1}^{N} \gamma_{ij}\tilde{x}_{ij} + u_{it} + (\widehat{\theta} - \theta^0)\tilde{w}_{it}, \tag{13}$$

where u_{it} is the error term, θ^0 is the actual value of θ and thus $(\widehat{\theta} - \theta^0)\tilde{w}_{it}$ is the estimation error from estimating θ[6]. The values γ_{ij} are calculated using the methods described in Section 3.1, where the variable y_{it} is replaced by $y_{it} - \hat{\theta} w_{it}$. The HAC weights, ϕ_{ij}^2, are based on the estimator of $\widehat{\mathbb{V}}(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\tilde{u}_{it} + (\widehat{\theta} - \theta^0)\tilde{w}_{it})\tilde{x}_{jt})$[7]. Note that for the real data the value θ^0 is unknown and thus it will be assumed that $\widehat{\theta} - \theta^0 = 0$.

3.3 Comparison between the models

As this paper will estimate different models to measure the spillover effects, the performance of these models is of the utmost importance. To determine which model provides the best estimates this paper will generate some data according to equation (1) where the values of $a_i, \beta_i, \gamma_{ij}$, and θ are generated according to a data generating process explained in Section 4.1. Then the multiple techniques discussed in section 3.1 and 3.2 will be used to estimate those parameters.

To determine the best-performing model, it is useful to consider which model can provide estimates that closely resemble the actual value of Γ. One way to do this is to see how many variables are correctly given a non-zero value. This can be done by dividing the number of correctly estimated non-zero values by the actual number of non-zero values. However, it could also be that a certain method sets too many variables to a non-zero value. To that end one should also see how many variables are correctly set to zero. This can be done by dividing the number of correctly set zero values by the actual number of zero-values. Note that for both of these percentages we would want it to be as close to 100% as possible. Another way to assess the effectiveness of an estimator is to compare the estimated values with the actual values. This is important because even if the estimator selects too many variables, it may still accurately

[5]The precise expressions for the weights and the initial weights of the HAC method can be found in the Appendix Section C.

[6]It should be noted that $\widehat{\theta}\tilde{w}_{it}$ is equal to $\sum_{d=1}^{D} \widehat{\theta}^d \tilde{w}_{it}^d$

[7]The precise expressions for the weights and the initial weights of the HAC method can be found in the Appendix Section C.

estimate the values of the actual non-zero variables while setting the values of the actual zero variables close to zero. It could also be that the model selects mostly correct variables, but that the estimates of these variables deviate significantly from the true values, which is undesired. To that end the Frobenius norm of the difference between the estimated model and the actual gamma matrix will be calculated using the formula

$$\|\Gamma - \widehat{\Gamma}^P\|_F = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} |\gamma_{ij} - \widehat{\gamma}_{ij}^P|^2}, \tag{14}$$

where γ_{ij} is the actual value of Γ and $\widehat{\Gamma}_{ij}^P$ is the estimated value of Γ. In this case it is preferred to have the Frobenius norm as low as possible, as a Frobenius norm of zero implies that the estimated model predicts the values with an accuracy of 100%. Note that this comparison with the real Γ is only possible for the simulated data and not for the actual data, as for the actual data Γ is unknown.

Another performance measure is the Mean Squared Error (MSE) which can be calculated with the formula

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \widehat{y}_{it})^2,$$

where y_{it} is the actual value of the observation and $\widehat{y}_{it}$ is the predicted value of y_{it} using the estimated model. To estimate $\widehat{y}_{it}$, $\widehat{\Gamma}^P$ is used together with $\widehat{\theta}$ to estimate the fixed effect $\widehat{a}_i$:

$$\widehat{a}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} - \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\widehat{\gamma}_{ij}^P x_{jt}\right) - \frac{1}{T}\sum_{t=1}^{T}\widehat{\theta}w_{it},$$

and then using all these estimates it becomes possible to calculate $\widehat{y}_{it}$:

$$\widehat{y}_{it} = \widehat{a}_i + \sum_{j=1}^{N}\widehat{\gamma}_{ij}^P x_{jt} + \widehat{\theta}w_{it}.$$

The MSE is useful to determine how close a model can predict values compared to its actual value. It can be used both on actual data and on simulated data.

Once these performance measures are computed for the Pooled estimator, this paper will choose the best performing Elastic-Net models and the best performing Lasso models. These models will then be used on the Double Pooled estimator and on the actual data to estimate the spillover effects. For the actual data, the MSE will be calculated. It is interesting to see whether the same model as before performs best in a real-life application, where there might be a larger dependence between observations and fewer assumptions hold.

4 Simulation

As explained in section 3 two main models are compared with different expressions for weights and penalty terms. It should be noted that in this section Elastic-Net is always estimated with $\alpha = 0.5$. The remaining of this section will be divided as follows. First the general data generating process (DGP) will be discussed. Then, the 20 different models as described in Table 2 will be estimated without control variables. In the simplified model, it is helpful to determine which estimation technique performs the best before moving on to more complicated methods. If a certain method is already performing poorly for the Pooled method, it might not be advisable to use it for the Double pooled method. Lastly, the three best performing Lasso and the three best performing Elastic-Net models will be chosen and the Double Pooled technique will be applied using these models. In the entire simulation studies $N = 263$, as this equals the actual amount of companies in the dataset. The number of years, T, will both be 21 years and 100 years. It is interesting to see how well the methods perform if the number of companies and the time period closely resembles the actual data. If the methods already perform poorly in the simulation, where the model we attempt to estimate follows the right model specification, we cannot expect to get very reasonable results when applying this model to the data. The reason for also focusing on a larger time frame, is to see whether the models perform better when more information is available and to thus determine whether the applied models are appropriate for estimating spillover effects.

4.1 Data generating process

The DGP is designed so it can be tested for different number of years, and different number of companies. The simulated data is created to closely mimic real-world data, while also offering reliable estimates. First, this subsection discusses the DGP of the variables followed by the DGP of the parameters.

First of all, the DGP of the variables is based on the logarithm of the actual data. Furthermore, to keep the model realistic, for each company a variable is taken from a normal distribution with mean and variance of the logaritm of the actual data. Then for each time period, it is assumed that the observations are again from a normal distribution, but with a smaller standard deviation. This is done such that for each company the observations over time are reasonably close together, as it is unlikely that there are big differences within a timespan of a few years. Furthermore, since the generated data is based on the logarithm of the actual data, only values of 0 and larger are being generated. As the actual data contains no negative numbers and $log(1 + x)$ is taken for all variables, all log-values are equal or larger than zero. The final DGP of the variables can be described by the following equations:

$$x_i \sim \mathcal{N}(4.7, 2.2),\ \forall i \in \{1, \dots, N\} \text{ and } x_{it} \sim \mathcal{N}(x_i, 1),\ \forall t \in \{1, \dots, T\},$$

$$w_i^1 \sim \mathcal{N}(2.1, 1.4),\ \forall i \in \{1, \dots, N\} \text{ and } w_{it}^1 \sim \mathcal{N}(w_i^1, 0.5),\ \forall t \in \{1, \dots, T\},$$

$$w_i^2 \sim \mathcal{N}(5.4, 2.1),\ \forall i \in \{1, \dots, N\} \text{ and } w_{it}^2 \sim \mathcal{N}(w_i^2, 1),\ \forall t \in \{1, \dots, T\},$$

$$u_{it} \sim \mathcal{N}(0, 1),\ \forall i \in \{1, \dots, N\}\ \forall t \in \{1, \dots, T\}.$$

It should be noted that all values for x_{it}, w_{it}^1 and w_{it}^2 which are smaller than zero, are set to zero.

This paragraph will discuss the parameter choices. The spillover effects $\gamma_{ij} \neq 0$ for $i \neq j$ are assumed in the simulation to only take place between firms with the same Standard Industrial Classification (SIC). To that end each company gets assigned a random SIC, which is a number ranging from 1 to 50. Thus, a spillover $\gamma_{ij} \neq 0$ is assigned to company i if company j and company i have the same value for SIC. This implies that there are on average 5 spillovers per company. Furthermore, the DGP for the parameters can be described by the following expressions:

$$a_i \sim \mathcal{N}(2, 1),\ \forall i \in \{1, \ldots, N\}$$

$$\beta_i = \gamma_{ii} \sim \mathcal{N}(3, 0.5),\ \forall i \in \{1, \ldots, N\}$$

$$\gamma_{ij} \sim \mathcal{N}(1, 0.2),\ (i, j) \in \{1, \ldots, N\},\ \text{for } i \neq j,\ \text{and } SIC_i = SIC_j$$

$$\theta^1 \sim \mathcal{N}(1, 0.5) \text{ and } \theta^2 \sim \mathcal{N}(0.5, 0.2)$$

These coefficients are chosen based on the assumption that the logarithm of y is not excessively large, thus the coefficients should also not be too large. However, the coefficients should not be made too small either because in that case Lasso and Elastic-Net might not detect these numbers as different from zero and might not select these numbers. Therefore, the numbers are set in such a way that Lasso and Elastic-Net can detect them without being set so extremely large that it becomes unrealistic.

4.2 The pooled estimator

Our initial focus is on evaluating the performance of the Pooled Lasso (Elastic-Net) estimator with varying weights and penalty parameters. To that end equation (2) is used to generate the y-variable and to estimate the final model. This equation is utilized to ensure the appropriate DGP is applied to the variable y, thereby avoiding the inclusion of any additional variables that may negatively impact the model's performance. An incorrect model specification could make models perform poorly, even when the model used for estimation is appropriate. The remaining of this subsection will be divided into multiple parts. First, the general observations will be discussed. Then the two penalty terms will be compared, followed by a discussion on the performance of the different types of weights. Subsequently, a comparison between the Elastic-Net and Lasso will be given. Finally, the three best Lasso and Elastic-Net models will be selected.

4.2.1 General observations

Applying all the models of Table 2 to the data, the performance measures in Table 3 are obtained. These performance measures include 21 years of data and these are the means of the performance measures taken over 100 simulations[8]. None of the models seem to perform very well in terms of accurately predicting spillovers, where of the actual values where $\gamma_{ij} \neq 0$ none of the models can set more than 40% of those parameters at a non-zero value. On top of that, it should be taken into consideration that the Post Pooled estimator forces the model to select the cross-products $\gamma_{ii}, \forall i \in \{1, \ldots, N\}$. Therefore, the percentage correct non-zero is inflated because some parameters are still selected, even if Lasso or Elastic-Net does not select them. Therefore also the measure correct spillovers is considered, which shows the percentage of the spillovers

[8]The corresponding standard deviations of these performance measures can be found in Table 8 in Section D in the Appendix.

Table 3: The mean of the performance measures using 100 simulations over 21 years of data

Method	F-norm	% non-zero	% spillovers	% zero	MSE
Panel A: Lasso methods					
EW static	44.01	19.90%	4.63%	99.79%	197.55
Var(x) static	70.11	19.96%	4.70%	98.08%	1,390.07
HAC static	55.57	19.21%	3.81%	99.26%	642.58
Lasso static	44.04	19.93%	4.67%	99.79%	198.69
Ridge static	126.66	37.95%	26.12%	95.03%	68,984.82
EW CV	299.47	38.04%	26.24%	95.74%	159,418.52
Var(x) CV	657.82	23.21%	8.58%	96.63%	908,440.68
HAC CV	321.63	30.74%	17.53%	95.83%	41,347.67
Lasso CV	50.64	32.69%	19.87%	97.97%	367.71
Ridge CV	489.59	37.93%	26.10%	94.82%	106,341.37
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	71.79	34.38%	21.87%	97.18%	3,526.01
Var(x) static	582.48	26.42%	12.40%	95.27%	240,295.10
HAC static	202.87	29.88%	16.51%	96.43%	72,266.26
Enet static	71.75	34.38%	21.87%	97.18%	3,525.13
Ridge static	13,243.09	39.85%	28.38%	93.57%	222,024,100.00
EW CV	2,298.26	37.39%	25.46%	94.41%	3,233,280.00
Var(x) CV	2,991.83	24.47%	10.08%	95.77%	24,320,420.00
HAC CV	2,895.14	31.65%	18.62%	94.69%	11,825,140.00
Enet CV	74.95	34.08%	21.52%	97.11%	984.88
Ridge CV	2,672.13	39.11%	27.51%	93.90%	53,656,300.00

Notes: This table shows the mean of the performance measures of the methods considered for 100 simulations over 21 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

$\gamma_{ij}, i \neq j$ the model does not set to zero, when the value is non-zero. This percentage is even smaller, so that all models correctly select less than 30% of the spillovers.

In contrast, the percentage of observations which are set correctly to zero is much larger, above 93% for all models. Nevertheless, it should be noted that there are far more observations for which $\gamma_{ij} = 0$, than for which $\gamma_{ij} \neq 0$. As there are approximately 5 spillovers between companies together with the fact that $\gamma_{ii} \neq 0$, $\forall i \in \{1, \ldots, N\}$, there are approximately 67,591 observations in the actual matrix Γ which are zero and only 1,578 observations which are non-zero. Therefore, an increase of 1% in correct non-zero implies that approximately 16 more variables are correctly set to non-zero, while a decrease of 1% in correct non-zero implies that approximately 676 variables are wrongly set to a non-zero value. Therefore, even when a model has higher percentages of correct non-zero and correct spillovers, a lower percentage of correctly identified zeros may result in a model that produces more miscalculated spillovers than accurate ones. In order to improve the model the increase in the percentage correct non-zero must thus be more than 40% to mitigate the effect of a 1% decrease in the percentage correct zero. Thus, when selecting a model more value should be given to the percentage correct zero compared to the percentage correct non-zero or correct spillovers.

Thus for 21 years of data none of the models considered in this paper seemed to perform very well. However, this could potentially be because of the small value of observations T. Therefore, it is important to determine whether the model performs better when a larger time horizon is used. To that end all the models are re-estimated with 100 years of data and 50 simulations[9]. The mean of the performance metrics of the 50 simulations can be found in Table 4[10]. When we increase the time horizon T, there is a noticeable increase in both the percentage of correct non-zero values and correct spillovers. For $T = 21$ all percentages correct non-zero and correct spillovers were below 40%, while for $T = 100$ these percentages are all higher than 54%. Furthermore, the Frobenius norm and the MSE have decreased a lot compared to 21 years of data. This is not unexpected, as more years of data imply more available information which leads to more accurate predictions of the coefficients, and thus also the outcome y. Further increasing the time horizon makes the results even better.

4.2.2 Comparison between the penalty terms

For deciding which models are the best models, first the two penalty terms are compared against each other. If we inspect a time period of 21 years for the Lasso methods, it immediately becomes clear that using the static formula for λ instead of using CV leads to a lower Frobenius norm and a lower MSE for all methods. The percentage for correct non-zero and correct spillovers is generally higher for the methods that use CV to compute λ for Lasso, but this is paired with a lower percentage for correct zero. For Lasso, the static λ thus seems to outperform the λ calculated using CV.

Then looking at a larger time period of 100 years for the Lasso methods, the λ calculated with CV outperforms the static λ in most cases for all performance measures except the percentage correct zero. The static value of λ seems only outperform almost all λ with CV methods in terms of percentage correct zero. Taking into account that the percentage correct zero has more value than the percentage correct non-zero and correct spillovers, the static value of the penalty parameter might still be a better fit for Lasso models. Nevertheless it should be taken into account that the overall best performing model from 100 years of data is the model which uses Lasso weights with CV for λ.

For Elastic-Net and 21 years of data, the Frobenius norm is generally lower for the methods that use the static λ. Also the MSE is generally lower for the static value of λ, with the exception of Ridge weights Elastic-Net weights. For the percentages correct non-zero, correct spillovers and correct not a certain value of the penalty parameter clearly outperforms the other. Most values are quite close with only a minimal difference between the two penalty parameters. Generally, the differences between the methods are small, except for using equal weights and the HAC weights. Therefore, for Elastic-Net there is no clearly best-performing penalty parameter.

[9]Only 50 simulations are used for 100 years of data because the estimation of the model takes longer when using 100 years instead of 21 years. In order to keep the estimation time feasible less simulations are used for 100 years of data.

[10]The corresponding standard deviations can be found in Table 9 in Section D in the Appendix.

Table 4: The mean of the performance measures using 50 simulations over 100 years of data

Method	F-norm	% non-zero	% spillovers	% zero	MSE
Panel A: Lasso methods					
EW static	23.18	62.74%	55.63%	99.98%	145.80
Var(x) static	25.02	61.80%	54.50%	99.53%	172.34
HAC static	23.44	62.78%	55.67%	99.91%	147.59
Lasso static	21.53	66.68%	60.32%	99.98%	138.41
Ridge static	36.22	98.58%	98.30%	89.51%	283.40
EW CV	15.28	99.92%	99.91%	89.63%	57.42
Var(x) CV	17.34	99.93%	99.92%	88.12%	70.92
HAC CV	16.13	99.94%	99.93%	89.18%	62.24
Lasso CV	6.01	99.14%	98.98%	99.86%	19.66
Ridge CV	14.41	97.26%	96.74%	95.49%	52.75
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	9.44	96.13%	95.40%	99.20%	36.50
Var(x) static	11.77	95.97%	95.20%	97.98%	46.72
HAC static	10.05	96.47%	95.80%	98.87%	38.26
Enet static	8.23	97.69%	97.25%	99.20%	29.75
Ridge static	288.37	98.85%	98.63%	82.01%	549,144.11
EW CV	17.94	99.91%	99.89%	85.31%	71.76
Var(x) CV	20.44	99.92%	99.90%	83.85%	92.99
HAC CV	19.08	99.92%	99.91%	84.82%	82.40
Enet CV	6.45	99.11%	98.94%	99.77%	20.86
Ridge CV	15.06	97.26%	96.74%	94.77%	55.36

Notes: This table shows the mean of the performance measures of the methods considered for 50 simulations over 100 years of data of 263 companies. The x-variables are i.i.d. generated variables. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

However, there seems to be a slight preference for the static λ, because when it outperforms the λ computed with CV, it mostly does it quite clearly.

Increasing the time period to 100 years does not give clearer results for Elastic-Net. For the Elastic-Net and Ridge weights, the λ computed with CV seems to be a better fit in terms of almost all performance measures. For the other three weights the static value of λ performs better for all performance measures except the percentages correct spillovers and correct non-zero. Generally, the static approach for λ seems to be slightly preferred, especially when we take into consideration that more value is given to the performance measure percentage correct zero. However, it should be taken into account that the Elastic-Net model with CV for λ and weights based on Elastic-Net seems to have the best performance overall.

All in all, the static method used in the paper of Manresa (2016) seems to generally perform best for most methods. However, for some specifications of the weights computing λ with CV actually outperforms all of the other methods, and thus this method should not be forgotten.

4.2.3 Comparison between the weights

Next, we compare the different weights that are used in this paper. If we inspect a time period of 21 years for the Lasso methods, the Lasso weights seem to mostly outperform all other methods in terms of Frobenius norm, percentage correct zero and MSE. The only method that performs better, using the static λ, is the equally weighted model, however this method performs poorly when using CV to compute λ. Using the weights of Ridge and the weights of the sample variance of x perform poorly for both values of λ. The HAC weights of Manresa (2016) do not perform very well, but also not exceptionally bad. Currently not many methods outperform the method without weights, but that could be due to the small value of T.

Looking at a larger time horizon of 100 years, the results do not change a lot. The Lasso weights outperform the other weights even more clearly now, for all performance measures except the percentage correct non-zero and correct spillovers. For the static value of λ the equally weighted model and the model with HAC weights now almost have the same performance and both perform quite well. Also noteworthy is that Ridge with λ computed with CV also performs suddenly quite well.

For the Elastic-Net methods and a time period of 21 years, the model with Elastic-Net weights outperforms all other weights for all performance measures except the percentage correct non-zero and correct spillovers. The performance of the equally weighted model is comparable to that of the Elastic-Net weights when using the static λ. However, it performs poorly when using the λ computed with CV. The Ridge weights seem to have perform the worst, followed by the weights based on the sample variance of x and the HAC weights. The underwhelming performance of the methods that incorporate weights, except for Elastic-Net weights, could once again be because of the low value of T.

Increasing the time period to 100 years, the results remain the same. The Elastic-Net weights still outperform all other models for all performance measures except the percentage correct non-zero and correct spillovers. For the static value of λ the equally weighted model and the model with HAC weights have become closer, with their performance slightly worse than the Elastic-Net weights.

4.2.4 Comparison between Lasso and Elastic-Net

The next step is to compare the Lasso methods with the Elastic-Net methods. For 21 years of data, the Lasso methods outperform the Elastic-Net models in terms of Frobenius norm, MSE and the percentage correct zero. The Elastic-Net models only perform better than Lasso for the percentage correct non-zero and correct spillovers. Thus, Elastic-Net can detect more spillovers compared to Lasso, but by doing so it also detects “false” spillovers, where it sets values not equal to zero while in actuality they are zero. As the percentage of correct spillovers for Elastic-Net is not large enough to mitigate the effect of the decrease in the percentage correct spillovers, the Lasso methods outperform the Elastic-Net models for 21 years of data.

For 100 years of data, Lasso no longer seems to clearly outperform Elastic-Net. For the static value of λ most of the Elastic-Net methods now outperform the Lasso methods for all performance measures except the percentage correct zero. The static Lasso methods have a higher percentage correct zero, however for the equally weighted method and the Elastic-Net weights the higher percentage correct spillovers for Elastic-Net is high enough to mitigate the effect of a lower percentage correct zero. Therefore for the static value of λ Elastic-Net seems to be the best method. However, if CV is used to compute λ the Lasso methods outperform the Elastic-Net methods for all performance measures.

The fact that the Elastic-Net methods do not consistently outperform the Lasso methods could be because up until now all x-variables are i.i.d. generated. Therefore the variables up until now did not contain multicollinearity, and Elastic-Net is mainly expected to outperform Lasso when there is multicollinearity in a model. In real life scenarios it is also more likely that variables are correlated, as events could happen that affect multiple companies in the same way. Therefore also two different scenarios with correlation between the x-variables are considered. The first scenario generates random values of x with correlation 0.5 between companies if the two companies have spillover effects and correlation 0.1 between companies if the two companies do not have spillover effects. When companies experience spillovers, it is probable that there is a stronger correlation between them. If there are no spillover effects between two companies, these companies are likely less related and may only respond similarly to significant changes in the economy. As a result, their correlation should be lower. In the second scenario, random values of x are generated with a correlation of 0.5 across all companies. Although less realistic in real life, it is interesting to observe whether the models can still perform well as the correlation increases.

Table 12 and Table 14 in Appendix Section D show the mean of the performance measures of the models where the x-variable is estimated with correlation of scenario one for 21 and 100 years of data respectively. For 21 years of data Elastic-Net only outperforms Lasso if Elastic-Net weights are used, and if equal weights are used with the static value of λ. Interestingly, Lasso does not seem to perform worse with this correlation structure, and still mostly outperforms Elastic-Net. For 100 years of data Lasso outperforms the Elastic-Net methods, except for Elastic-Net weights with CV for λ. Once again both Lasso and Elastic-Net methods perform better when these correlations are used, even though surprisingly Lasso still outperforms Elastic-Net.

Next we observe Table 16 and Table 18 in Appendix Section D, which show the mean of the performance measures of the models where the x-variable is estimated with a correlation of 0.5 between all companies for 21 and 100 years of data respectively. In this case most models actually perform worse compared to not including correlations. Furthermore, in almost all cases Lasso outperforms Elastic-Net with this correlation structure, the only exception is Elastic-Net weights with CV for λ. For 100 years of data actually many models perform better with correlations compared to no correlations, especially the models with a static λ. However, the Lasso models mostly still outperform the Elastic-Net models.

It is quite surprising that Elastic-Net does not seem to clearly outperform the Lasso methods when correlations are introduced, as that is what one would expect based on theory (Zou & Hastie, 2005). However, up until now only $\alpha = 0.5$ is considered. It is possible that the current value of α may not be appropriate, and other values of α might result in Elastic-Net performing better than Lasso in the presence of correlations. Further research could potentially find a better value of α.

Now that all models are compared for different values of T, we choose the three best performing Lasso and the three best performing Elastic-Net models. For Lasso the methods Lasso static and Lasso CV and for Elastic-Net the methods Enet static and Enet CV are chosen because in all simulations they consistently showed to be one of the best estimators for $\widehat{\gamma}^P$. Furthermore, for both Elastic-Net as well as Lasso the method HAC static is chosen. This method is chosen as the paper of Manresa (2016) specifically applied it to the Double Pooled estimator, and thus we expect it to perform quite well. Furthermore, this method has done quite well when the value of T increased, showing that asymptotically this method provides good estimates.

4.3 The double pooled estimator

For the double pooled estimator equation (1) is used where y_{it} is generated with control variables. Once again the estimation is done with 263 companies. First general observations are discussed, followed by a comparison of the different models. Finally, the Lasso and Elastic-Net models will be compared against one another.

Table 5: The mean of 100 simulations of the performance measures of the Double Pooled methods over 21 years of data

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	55.49	19.17%	3.69%	99.27%	628	0.39	0.19
HAC sc static	55.48	19.18%	3.70%	99.27%	629	0.39	0.19
Lasso static	43.86	20.03%	4.73%	99.79%	195	0.27	0.13
Lasso CV	63.14	31.78%	18.73%	97.90%	2,947	0.80	0.38
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	3,438.04	29.81%	16.38%	96.29%	475,115,200	0.90	0.43
HAC sc static	3,438.04	29.81%	16.38%	96.29%	475,115,200	0.90	0.43
Enet static	59.96	34.28%	21.70%	97.01%	523	0.87	0.42
Enet CV	126.78	33.39%	20.64%	97.01%	7,760	0.87	0.41

Notes: This table shows the mean of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

4.3.1 General observations

Table 5 shows the results of the Double Pooled method using 21 years of data[11]. Similar to the Pooled method the percentage correct spillovers is quite small, below 22% for all methods.

[11] The standard deviation of this table can be found in Table 20 in Appendix Section D.

Furthermore, the estimates of the θ are quite inaccurate, especially for the Elastic-Net methods. If we take into account that $\theta_1 \sim \mathcal{N}(1, 0.5)$ and $\theta_2 \sim \mathcal{N}(0.5, 0.2)$, then the mistake in calculating θ is indeed quite big. A possible explanation for the big calculating mistake in Elastic-Net could be that Elastic-Net might set too many values of $\lambda_{ij} \neq 0$, thus implying that x_j and w_i are related when they are not in reality, which then leads to a miscalculated value of θ. Additionally, we also observe that the HAC static model and the HAC scaling static model have almost equal performance. This is a good sign as that implies that using the estimated error without using the actual parameters of the original model can provide just as good estimations as estimating the errors with the actual parameters. So even when the difference between the actual and estimated parameter is quite large it does not seem to seriously alter the performance of the estimated weight matrix.

When we re-estimate the models using a time horizon of 100 years the results in Table 6 are obtained[12]. Once again all the performance metrics improve when using a larger time horizon. The two HAC methods are still approximately equal, thus even for larger time periods the inaccuracy in estimating the θ does not seem to lead in very different estimates for the weights ϕ_{ij}.

Table 6: The mean of 50 simulations of the performance measures of the Double Pooled methods over 100 years of data

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	23.78	62.24%	55.12%	99.90%	158.29	0.10	0.06
HAC sc static	23.78	62.24%	55.12%	99.90%	158.22	0.10	0.06
Lasso static	21.81	66.17%	59.78%	99.98%	146.86	0.09	0.05
Lasso CV	6.18	99.04%	98.86%	99.85%	20.55	0.12	0.07
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	10.32	96.35%	95.66%	98.82%	40.24	0.19	0.11
HAC sc static	10.32	96.35%	95.66%	98.82%	40.24	0.19	0.11
Enet static	8.51	97.53%	97.06%	99.17%	31.43	0.17	0.10
Enet CV	6.69	98.99%	98.80%	99.76%	21.75	0.14	0.08

Notes: This table shows the mean of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

4.3.2 Comparison between the different models

Next the different methods are compared. For 21 years of data and for the Lasso methods, the Lasso static method outperforms the other three methods for all performance measures

[12]The standard deviation of this table can be found in Table 21 in Appendix Section D.

except the percentage correct non-zero and correct spillovers. Lasso with CV has the highest percentage of correct non-zero and correct spillovers, however this comes at the cost of a much lower percentage of correct zero. Lasso with CV performs worst overall, and also estimates the value of θ quite poorly. A possible explanation could be that because CV is used for λ, the estimates of λ_{ij} could be very inaccurate leading to a bad estimation of θ. All in all, the Lasso static method clearly outperforms all other methods.

For 100 years of data, Lasso static no longer seems to perform best. In this case the Lasso method with CV outperforms all other methods in terms of all performance measures except the percentage correct zero and the accuracy of θ. The other three methods have a higher percentage correct zero, but that difference is minimal compared to the huge increase in the percentage correct spillovers that the Lasso weighted model with CV for λ achieves. Furthermore, the three other models give better estimates for the values of θ, but again this difference is minimal. Lasso static is now the second best performing model, as Lasso with CV is in this case the best model.

For Elastic-Net and 21 years of data the Elastic-Net static method outperforms the other methods for almost all performance measures. Only Elastic-Net with CV for λ can estimate θ_2 slightly more accurate compared to Elastic-Net with a static penalty parameter. The HAC weights with a static λ truly underperform and give the worst estimates for all performance measures. Thus, using Elastic-Net weights with a static λ seems to perform best for a small sample size.

For 100 years of data, once again the static Elastic-Net model no longer performs best but rather the Elastic-Net model that uses CV for computing λ outperforms all other models for all performance measures. This method outperforms all other methods for all of the performance measures. Hence, this is clearly the best method for Elastic-Net. Furthermore, the static method which uses HAC weights have the worst performance. Using Lasso and Elastic-Net weights give the best performance and if the time period is small a static value of λ should be used, while for a larger time horizon λ should be computed with CV.

4.3.3 Comparison between Lasso and Elastic-Net

Comparing the Elastic-Net and Lasso methods for 21 years of data, the Lasso methods seem to outperform the Elastic-Net methods in terms of almost all performance measures. Elastic-Net only outperforms the Lasso methods in terms of percentage correct non-zero and percentage spillovers, but this higher percentage does not mitigate the much lower percentage of correct zero. Thus, Lasso clearly outperforms Elastic-Net. For 100 years of data the comparison is more complicated. For the three static methods, Elastic-Net outperforms Lasso for most performance measures, except the percentage correct zero and accuracy of θ. Thus, for a static value of λ Elastic-Net seems to be preferred. For the method that uses CV, Lasso seems to outperform Elastic-Net in all performance measures. This is also the best performing model, and thus overall Lasso seems to provide the best estimates for the spillover effects.

Up until now the x-variables were i.i.d. and thus had a minimal amount of correlation between variables. However, one would expect Elastic-Net to outperform Lasso when correlations between the x-variables are introduced and therefore the same scenarios of correlations as discussed in section 4.2 will be applied here.

For the scenario where there is a correlation of 0.5 between spillover companies and 0.1 otherwise, Table 22 and Table 24 in the Appendix Section D are generated with 21 and 100 years of data respectively. For 21 years of data, the two HAC methods estimated using Lasso, outperform the two HAC Elastic-Net models. However, the two models that use Elastic-Net weights seem to have a slightly better performance compared to the two models that have Lasso weights. However, it should be noted that this difference is minimal. For 100 years of data Elastic-Net only outperforms Lasso if Elastic-Net weights are used and λ is computed with CV. However, the best performing model is a Lasso model with HAC weights and a static λ. Thus, even when this type of correlations are introduced Elastic-Net does not (consistently) outperform Lasso.

For the scenario with equal correlations of 0.5 between all companies Table 28 and Table 30 in the Appendix Section D are generated with 21 and 100 years of data respectively. In this case for 21 years almost all Lasso methods outperforms almost all the Elastic-Net models, except that Elastic-Net with CV for λ performs slightly better for some performance measures. For 100 years of data the static methods perform better when using Lasso instead of Elastic-Net. However, the Elastic-Net model with CV slightly outperforms the Lasso model with CV for most performance measures. However, this difference is so small that we can conclude that with this type of correlations Elastic-Net does still not outperform Lasso.

Thus similar to before and against common expectations, Lasso seems to perform better than Elastic-Net when there are correlations between the variables and this effect persists for different values of T. However, once again one should bear in mind that only $\alpha = 0.5$ is used. Hence, attempting different values of α might enhance the performance of Elastic-Net compared to Lasso.

Finally, there does not seem to be one best performing model which consistently outperforms all other models. However, based on the small sample size in the actual data one could expect that the Lasso static or the Lasso CV method will have the best performance. However, if one attempts different values of α for Elastic-Net, it could possibly be that the static Elastic-Net approach outperforms (most) of the other methods.

5 Real life application

The models discussed in section 3 will be applied on company data. The main assumption here is that the estimation structure follows a Cobb Douglas production function:

$$Y_{it} = L_{it}^{\theta_L} C_{it}^{\theta_C} K_{it}^{\beta_i} SK_{it} A_{it}, \tag{15}$$

where Y is a measure for output, in this case sales. L is a measure for labor, measured by the number of employees of a company. C is physical or tangible capital, measured by the net stock of property, plant, and equipment. K is knowledge capital, measured by capital stock of R&D at time $t-1$. SK is a measure for knowledge spillovers, and A represents technological progress. Similar to before subscript i represents a company and subscript t represents the time period. The parameters θ_L, θ_C, and β_i are elasticities, and thus represents the sensitivity of the output Y to one of the inputs, L, C, and K. $\theta_L = \varepsilon^{Y_i}_{L_i}$ represents firm i's output elasticity to labor, while $\theta_C = \varepsilon^{Y_i}_{C_i}$ represents firm i's output elasticity to capital. $\beta_i = \varepsilon^{Y_i}_{K_i}$ is then firm i's output elasticity to its own knowledge capital. For the spillover effects SK_{it}, a specification is used where other firm's knowledge capital generates spillovers for another company, such that the production function in equation (15) becomes:

$$Y_{it} = L_{it}^{\theta_L} C_{it}^{\theta_C} K_{it}^{\beta_i} \prod_{j \neq i} K_{jt}^{\gamma_{ij}} A_{it}. \tag{16}$$

Within this equation $\gamma_{ij} = \varepsilon^{Y_i}_{K_j}$ represents firm i's output elasticity to the knowledge capital of firm j. Note that within this equation the technological progress A_{it} has both a subscript i and a subscript t. This implies that the technological progress factor has both a time-invariant component which captures a firm-specific shock across firms as well as a time-varying component which captures time shocks to productivity. Furthermore, A also includes a time and firm-specific idiosyncratic component, which will be reflected in the error term. For the sake of simplicity, the time-varying aspect of A will be disregarded in this paper due to the relatively short time period. Therefore, A_{it} will be simplified to A_i. Finally, in order to estimate this model the log is taken such that instead of a multiplication of the different production inputs, the production inputs become additive:

$$y_{it} = \alpha_i + \beta_i k_{it} + \sum_{j \neq i} \gamma_{ij} k_{jt} + \theta_L l_{it} + \theta_C c_{it} + u_{it}, \tag{17}$$

where the lower-case letters denote the log of the capital letters in equation (16). Note that α_i is the time-invariant firm-specific component of technological progress and is unknown thus it should be estimated and u_{it} is the time and firm-specific idiosyncratic component.

It is interesting to compare model (17) with model (1) described in section 3. In this case the variable x_{it} in equation (1) are replaced by k_{it}, the R&D stock of company i. Also equation (17) contains two control variables w^d_{it}, which are l_{it}, the labor of the company, and c_{it}, the capital stock of the company. Thus, θ consists of two values which should be estimated, namely $\widehat{\theta}_C$ and $\widehat{\theta}_L$. Next, we estimate the models and compare the performance followed by an economical interpretation of the obtained results.

5.1 The results

Applying the model to the data two different structures are applied. Firstly, the Pooled Lasso and Elastic-Net approach are used, thus the approach without control variables. This is done in order to determine whether a more complicated approach is truly necessary and is able to improve the MSE of the different models. Also the Double Pooled method is applied using equation (17). The results of these models can be found in Table 7.

As Table 7 shows, the MSE is for almost all methods lower when switching from the Pooled estimation approach toward the Double Pooled approach. Only for the static λ and the Elastic-Net weights this is not the case, but this method performs poorly in the real life application and thus not too much value should be taken from this. That the MSE decreases when using the Double Pooled estimator is a strong indicator of the fact that including control variables is

Table 7: The MSE and other measures of the different models applied to the data

Panel A: Lasso methods				
Method	**MSE**	**% non-zero**	$\widehat{\theta}_L$	$\widehat{\theta}_C$
Pooled HAC static	28,771.24	1.13%	0	0
Pooled Lasso static	0.08	0.38%	0	0
Pooled Lasso CV	1,614.58	2.05%	0	0
Double Pooled HAC static	8,858.45	0.73%	0.68	0.45
Double Pooled Lasso static	0.06	0.38%	1.11	0.65
Double Pooled Lasso CV	864.80	2.00%	0.43	0.15
Panel B: Elastic-Net methods				
Method	**MSE**	**% non-zero**	$\widehat{\theta}_L$	$\widehat{\theta}_C$
Pooled HAC static	32,469.47	2.35%	0	0
Pooled Elastic-Net static	5,711,590,496.72	0.69%	0	0
Pooled Elastic-Net CV	2,296.42	2.65%	0	0
Double Pooled HAC static	10,960.29	1.68%	0.45	0.27
Double Pooled Elastic-Net static	753,046,025,756.13	7.60%	3,155.95	0.05
Double Pooled Elastic-Net CV	1,147.77	2.54%	0.46	0.16

Notes: This table shows different measures of the methods considered of the model applied to the data. This entire table is based on in-sample estimation with $\alpha = 0.5$ for Elastic-Net. % non-zero represents the percentage of parameters γ_{ij} which are given a non-zero value. $\widehat{\theta}_L$ is the estimated value of θ_L and $\widehat{\theta}_C$ is the estimated value of θ_C. Pooled implies the model is estimated without the control variables, $\theta = 0$, while Double Pooled implies the model is estimated with control variables. Static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) is used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ is used for weights and for Elastic-Net the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

necessary to make models more accurate. Thus, it seems that the Double Pooled approach does not over-complicate the model, but rather improves the model.

Another observation that is quite peculiar is that the Lasso static model obtains a much lower MSE compared to all other models. It even obtains a lower MSE than in the simulation, which is really strange as in that case we know that the model to be estimated follows the right model structure. The reason for the low MSE can be explained by the percentage of non-zero values. The Lasso static model sets all values $\gamma_{ij} = 0$ if $i \neq j$, it thus estimates no spillover effects. Remember that the Post Pooled estimator always selects the variables x_{ii} and thus $\gamma_{ii} \neq 0$. If there are no estimated spillover effects, there are also no "false" spillovers and the estimator can set each value of γ_{ii} in such a way that the MSE is very small. Another possible explanation for the low MSE could be that there are simply no spillover effects between the companies considered in this dataset. If that were to be the case then it is of course quite logical that the model that estimates no spillover effects obtains a low MSE.

The MSE of the other methods is still quite sizable. A model with a high MSE is an issue because such a model cannot accurately predict the dependent variable y_{it} and thus the estimates are not very accurate. Thus, since the model is not predicting the variable y very well, this could mean that the estimates for $\widehat{\gamma}_{ij}^P$ are also not reliable, which is something to bear in mind when drawing conclusions from the data. A likely explanation for this result is that the time horizon which is available for estimation, 21 years, is simply not large enough for obtaining an accurate estimation. This also can be seen in section 4, where indeed for 21 years of data the Double Pooled estimator has quite a sizable MSE even when the right model structure is applied.

However, even though the fact that there is not enough data available for a reliable estimator there might be also other factors present which inflate the MSE. Comparing these values of MSE with those in section 4, the MSE in this table is a lot larger compared to the MSE in section 4. Therefore another possible explanation for the high MSE could be that even though two control variables are included, there might be even more control variables which were not included. Thus, there might be omitted variable bias within this model.

Another factor which might be the cause of the high MSE could be the fact that there is only a common effect θ for the control variables, which is the same for all companies over all time periods. It seems quite unlikely that this assumption holds. For more labor-intensive industries, labor has likely a larger effect on the sales of a company compared to capital-intensive industries. The same sort of reasoning can also be applied to capital.

Table 7 also shows that for both the Pooled as well as the Double Pooled method, the MSE of Lasso is lower than that of Elastic-Net. This implies that the estimates obtained by Lasso can predict the y-variable more accurately compared to the estimates from Elastic-Net. Furthermore, Elastic-Net selects more spillovers than Lasso, as the percentage of non-zero values is higher for all of the Elastic-Net models. However, given that the MSE of Elastic-Net is higher than that of Lasso this implies that Elastic-Net likely select more "false" spillovers, thus values for γ_{ij} that should actually equal zero, compared to accurate spillovers.

The values of θ_L and θ_C differs quite a lot between different methods, which make it hard to determine which value is the most accurate. The values of θ_C and θ_L are all positive and thus it seems that there is a positive relationship between labor and sales and capital and sales. This is not unexpected, as companies with more capital and more employees are generally bigger and thus should also produce more sales. Furthermore, $\widehat{\theta}_L > \widehat{\theta}_C$ for all methods and that thus implies that $\varepsilon_{L_i}^{Y_i} > \varepsilon_{C_i}^{Y_i}$ for these companies. Thus gaining more employees will lead to a larger increase in sales compared to gaining more capital. However it should be taken into account that for 21 years of data the estimates of θ can be quite inaccurate as was seen in Section 4. Nevertheless, because the same pattern is seen across different methods there is some explanatory power.

Thus, based on the results in Table 7 it seems like the Double Pooled Lasso method outperforms the Double Pooled Elastic-Net method for all methods. However, it should be taken into account that these values are all calculated with $\alpha = 0.5$ and different values of α might perform better. Therefore we re-estimate the three Elastic-Net models for different values of α, from 0.1 to 1 with an interval of 0.1. The results of the static HAC method can be found in Table 32 in Appendix Section D. Here we find lower values for the MSE compared to Lasso for $\alpha = 0.6$ and $\alpha = 0.7$. However, the MSE is only slightly smaller and unlikely to be significant. For the Elastic-Net model which uses CV for computing λ Table 33 can be found in Appendix Section D. In this case Lasso has the lowest MSE for all values of α. We observe that in this case the MSE decreases as α becomes lower, implying that Lasso is truly the best method when using CV for computing λ. The static Elastic-Net model with Elastic-Net weights can be found in Table 34 in Appendix Section D. Here $\alpha = 0.7$ gives a lower MSE compared to Lasso, but this is only because it estimates the value for θ slightly different and not because it estimates more spillover effects. Therefore the Lasso models seem to be preferred, as it is a simpler alternative to Elastic-Net and delivers comparable or even better performance.

5.2 Interpreting spillover effects

It would be beneficial to have a clear understanding of the industries and companies that generate spillover effects, as this could also perhaps help the government assign subsidies to companies who generate the most spillover effects in times of economical difficulties. Furthermore, it is interesting to see if the findings in this paper are in accordance with other literature. Therefore, we will interpret the spillover effect and the private effect of R&D spillovers on the sales of a company. However, based on what was seen in the simulation in section 4 and the high MSE in Table 7 the estimation might not be very precise. It is very well possible that of the non-zero values there are a lot of "falsely" estimated spillover effect, parameters that should actually be zero but are not set to zero. The number of correct spillovers could even potentially be smaller than the false spillover effects. Thus, the interpretation of the spillovers should be taken with caution, as they could be quite inaccurate.

We interpret only the spillovers estimated with the Double Pooled estimator, as they provide the best predictions overall. Furthermore, we use $\alpha = 0.5$ for Elastic-Net for interpreting spillovers. First we focus on interpreting the private effect of R&D investment on the sales of a company. For the Lasso static method we find that the average private effect of a company's own R&D stock on sales is -0.18. This would thus imply that investing in R&D stock would actually lead to less sales, which is counter-intuitive. Because if a higher R&D stock leads to lower sales, it makes no sense for a company to invest in R&D. This also contradicts earlier research of, among others, Jefferson et al. (2006), who find that R&D expenditure leads to higher sales. However, it should be taken into account that this results might not be completely accurate. Furthermore, for the Lasso static method the estimated values of θ_L and θ_C are much larger compared to the other methods. Thus it could be that in the Lasso static method these values are overestimated, and then OLS will set the spillovers to a much smaller value, and apparently even negative in order to minimize the squared residuals.

For Lasso with λ computed with CV the average private effect is -0.01 and for Elastic-Net with CV for λ the average private effect is -0.02. Even though the magnitude of the private effect has decreased, it is still negative which is not in accordance with other literature. So even when the value θ has been reduced in more than half, the private effect of R&D stock on sales remains negative. However, both Lasso and Elastic-Net get a positive value, 0.05 and 0.11 respectively, for the average private effect when the static λ is used together with HAC weights. The model with Elastic-Net weights and a static λ also gets a positive value of 1410.23 for the average private effect, but this number is so large that it is not realistic. Thus, three methods provide positive estimates for the private effect, albeit one is quite unrealistic, and three methods provide negative estimated for the private effect. This makes it difficult to make a final conclusion on the private effect of R&D investment on sales. The positive private effect seems more likely based on economic theory, because companies should only invest in R&D if they can profit off it, and on earlier literature. However, no final conclusion can be made about the direction of the private effect of R&D investment.

Next we try to interpret the spillovers to see whether a certain pattern can be found and whether different methods find similar companies who produce/receive spillovers. To this end we focus on all of the models except Lasso static and Elastic-Net static. We do not focus on Lasso static, because in this model there are no spillovers, that is $\gamma_{ij} = 0,\ \forall i \neq j$. Elastic-Net static is also not focused on as it produces such bad estimates which are highly unlikely and thus interpreting spillovers of this model seems to be pointless. We start with interpreting companies who produce spillovers. To that end we only focus on companies that produce more than 30 spillovers and are present in at least two of the four methods[13]. Most of the spillovers that were estimated were negative, that is that if a companies invests in R&D the sales of another company declines. Only three companies provided on average positive estimates for the spillover effects. Two of these companies were large companies, with values of sales, R&D investment, labor and capital above the mean. It should be noted though that for one of these companies the HAC weights with Elastic-Net provided positive spillovers, while the HAC weights with Lasso provided negative spillovers and the latter effect was stronger. However, this would be in line with earlier research that large companies provide spillover effects (Coe & Helpman, 1995). For all companies that have negative spillovers the values of sales, R&D stock, labor and capital were mostly below the mean. Thus small companies seem to provide negative spillovers.

Next we focus on companies that receive spillovers. Generally spillovers seem to be equally divided and not many companies receive more than 10 spillovers. Hence we focus on companies that receive at least 10 spillover effects and these companies must be found in at least two methods[14]. The two companies who receive positive spillovers are relatively small in terms of labor, sales, capital and R&D stock. This is in line with literature which states that small companies receive the most spillover effects (Coe & Helpman, 1995). The company that received a negative spillover effect had a relatively high sales and high R&D stock, but a low value of capital and labor.

All in all the fact that there were more negative spillovers compared to positive spillovers is in contrast with conclusions found in Bloom et al. (2013). However, the fact that the positive spillovers seemed to come from larger companies is in line with earlier research (Coe & Helpman, 1995). Furthermore, the private effect of the spillovers was for most methods negative, which was also unexpected and not in accordance with earlier research (Jefferson et al., 2006) However, one should bear in mind that the spillover effect that were estimated are likely quite unreliable, because of the small sample size.

6 Conclusion

The main interest in this paper is to find a model which provides the best estimates for measuring R&D spillovers. To answer this question this paper considered a total of 20 models, of which 10 used Elastic-Net and 10 used Lasso methods. The Elastic-Net and Lasso models were calculated with five different types of weights and two different types of penalty parameters. The simulation and the results section in the real life application showed that the Elastic-Net models do no consistently outperform the Lasso methods. Thus, using a more complicated Elastic-Net

[13] There are in total eight companies with CUSIP code 413875, 827079, 806857, 281347, 30710, 949765, 755111, and 858586.

[14] There are in total three companies with CUSIP code 127055, 63934E, and 253651.

approach does not seem to be beneficial for the performance. Additionally, there is no simple best method for estimating the spillovers. Generally, the method which uses Lasso or Elastic-Net for estimating the weights outperform the other methods. Furthermore, using a static value of λ or a λ computed with CV also differs in performance depending on the time horizon and the correlation between variables. Generally, for shorter time horizons, using a static λ might be beneficial, while for larger sample sizes CV is mostly preferred. Therefore it might be a good idea to experiment with simulation depending on the sample size, correlations, and time horizon which method will give the best performance in that specific case.

This research builds on the paper of Manresa (2016) as it uses the method proposed in this paper to estimate the spillover effects. However, this paper considers more expressions for the weights and the penalty parameter compared to the paper of Manresa (2016) and it also applies all these different expressions on Elastic-Net, instead of only Lasso. Furthermore, this paper provides a large scale simulation study which shows when certain methods work well and when certain methods seem to fail. This paper provides some experimental evidence for the models that were proposed in the paper of Manresa (2016). However, this paper simplifies the model which is applied to the data, as it does not consider time effects which were considered in the paper of Manresa (2016).

This paper has some limitations which could be addressed in further research. As mentioned before when applying the model to the data the time effects were ignored in the model, while they might have been present. Therefore further research could focus on finding an appropriate expression for the time effects and put it in the model. Additionally, because the number of years available to estimate the coefficients is quite small, the estimates are not very reliable as demonstrated in section 4. Therefore, the estimated spillovers that are interpreted in this paper, are likely to be inaccurate. Further research should apply this model to a dataset with either a larger time period or with fewer companies. Then they could also interpret the spillovers that were calculated as they become more accurate as the sample size increases.

Another limitation might be that there are some missing control variables in the production function or that the expression for the control variables is not well-defined. Currently, the model includes only labor and capital as control variables. However, in reality there are most likely more control variables which influence sales. Further research could focus on finding more control variables. Furthermore, the expression for the control variables might not be completely accurate. In reality it is quite unlikely that the labor or capital of a company influences all companies in the exact same magnitude. Hence, further research could try to estimate the variable θ for each company separately in order to have a closer resemblance with real life circumstances.

Another limitation is that Elastic-Net is mostly estimated throughout this paper with an $\alpha = 0.5$. However, Elastic-Net might perform a lot better with different values of α. Hence, further research could try to find the optimal value of α in different models with a different DGP. Finally, the last limitation is that the DGP with correlations in x is rather simplistic and unlikely to be found in real life. Only two simple choices were given for generating x with correlations, while in reality the structure of correlations for a variable is likely more complicated. Therefore, future research could focus on different correlation structures.

References

Arcidiacono, P., Foster, G., Goodpaster, N. & Kinsler, J. (2012). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics*, *3*(3), 421–470.

Audretsch, D. B. & Feldman, M. P. (1996). R&d spillovers and the geography of innovation and production. *The American economic review*, *86*(3), 630–640.

Audretsch, D. B. & Vivarelli, M. (1996). Firms size and r&d spillovers: Evidence from italy. *Small Business Economics*, *8*, 249–258.

Bayoumi, T., Coe, D. T. & Helpman, E. (1999). R&d spillovers and global growth. *Journal of International Economics*, *47*(2), 399–428.

Belloni, A., Chen, D., Chernozhukov, V. & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*(6), 2369–2429.

Belloni, A. & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models.

Belloni, A., Chernozhukov, V. & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650.

Bernstein, J. I. & Nadiri, M. I. (1988). *Interindustry r&d spillovers, rates of return, and production in high-tech industries.* National Bureau of Economic Research Cambridge, Mass., USA.

Bloom, N., Schankerman, M. & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347–1393.

Bonaldi, P., Hortaçsu, A. & Kastl, J. (2015). *An empirical analysis of funding costs spillovers in the euro-zone with application to systemic risk* (Tech. Rep.). National Bureau of Economic Research.

Chan, K.-S. & Chen, K. (2011). Subset arma selection via the adaptive lasso. *Statistics and its Interface*, *4*(2), 197–205.

Coe, D. T. & Helpman, E. (1995). International r&d spillovers. *European economic review*, *39*(5), 859–887.

Dimelis, S. P. (2005). Spillovers from foreign direct investment and firm growth: Technological, financial and market structure effects. *International Journal of the Economics of Business*, *12*(1), 85–104.

Douglas, P. H. (1976). The cobb-douglas production function once again: its history, its testing, and some new empirical values. *Journal of political economy*, *84*(5), 903–915.

Görg, H. & Strobl, E. (2001). Multinational companies and productivity spillovers: A meta-analysis. *The economic journal*, *111*(475), 723–739.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Jefferson, G. H., Huamao, B., Xiaojing, G. & Xiaoyun, Y. (2006). R&d performance in chinese industry. *Economics of innovation and new technology*, *15*(4-5), 345–366.

Khan, M. H. R. & Shaw, J. E. H. (2016). Variable selection for survival data with a class of adaptive elastic net techniques. *Statistics and Computing*, *26*(3), 725–741.

Manresa, E. (2016). Estimating the structure of social interactions using panel data.

Newey, W. K. & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

O'Mahony, M. & Vecchi, M. (2009). R&d, knowledge spillovers and company productivity performance. *Research Policy*, *38*(1), 35–44.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Vasylieva, T. A., Lieonov, S. V., Liulov, O. V. & Kyrychenko, K. I. (2018). Macroeconomic stability and its impact on the economic growth of the country.

Yuan, C., Liu, S. & Wu, J. (2009). Research on energy-saving effect of technological progress based on cobb–douglas production function. *Energy Policy*, *37*(8), 2842–2846.

Zellner, A., Kmenta, J. & Dreze, J. (1966). Specification and estimation of cobb-douglas production function models. *Econometrica: Journal of the Econometric Society*, 784–795.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Zou, H. & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics*, *37*(4), 1733.

Appendix A Data cleaning process

Some companies and observations have been removed. Firstly, all observations before the year 1980 and after the year 2001 are removed. Furthermore, observations are removed with missing information in employment, sales, capital stock, and R&D stock. Next, also the companies are removed who have a value of zero for R&D stock for more than 10 years. This is done because if there is not enough variability within the x-variables over a 21-year time period it becomes almost impossible to accurately estimate the model. Therefore removing companies who remain at a zero R&D stock over time will improve the model. Lastly, as for some companies some observations are missing in some years, there might be some issues with missing data for estimating the model. To avoid this issue, any company that does not have complete data for all years is excluded from the dataset. Finally, the final dataset contains 5,786 observations of 263 companies for 22 years of data. However, it should be noted that because in the final model R&D stock is lagged, only 21 years of R&D stock is used and 21 years of the other variables is used. The summary statistics of the adapted dataset can be found in Table 1. Note that this adapted data set contains 22 years of data.

Appendix B The iterative strategy of Belloni et al. (2012)

As mentioned in section 3.1.2 the weights ϕ_{ij}^2 are calculated using HAC type estimator proposed in Newey and West (1986). However, the error term $\widehat{\widehat{u}}_{it}$ is not defined from the data and therefore the iterative strategy from Belloni et al. (2012) is used. In this method, the model is first estimated using the weights:

$$\phi_{ij}^{2(0)} = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2\tilde{y}_{it}^2 + \frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}\tilde{y}_{it}\tilde{y}_{it-1}. \tag{18}$$

Next, using the obtained weights from equation (18), the Pooled Lasso estimator gives estimates for $\widehat{\Gamma}$ and these estimates are used to calculate the residuals: $\widehat{\tilde{u}}_{it} = \widehat{y}_{it} - \sum_{j=1}^{N}\widehat{\gamma}_{ij}^{(0)}\tilde{x}_{jt}$. These residuals can then be plugged into equation (8) to obtain the new weights which can be used to re-estimate the parameter $\widehat{\Gamma}$. This procedure can be iterated to obtain the best estimates for $\widehat{\Gamma}$. In the application in this paper this procedure is iterated until no element in the matrix $\widehat{\Gamma}$ changes with more than 0.001.

Appendix C The Double Pooled Estimator weights

For equation (10) the weights are calculated using the iterative strategy described in appendix B. However, because there are some different parameters the formula's for the weights also change. In this case the initial weights are calculated using the formula:

$$\phi_{ij}^{2(0)} = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2\tilde{w}_{it}^{d2} + \frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}\tilde{w}_{it}^{d}\tilde{w}_{it-1}^{d}.$$

The error terms in the iterative strategy are calculated as $\widehat{\tilde{e}}_{it}^{d} = \widehat{w}_{it}^{d} - \sum_{j=1}^{N}\widehat{\lambda}_{ij}^{d(0)}\tilde{x}_{jt}$. Finally, once the error term is obtained the weights are calculated using the formula

$$\phi_{ij}^{d\,2} = \phi_{j}^{d2} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2\hat{\tilde{e}}_{it}^{d\,2} + \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}\hat{\tilde{e}}_{it}^{d}\hat{\tilde{e}}_{it-1}^{d}\right).$$

For equation (11) the weights are once again estimated using the iterative strategy described in Appendix B. In this case the initial weights are calculated using formula (18). The error terms are then obtained by the formula: $\widehat{\tilde{v}}_{it} = \widehat{y}_{it} - \sum_{j=1}^{N}\widehat{\nu}_{ij}^{(0)}\tilde{x}_{jt}$. Finally, after the first iteration, the weights are then calculated using the formula

$$\phi_{ij}^{2} = \phi_{j}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2\widehat{\tilde{v}}_{it}^{2} + \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}\widehat{\tilde{v}}_{it}\widehat{\tilde{v}}_{it-1}\right).$$

Finally for equation (13) the iterative strategy in Appendix B is used with the initial weights defined as

$$\begin{aligned}\phi_{ij}^{2(0)} &= \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2(\tilde{y}_{it}-\widehat{\theta}\tilde{w}_{it}+(\widehat{\theta}-\theta^0)\tilde{w}_{it})^2+\frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}(\tilde{y}_{it}-\widehat{\theta}\tilde{w}_{it}+(\widehat{\theta}-\theta^0)\tilde{w}_{it})(\tilde{y}_{it-1}-\widehat{\theta}\tilde{w}_{it-1}+(\widehat{\theta}-\theta^0)\tilde{w}_{it-1})\\ &= \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2(\tilde{y}_{it}-\theta^0\tilde{w}_{it})^2 + \frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}(\tilde{y}_{it}-\theta^0\tilde{w}_{it})(\tilde{y}_{it-1}-\theta^0\tilde{w}_{it-1}).\end{aligned}$$

Note that for this equation you need the true value of θ, θ^0, which is only available if one has performed a simulation where this value is known beforehand. But in case this model is applied to the data the original θ^0 is unknown and then the initial weights become simply:

$$\phi_{ij}^{2^{(0)}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2(\tilde{y}_{it} - \widehat{\theta}\tilde{w}_{it})^2 + \frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}(\tilde{y}_{it} - \widehat{\theta}\tilde{w}_{it})(\tilde{y}_{it-1} - \widehat{\theta}\tilde{w}_{it-1}).$$

The error terms are then obtained after Pooled Lasso (Elastic-Net) using the formula: $\widehat{\tilde{u}}_{it} = \widehat{y}_{it} - \widehat{\theta}\tilde{w}_{it} - \sum_{j=1}^{N}\widehat{\gamma}_{ij}^{(0)}\tilde{x}_{jt}$. Finally, new weights can be calculated using the formula:

$$\phi_{ij}^2 = {\phi_j}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2(\widehat{\tilde{u}}_{it} + (\widehat{\theta} - \theta^0)\tilde{w}_{it})^2 + \frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}(\widehat{\tilde{u}}_{it} + (\widehat{\theta} - \theta^0)\tilde{w}_{it})(\widehat{\tilde{u}}_{it-1} + (\widehat{\theta} - \theta^0)\tilde{w}_{it-1})\right).$$

Once again, in the data θ^0 is undefined and then the weights will be simplified to

$$\phi_{ij}^2 = {\phi_j}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{jt}^2\widehat{\tilde{u}}_{it}^2 + \frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{jt}\tilde{x}_{jt-1}\widehat{\tilde{u}}_{it}\widehat{\tilde{u}}_{it-1}\right).$$

Appendix D Tables & Figures

Table 8: The standard deviation of 100 simulations of the performance measures of the models considered for 21 years of data

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	1.27	1.01%	1.10%	0.04%	47.50
Var(x) static	6.30	1.02%	1.03%	0.20%	719.96
HAC static	3.67	0.95%	1.00%	0.09%	344.89
Lasso static	1.28	1.01%	1.10%	0.03%	47.01
Ridge static	333.36	2.03%	2.26%	0.09%	660,763.10
EW CV	634.68	2.24%	2.47%	0.18%	895,499.40
Var(x) CV	2,738.55	1.64%	1.78%	0.29%	6,201,708.00
HAC CV	476.15	1.89%	2.09%	0.20%	135,143.80
Lasso CV	6.30	1.92%	2.09%	0.11%	113.10
Ridge CV	2,062.05	2.02%	2.21%	0.12%	450,589.20
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	107.76	2.02%	2.30%	0.10%	27,835.73
Var(x) static	1,406.40	1.53%	1.66%	0.15%	1,349,977.00
HAC static	784.09	1.83%	2.06%	0.10%	660,683.10
Enet static	107.77	2.02%	2.30%	0.10%	27,835.84
Ridge static	116,435.37	2.09%	2.31%	0.06%	2,200,229,000.00
EW CV	4,772.95	1.97%	2.14%	0.16%	12,655,310.00
Var(x) CV	9,304.73	1.53%	1.69%	0.35%	203,749,200.00
HAC CV	7,866.34	1.83%	2.02%	0.20%	86,041,530.00
Enet CV	95.75	1.88%	2.07%	0.12%	2,621.86
Ridge CV	13,987.26	2.08%	2.28%	0.09%	525,959,600.00

Notes: This table shows the standard deviation of the performance measures of the methods considered for 100 simulations over 21 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall(i,j) \in 1,...,N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 9: The standard deviation of 50 simulations of the performance measures of the models considered for 100 years of data

Method	F-norm	% non-zero	% spillovers	% zero	MSE
Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	1.15	2.72%	3.17%	0.01%	21.10
Var(x) static	1.19	2.48%	2.88%	0.15%	33.60
HAC static	1.16	2.67%	3.10%	0.03%	23.41
Lasso static	1.17	2.55%	2.96%	0.01%	20.73
Ridge static	45.55	0.38%	0.45%	0.43%	736.52
EW CV	0.33	0.07%	0.08%	0.44%	6.47
Var(x) CV	0.59	0.06%	0.07%	0.52%	8.96
HAC CV	0.42	0.06%	0.07%	0.43%	7.14
Lasso CV	0.27	0.27%	0.32%	0.02%	2.32
Ridge CV	0.70	0.66%	0.78%	0.23%	8.71
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.93	1.02%	1.20%	0.10%	7.64
Var(x) static	1.12	1.13%	1.33%	0.23%	11.00
HAC static	0.91	0.95%	1.12%	0.11%	8.47
Enet static	0.72	0.69%	0.81%	0.10%	5.67
Ridge static	1,462.72	0.33%	0.39%	0.40%	3,859,044.00
EW CV	1.14	0.08%	0.10%	0.48%	8.78
Var(x) CV	2.88	0.07%	0.08%	0.48%	28.15
HAC CV	2.12	0.06%	0.07%	0.46%	25.82
Enet CV	0.30	0.28%	0.33%	0.03%	2.21
Ridge CV	0.71	0.66%	0.78%	0.25%	8.71

Notes: This table shows the standard deviation of the performance measures of the methods considered for 50 simulations over 100 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 10: The mean of the performance measures using 50 simulations over 200 years of data

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	9.60	91.51%	89.90%	100.00%	46.73
Var(x) static	9.12	92.46%	91.03%	99.77%	42.55
HAC static	8.70	92.55%	91.14%	99.98%	40.20
Lasso static	7.73	94.21%	93.11%	100.00%	35.06
Ridge static	19.44	99.95%	99.94%	86.33%	114.75
EW CV	10.78	100.00%	99.99%	91.32%	35.33
Var(x) CV	11.75	100.00%	99.99%	90.12%	37.64
HAC CV	11.25	100.00%	100.00%	90.97%	36.13
Lasso CV	3.28	99.72%	99.66%	99.99%	14.58
Ridge CV	6.03	99.38%	99.26%	98.92%	19.75
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	3.80	99.75%	99.70%	99.84%	15.23
Var(x) static	5.35	99.86%	99.83%	99.08%	17.35
HAC static	4.31	99.86%	99.83%	99.68%	15.69
Enet static	3.69	99.86%	99.84%	99.84%	14.88
Ridge static	38.92	99.95%	99.94%	74.91%	2,625.22
EW CV	12.24	100.00%	99.99%	87.11%	41.80
Var(x) CV	13.19	100.00%	100.00%	85.81%	45.21
HAC CV	12.67	100.00%	100.00%	86.75%	42.67
Enet CV	3.29	99.75%	99.70%	99.99%	14.56
Ridge CV	6.22	99.37%	99.25%	98.73%	20.27

Notes: This table shows the mean of 50 simulations of the performance measures of the methods considered for 200 years of data of 263 companies. The x-variables are i.i.d. generated variables. F-norm represents the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 11: The standard deviation of the performance measures using 50 simulations over 200 years of data

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.95	1.42%	1.69%	0.00%	10.85
Var(x) static	0.84	1.18%	1.41%	0.09%	7.88
HAC static	0.92	1.34%	1.59%	0.01%	8.62
Lasso static	0.71	0.91%	1.08%	0.00%	7.97
Ridge static	10.72	0.06%	0.07%	0.74%	176.84
EW CV	0.19	0.02%	0.02%	0.38%	3.72
Var(x) CV	0.33	0.02%	0.02%	0.40%	3.99
HAC CV	0.24	0.01%	0.02%	0.39%	3.71
Lasso CV	0.14	0.13%	0.16%	0.00%	1.88
Ridge CV	0.33	0.26%	0.31%	0.10%	2.57
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.18	0.13%	0.15%	0.03%	2.07
Var(x) static	0.43	0.10%	0.12%	0.21%	2.35
HAC static	0.24	0.10%	0.12%	0.06%	1.99
Enet static	0.13	0.08%	0.10%	0.03%	1.99
Ridge static	85.16	0.08%	0.10%	0.78%	16,675.79
EW CV	0.18	0.02%	0.02%	0.38%	3.97
Var(x) CV	0.33	0.01%	0.02%	0.45%	4.40
HAC CV	0.24	0.01%	0.01%	0.39%	4.18
Enet CV	0.14	0.13%	0.15%	0.01%	1.90
Ridge CV	0.34	0.26%	0.31%	0.11%	2.83

Notes: This table shows the standard deviation of 50 simulations of the performance measures of the methods considered for 200 years of data of 263 companies. The x-variables are i.i.d. generated variables. F-norm represents the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall(i,j) \in 1,...,N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 12: The mean of 100 simulations of the performance measures of the models considered for 21 years of data with unequal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	30.88	67.43%	61.23%	99.94%	349.74
Var(x) static	74.02	47.07%	36.97%	97.98%	5,823.09
HAC static	46.92	59.37%	51.62%	99.33%	1,638.75
Lasso static	30.88	67.43%	61.23%	99.94%	349.74
Ridge static	164.39	93.15%	91.85%	97.03%	23,601.58
EW CV	221.45	92.41%	90.97%	96.40%	90,841.12
Var(x) CV	510.67	60.34%	52.78%	95.49%	321,283.93
HAC CV	456.24	83.41%	80.25%	95.78%	239,099.53
Lasso CV	29.89	76.96%	72.57%	99.48%	204.39
Ridge CV	58.06	93.76%	92.58%	97.57%	6,669.21
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	25.50	93.67%	92.47%	98.98%	124.74
Var(x) static	1,355.08	68.51%	62.50%	96.00%	528,806.70
HAC static	110.41	89.61%	87.63%	97.80%	3,796.93
Enet static	25.50	93.67%	92.47%	98.98%	124.75
Ridge static	542.22	93.49%	92.25%	95.54%	179,262.20
EW CV	1,997.05	91.84%	90.29%	95.16%	2,295,043.00
Var(x) CV	5,021.30	66.91%	60.61%	94.51%	45,332,950.00
HAC CV	8,456.08	85.44%	82.67%	94.70%	112,743,600.00
Enet CV	27.79	83.69%	80.59%	99.30%	164.41
Ridge CV	191.69	95.58%	94.73%	97.01%	113,293.50

Notes: This table shows the mean of the performance measures of the methods considered for 100 simulations over 21 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between companies with spillovers and 0.1 between companies without spillovers. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1,...,N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 13: The standard deviation of 100 simulations of the performance measures of the models considered for 21 years of data with unequal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	1.50	2.58%	3.06%	0.02%	78.65
Var(x) static	11.39	2.89%	3.45%	0.29%	2,894.58
HAC static	5.07	2.74%	3.26%	0.15%	807.73
Lasso static	1.50	2.58%	3.06%	0.02%	78.65
Ridge static	504.01	1.31%	1.54%	0.09%	169,461.10
EW CV	924.15	1.30%	1.53%	0.13%	850,457.50
Var(x) CV	1,023.78	5.13%	6.07%	0.18%	2,057,395.00
HAC CV	1,081.16	5.69%	6.75%	0.14%	1,376,621.00
Lasso CV	1.42	1.72%	1.96%	0.05%	57.73
Ridge CV	250.14	1.22%	1.44%	0.14%	64,224.15
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	4.57	1.42%	1.68%	0.14%	36.53
Var(x) static	4,007.79	3.78%	4.49%	0.30%	1,849,943.00
HAC static	223.55	2.14%	2.53%	0.22%	14,492.22
Enet static	4.57	1.42%	1.68%	0.14%	36.55
Ridge static	828.18	1.19%	1.40%	0.07%	956,687.90
EW CV	5,384.32	1.27%	1.50%	0.10%	9,160,996.00
Var(x) CV	13,923.85	4.48%	5.30%	0.17%	283,733,600.00
HAC CV	48,061.14	3.92%	4.65%	0.11%	765,468,100.00
Enet CV	1.28	1.47%	1.69%	0.07%	42.43
Ridge CV	1,012.00	1.02%	1.20%	0.15%	1,096,810.00

Notes: This table shows the standard deviation of the performance measures of the methods considered for 100 simulations over 21 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between companies with spillovers and 0.1 between companies without spillovers. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 14: The mean of 50 simulations of the performance measures of the models considered for 100 years of data with unequal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	6.12	99.44%	99.33%	100.00%	16.84
Var(x) static	7.52	98.88%	98.67%	99.70%	39.39
HAC static	6.00	99.68%	99.62%	99.99%	17.13
Lasso static	6.12	99.44%	99.33%	100.00%	16.84
Ridge static	148.09	99.87%	99.85%	91.29%	15,135.72
EW CV	11.43	99.99%	99.99%	96.20%	31.23
Var(x) CV	13.73	100.00%	99.99%	95.00%	41.87
HAC CV	12.38	100.00%	100.00%	95.88%	35.64
Lasso CV	6.94	98.04%	97.67%	100.00%	19.34
Ridge CV	7.17	100.00%	100.00%	99.66%	18.88
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	6.22	99.99%	99.99%	99.91%	16.82
Var(x) static	8.35	100.00%	100.00%	99.11%	49.37
HAC static	7.05	100.00%	100.00%	99.74%	27.85
Enet static	6.22	99.99%	99.99%	99.91%	16.82
Ridge static	252.67	99.79%	99.74%	84.18%	43,428.86
EW CV	12.38	100.00%	100.00%	94.75%	33.65
Var(x) CV	14.69	100.00%	100.00%	93.51%	45.33
HAC CV	13.37	100.00%	100.00%	94.37%	38.16
Enet CV	6.49	98.75%	98.51%	99.99%	18.01
Ridge CV	7.29	100.00%	100.00%	99.61%	19.20

Notes: This table shows the mean of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between companies with spillovers and 0.1 between companies without spillovers. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall(i,j) \in 1,...,N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 15: The standard deviation of 50 simulations of the performance measures of the models considered for 100 years of data with unequal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.29	0.24%	0.29%	0.00%	1.92
Var(x) static	0.67	0.33%	0.39%	0.17%	13.07
HAC static	0.24	0.13%	0.15%	0.01%	2.46
Lasso static	0.29	0.24%	0.29%	0.00%	1.92
Ridge static	381.57	0.17%	0.21%	0.36%	49,230.78
EW CV	0.22	0.02%	0.02%	0.23%	3.01
Var(x) CV	0.78	0.02%	0.02%	0.34%	5.73
HAC CV	0.49	0.01%	0.01%	0.24%	3.82
Lasso CV	0.37	0.38%	0.45%	0.00%	2.59
Ridge CV	0.23	0.01%	0.02%	0.03%	2.03
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.22	0.02%	0.02%	0.02%	1.80
Var(x) static	0.74	0.02%	0.02%	0.28%	14.08
HAC static	0.42	0.00%	0.00%	0.08%	7.36
Enet static	0.22	0.02%	0.02%	0.02%	1.80
Ridge static	513.94	0.22%	0.27%	0.34%	161,495.80
EW CV	0.24	0.01%	0.01%	0.25%	3.88
Var(x) CV	0.79	0.01%	0.01%	0.39%	6.77
HAC CV	0.45	0.01%	0.01%	0.29%	5.05
Enet CV	0.35	0.30%	0.36%	0.00%	2.27
Ridge CV	0.23	0.01%	0.01%	0.03%	2.12

Notes: This table shows the standard deviation of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between companies with spillovers and 0.1 between companies without spillovers. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall(i,j) \in 1,...,N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 16: The mean of 100 simulations of the performance measures of the models considered for 21 years of data with equal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	55.17	30.72%	17.41%	97.63%	1,568.95
Var(x) static	86.68	22.07%	7.10%	96.80%	31,923.74
HAC static	87.21	23.54%	8.85%	96.84%	25,794.61
Lasso static	55.17	30.71%	17.41%	97.63%	1,569.10
Ridge static	953.45	37.87%	25.94%	95.04%	3,941,073.94
EW CV	1,109.02	47.56%	37.50%	94.25%	590,896.93
Var(x) CV	517.10	30.28%	16.90%	94.42%	139,442.39
HAC CV	1,209.28	34.28%	21.66%	94.16%	1,365,209.96
Lasso CV	54.43	35.22%	22.78%	98.00%	728.66
Ridge CV	328.62	42.34%	31.27%	94.80%	48,876.55
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	11,377.82	41.94%	30.80%	93.71%	346,007,900.00
Var(x) static	1,903.64	28.50%	14.77%	94.42%	1,949,251.00
HAC static	21,793.30	32.48%	19.51%	93.93%	5,845,696,000.00
Enet static	12,520.03	41.95%	30.81%	93.71%	352,364,700.00
Ridge static	2,015.40	40.42%	28.98%	93.72%	2,081,421.00
EW CV	3,823.18	44.84%	34.25%	93.40%	6,643,204.00
Var(x) CV	4,421.20	32.50%	19.54%	93.49%	22,761,340.00
HAC CV	20,228.89	36.16%	23.91%	93.32%	105,491,600.00
Enet CV	54.36	38.09%	26.20%	97.22%	632.09
Ridge CV	3,032.76	43.74%	32.94%	94.03%	15,303,070.00

Notes: This table shows the mean of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between all companies. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i, j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 17: The standard deviation of 100 simulations of the performance measures of the models considered for 21 years of data with equal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	1.89	2.28%	2.60%	0.34%	293.78
Var(x) static	15.69	1.11%	1.26%	0.27%	36,680.43
HAC static	14.50	1.33%	1.55%	0.36%	29,434.80
Lasso static	1.89	2.28%	2.60%	0.34%	293.79
Ridge static	5,737.52	2.02%	2.25%	0.09%	37,453,270.00
EW CV	3,377.50	2.36%	2.60%	0.10%	3,205,108.00
Var(x) CV	905.92	2.42%	2.80%	0.37%	584,003.60
HAC CV	3,068.59	3.00%	3.43%	0.22%	8,077,505.00
Lasso CV	1.95	1.94%	2.05%	0.14%	181.16
Ridge CV	523.96	2.14%	2.37%	0.11%	252,864.80

Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	64,195.14	2.23%	2.46%	0.28%	3,278,305,000.00
Var(x) static	3,730.03	1.60%	1.86%	0.31%	7,610,477.00
HAC static	156,089.10	2.02%	2.29%	0.30%	58,108,140,000.00
Enet static	64,689.68	2.23%	2.45%	0.28%	3,278,186,000.00
Ridge static	5,081.53	2.09%	2.32%	0.08%	8,817,070.00
EW CV	6,546.15	2.38%	2.63%	0.06%	43,251,310.00
Var(x) CV	10,557.79	2.31%	2.64%	0.26%	143,000,300.00
HAC CV	122,798.70	2.87%	3.25%	0.09%	646,857,000.00
Enet CV	1.64	1.97%	2.16%	0.12%	143.38
Ridge CV	9,522.36	2.31%	2.57%	0.10%	123,449,200.00

Notes: This table shows the standard deviation of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between all companies. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 18: The mean of 50 simulations of the performance measures of the models considered for 100 years of data with equal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	13.04	98.23%	97.89%	96.70%	47.10
Var(x) static	25.25	79.33%	75.40%	94.68%	570.20
HAC static	17.94	92.10%	90.60%	95.22%	183.68
Lasso static	13.04	98.23%	97.89%	96.70%	47.09
Ridge static	68.27	96.49%	95.83%	91.53%	708.45
EW CV	17.63	99.81%	99.77%	91.20%	54.17
Var(x) CV	18.87	99.65%	99.58%	91.29%	109.49
HAC CV	18.58	99.81%	99.78%	91.01%	61.55
Lasso CV	9.49	95.11%	94.18%	99.95%	27.34
Ridge CV	14.47	98.16%	97.81%	96.33%	51.09
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	15.11	99.40%	99.29%	90.73%	70.00
Var(x) static	18.18	96.62%	95.98%	91.24%	240.09
HAC static	16.65	99.08%	98.90%	90.82%	199.45
Enet static	15.11	99.40%	99.28%	90.74%	69.97
Ridge static	227.54	97.27%	96.75%	85.10%	65,831.57
EW CV	20.25	99.79%	99.75%	88.63%	63.39
Var(x) CV	20.49	99.61%	99.54%	89.33%	129.19
HAC CV	21.02	99.80%	99.76%	88.59%	74.06
Enet CV	9.40	95.68%	94.86%	99.92%	26.46
Ridge CV	14.76	98.15%	97.80%	95.55%	51.08

Notes: This table shows the mean of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between all companies. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1, ..., N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 19: The standard deviation of 50 simulations of the performance measures of the models considered for 100 years of data with equal correlations

Panel A: Lasso methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.48	0.69%	0.83%	0.33%	7.81
Var(x) static	2.17	3.56%	4.27%	0.42%	251.40
HAC static	1.91	2.89%	3.43%	0.49%	47.19
Lasso static	0.48	0.69%	0.83%	0.33%	7.80
Ridge static	76.46	0.77%	0.90%	0.40%	1,277.38
EW CV	0.44	0.12%	0.15%	0.27%	7.77
Var(x) CV	1.05	0.19%	0.22%	0.41%	30.08
HAC CV	0.57	0.11%	0.13%	0.24%	9.17
Lasso CV	0.58	0.76%	0.88%	0.02%	4.07
Ridge CV	0.52	0.45%	0.53%	0.21%	7.61
Panel B: Elastic-Net methods					
Method	**F-norm**	**% non-zero**	**% spillovers**	**% zero**	**MSE**
EW static	0.45	0.26%	0.31%	0.40%	14.43
Var(x) static	0.85	0.93%	1.10%	0.43%	74.77
HAC static	0.61	0.41%	0.49%	0.41%	76.49
Enet static	0.45	0.26%	0.31%	0.40%	14.43
Ridge static	530.27	0.69%	0.81%	0.46%	433,087.90
EW CV	0.47	0.12%	0.14%	0.31%	9.27
Var(x) CV	1.22	0.23%	0.28%	0.49%	39.27
HAC CV	0.68	0.13%	0.16%	0.35%	12.70
Enet CV	0.54	0.67%	0.78%	0.03%	3.81
Ridge CV	0.51	0.45%	0.54%	0.22%	7.53

Notes: This table shows the standard deviation of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated, where the correlation is 0.5 between all companies. F-norm refers to the Frobenius norm and % non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. % spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. % zero gives the percentage of variables γ_{ij} which are correctly set to zero. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. EW means that weights $\phi_{ij} = 1$ is used $\forall (i,j) \in 1,...,N$. Var(x) implies that equation (7) is used for the weights, and HAC implies that equation (8) are used for the weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights, for Ridge the weights $\phi_{ij} = (\widehat{\gamma}(ridge)_{ij})^{-1}$ are used and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 20: The standard deviation of 100 simulations of the performance measures of the Double Pooled methods for 21 years of data

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	4.91	0.78%	0.89%	0.13%	368	0.20	0.09
HAC sc static	4.91	0.79%	0.89%	0.13%	366	0.20	0.09
Lasso static	1.35	0.95%	1.09%	0.04%	55	0.14	0.07
Lasso CV	105.58	1.82%	2.01%	0.11%	25,525	0.36	0.16
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	33,206.28	1.76%	2.02%	0.15%	4,751,095,000	0.42	0.17
HAC sc static	33,206.28	1.76%	2.02%	0.15%	4,751,095,000	0.42	0.17
Enet static	22.68	1.91%	2.19%	0.15%	521	0.41	0.17
Enet CV	279.05	1.83%	2.05%	0.14%	36,721	0.39	0.16

Notes: This table shows the standard deviation of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 21: The standard deviation of 50 simulations of the performance measures of the Double Pooled methods over 100 years of data

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	1.27	2.59%	3.05%	0.04%	28.12	0.05	0.03
HAC sc static	1.27	2.59%	3.05%	0.04%	28.15	0.05	0.03
Lasso static	1.12	2.35%	2.77%	0.01%	22.26	0.04	0.02
Lasso CV	0.36	0.26%	0.31%	0.03%	2.55	0.07	0.04
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	1.03	0.95%	1.12%	0.14%	9.30	0.10	0.06
HAC sc static	1.03	0.95%	1.12%	0.14%	9.30	0.10	0.06
Enet static	0.80	0.72%	0.85%	0.09%	7.20	0.10	0.06
Enet CV	0.38	0.27%	0.32%	0.04%	2.76	0.08	0.05

Notes: This table shows the standard deviation of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are i.i.d. generated variables. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 22: The mean of 100 simulations of the performance measures of the Double Pooled methods for 21 years of data with unequal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	47.91	58.90%	51.10%	99.30%	1,871	0.51	0.22
HAC sc static	47.93	58.90%	51.10%	99.30%	1,877	0.51	0.22
Lasso static	31.30	67.42%	61.23%	99.93%	357	0.38	0.16
Lasso CV	32.97	74.00%	69.07%	99.34%	242	0.76	0.34
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	131.44	88.41%	86.21%	97.64%	13,973	0.83	0.37
HAC sc static	131.44	88.41%	86.21%	97.64%	13,973	0.83	0.37
Enet static	28.78	92.95%	91.61%	98.83%	153	0.71	0.31
Enet CV	31.25	81.04%	77.45%	99.11%	199	0.80	0.35

Notes: This table shows the mean of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 with companies with spillovers, and a correlation of 0.1 for companies without spillovers. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 23: The standard deviations of the performance measures of the Double Pooled methods using 100 simulations for 21 years of data with unequal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	6.11	2.75%	3.25%	0.17%	1,236	0.25	0.10
HAC sc static	6.12	2.74%	3.25%	0.17%	1,235	0.25	0.10
Enet static	1.64	2.70%	3.23%	0.03%	75	0.18	0.07
Enet CV	2.24	2.32%	2.69%	0.11%	61	0.38	0.14
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	244.71	2.29%	2.72%	0.24%	60,352	0.42	0.15
HAC sc static	244.71	2.29%	2.72%	0.24%	60,352	0.42	0.15
Lasso static	3.51	1.67%	1.99%	0.18%	42	0.38	0.13
Lasso CV	2.44	2.12%	2.48%	0.16%	50	0.41	0.14

Notes: This table shows the standard deviation of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 with companies with spillovers, and a correlation of 0.1 for companies without spillovers. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 24: The mean of 50 simulations of the performance measures of the Double Pooled methods over 100 years of data with unequal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	6.03	99.67%	99.61%	99.99%	17.51	0.06	0.03
HAC sc static	6.03	99.67%	99.61%	99.99%	17.52	0.06	0.03
Lasso static	6.12	99.45%	99.34%	100.00%	17.47	0.06	0.03
Lasso CV	7.00	97.95%	97.56%	100.00%	19.80	0.07	0.03
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	7.19	100.00%	99.99%	99.72%	29.52	0.16	0.07
HAC sc static	7.19	100.00%	99.99%	99.72%	29.52	0.16	0.07
Enet static	6.34	99.98%	99.98%	99.90%	17.59	0.15	0.07
Enet CV	6.51	98.72%	98.47%	99.99%	18.47	0.07	0.03

Notes: This table shows the mean of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 with companies with spillovers, and a correlation of 0.1 for companies without spillovers. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 25: The standard deviation of 50 simulations of the performance measures of the Double Pooled methods over 100 years of data with unequal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	0.26	0.15%	0.18%	0.01%	1.83	0.03	0.01
HAC sc static	0.26	0.15%	0.18%	0.01%	1.84	0.03	0.01
Lasso static	0.29	0.25%	0.29%	0.00%	2.03	0.03	0.01
Lasso CV	0.37	0.37%	0.44%	0.00%	2.34	0.04	0.02
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	0.39	0.02%	0.02%	0.08%	7.65	0.10	0.04
HAC sc static	0.39	0.02%	0.02%	0.08%	7.65	0.10	0.04
Enet static	0.26	0.04%	0.05%	0.03%	1.79	0.11	0.05
Enet CV	0.33	0.33%	0.39%	0.00%	2.12	0.04	0.02

Notes: This table shows the standard deviation of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 with companies with spillovers, and a correlation of 0.1 for companies without spillovers. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 26: The mean of 100 simulations of the performance measures of the Double Pooled methods for 21 years of data with equal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	85.66	23.76%	9.18%	96.92%	21,531	0.66	0.37
HAC sc static	88.08	23.77%	9.19%	96.91%	21,596	0.66	0.37
Lasso static	56.14	30.23%	16.88%	97.71%	1,476	0.65	0.37
Lasso CV	56.45	33.78%	21.12%	97.85%	662	0.84	0.47
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	3,186.07	32.47%	19.56%	94.00%	3,640,095	0.90	0.50
HAC sc static	3,186.07	32.47%	19.56%	94.00%	3,640,095	0.90	0.50
Enet static	18,429.89	41.21%	29.97%	93.81%	1,137,144,000	0.91	0.51
Enet CV	56.25	36.55%	24.42%	97.10%	583	0.87	0.49

Notes: This table shows the mean of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 between all companies. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 27: The standard deviation of 100 simulations of the performance measures of the Double Pooled methods for 21 years of data with equal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	8.64	1.48%	1.73%	0.42%	9,878	0.31	0.13
HAC sc static	24.63	1.48%	1.74%	0.42%	10,001	0.31	0.13
Lasso static	1.88	2.55%	2.96%	0.41%	271	0.31	0.13
Lasso CV	2.11	2.24%	2.41%	0.18%	154	0.38	0.17
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	7,440.64	1.91%	2.16%	0.34%	16,521,610	0.41	0.18
HAC sc static	7,440.64	1.91%	2.16%	0.34%	16,521,610	0.41	0.18
Enet static	90,196.77	2.24%	2.48%	0.35%	10,507,590,000	0.42	0.18
Enet CV	1.89	2.39%	2.63%	0.17%	121	0.40	0.17

Notes: This table shows the standard deviation of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 between all companies. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 28: The mean of 100 simulations of the performance measures of the Double Pooled methods for 21 years of data with equal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	85.66	23.76%	9.18%	96.92%	21,531	0.66	0.37
HAC sc static	88.08	23.77%	9.19%	96.91%	21,596	0.66	0.37
Lasso static	56.14	30.23%	16.88%	97.71%	1,476	0.65	0.37
Lasso CV	56.45	33.78%	21.12%	97.85%	662	0.84	0.47
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	3,186.07	32.47%	19.56%	94.00%	3,640,095	0.90	0.50
HAC sc static	3,186.07	32.47%	19.56%	94.00%	3,640,095	0.90	0.50
Enet static	18,429.89	41.21%	29.97%	93.81%	1,137,144,000	0.91	0.51
Enet CV	56.25	36.55%	24.42%	97.10%	583	0.87	0.49

Notes: This table shows the mean of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 between all companies. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 29: The standard deviation of 100 simulations of the performance measures of the Double Pooled methods for 21 years of data with equal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	8.64	1.48%	1.73%	0.42%	9,878	0.31	0.13
HAC sc static	24.63	1.48%	1.74%	0.42%	10,001	0.31	0.13
Lasso static	1.88	2.55%	2.96%	0.41%	271	0.31	0.13
Lasso CV	2.11	2.24%	2.41%	0.18%	154	0.38	0.17
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	7,440.64	1.91%	2.16%	0.34%	16,521,610	0.41	0.18
HAC sc static	7,440.64	1.91%	2.16%	0.34%	16,521,610	0.41	0.18
Enet static	90,196.77	2.24%	2.48%	0.35%	10,507,590,000	0.42	0.18
Enet CV	1.89	2.39%	2.63%	0.17%	121	0.40	0.17

Notes: This table shows the standard deviation of 100 simulations of the performance measures of the methods considered over 21 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 between all companies. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 30: The mean of 50 simulations of the performance measures of the Double Pooled methods over 100 years of data with equal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	18.28	91.65%	90.04%	95.29%	201.65	0.22	0.12
HAC sc static	18.30	91.60%	89.99%	95.29%	201.75	0.22	0.12
Lasso static	13.20	98.04%	97.66%	96.76%	47.66	0.21	0.11
Lasso CV	9.49	95.13%	94.20%	99.95%	27.13	0.10	0.05
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	17.00	98.96%	98.76%	90.96%	228.05	0.26	0.14
HAC sc static	17.00	98.96%	98.76%	90.96%	228.05	0.26	0.14
Enet static	15.31	99.30%	99.17%	90.88%	70.50	0.25	0.13
Enet CV	9.43	95.64%	94.81%	99.91%	26.57	0.10	0.05

Notes: This table shows the mean of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 between all companies. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 31: The standard deviation of 50 simulations of the performance measures of the Double Pooled methods over 100 years of data with equal correlations

Panel A: Lasso methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	1.85	2.70%	3.23%	0.43%	54.79	0.12	0.05
HAC sc static	1.86	2.71%	3.24%	0.43%	53.99	0.12	0.05
Lasso static	0.53	0.69%	0.82%	0.28%	7.70	0.12	0.05
Lasso CV	0.50	0.76%	0.91%	0.02%	3.30	0.06	0.03
Panel B: Elastic-Net methods							
Method	**F-norm**	**%non-zero**	**%spillovers**	**%zero**	**MSE**	$\lvert\theta_1^0 - \widehat{\theta}_1\rvert$	$\lvert\theta_2^0 - \widehat{\theta}_2\rvert$
HAC static	0.74	0.38%	0.45%	0.39%	89.54	0.14	0.06
HAC sc static	0.74	0.38%	0.45%	0.39%	89.54	0.14	0.06
Enet static	0.46	0.29%	0.35%	0.39%	11.20	0.14	0.06
Enet CV	0.47	0.64%	0.76%	0.02%	3.12	0.07	0.03

Notes: This table shows the standard deviation of 50 simulations of the performance measures of the methods considered over 100 years of data for 263 companies. The x-variables are correlated with a correlation of 0.5 between all companies. F-norm represents the Frobenius norm and %non-zero represents the percentage of variables γ_{ij} which are correctly given a non-zero value. %spillovers gives the percentage of variables $\gamma_{ij}, i \neq j$ which are correctly estimated to be non-zero. %zero gives the percentage of variables γ_{ij} which are correctly set to zero. $\widehat{\theta}$ references to the estimated value of θ, while θ^0 references to the actual value of θ. For the abbreviations of the methods, static implies that the λ of equation (6) is used, while CV implies that λ is calculated with cross-validation. HAC implies that equation (8) are used for the weights and θ^0 is not used to estimate the final weights. HAC sc implies that equation (8) are used for the weights and θ^0 is used to estimate the final weights. Lasso implies that $\phi_{ij} = (\widehat{\gamma}(lasso)_{ij})^{-1}$ are used for weights and for Enet the weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ are used.

Table 32: The results of the Double Pooled Elastic-Net model with the weights and λ of Manresa (2016) for different values of α

α	**MSE**	**% non-zero**	$\widehat{\theta}_L$	$\widehat{\theta}_C$
0.1	30,643,288,065.64	7.22%	0.08	0.12
0.2	6,228,329.16	5.36%	0.24	0.13
0.3	27,595.61	3.39%	0.33	0.21
0.4	18,204.80	2.27%	0.39	0.26
0.5	10,960.29	1.68%	0.45	0.27
0.6	8,840.34	1.31%	0.50	0.33
0.7	8,673.80	1.10%	0.53	0.35
0.8	9,387.61	0.92%	0.60	0.40
0.9	9,978.32	0.83%	0.64	0.41
1.0	8,858.45	0.73%	0.68	0.45

Notes: This table shows the MSE, the percentage of non-zero values within the matrix Γ, and the estimated values of θ_L and θ_C for the Double Pooled Elastic-Net model with weights of equation (8) and λ according to equation (6). It shows these results for different values of α and it should be noted that $\alpha = 1$ is equal to the Lasso estimator.

Table 33: The results of the Double Pooled Elastic-Net model with Elastic-Net weights and λ computed with cross-validation

α	**MSE**	**% non-zero**	$\widehat{\theta}_L$	$\widehat{\theta}_C$
0.1	17,770.01	5.46%	0.30	0.07
0.2	1,859.01	3.83%	0.41	0.13
0.3	1,426.05	3.15%	0.45	0.15
0.4	1,262.61	2.80%	0.48	0.17
0.5	1,147.77	2.54%	0.46	0.16
0.6	1,148.06	2.36%	0.46	0.16
0.7	1,026.40	2.22%	0.46	0.16
0.8	1,007.20	2.12%	0.44	0.15
0.9	951.52	2.06%	0.44	0.15
1	864.80	2.00%	0.43	0.15

Notes: This table shows the MSE, the percentage of non-zero values within the matrix Γ, and the estimated values of θ_L and θ_C for the Double Pooled Elastic-Net model with weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ and λ computed with cross-validation. It shows these results for different values of α and it should be noted that $\alpha = 1$ is equal to the Lasso estimator.

Table 34: The results of the Double Pooled Elastic-Net model with Elastic-Net weights and a static λ

α	**MSE**	**% non-zero**	$\widehat{\theta}_L$	$\widehat{\theta}_C$
0.1	887,500.74	5.75%	0.05	0.05
0.2	641,638.02	2.94%	0.24	-0.07
0.3	1,546,167.99	1.65%	0.00	-0.01
0.4	351,091,236.78	2.46%	4.72	-0.08
0.5	753,046,025,756.13	7.60%	3155.95	0.05
0.6	229,693,958.62	5.58%	-28.55	-7.47
0.7	0.05	0.38%	1.07	0.63
0.8	0.06	0.38%	1.11	0.65
0.9	0.06	0.38%	1.11	0.65
1	0.06	0.38%	1.11	0.65

Notes: This table shows the MSE, the percentage of non-zero values within the matrix Γ, and the estimated values of θ_L and θ_C for the Double Pooled Elastic-Net model with weights $\phi_{ij} = (\widehat{\gamma}(enet)_{ij})^{-1}$ and λ calculated with equation (6). It shows these results for different values of α and it should be noted that $\alpha = 1$ is equal to the Lasso estimator.