Web Appendix material for "Dual gravity: Using spatial econometrics to control for multilateral resistance"

Kristian Behrens* Cem Ertur† Wilfried Koch‡

September 20, 2010

*Canada Research Chair, Département des Sciences Economiques, Université du Québec à Montréal (UQAM), Canada; CIRPÉE, Canada; and CEPR, London, UK. E-mail address: behrens.kristian@uqam.ca

†LEO, Faculté de Droit d'Economie et de Gestion, Université d'Orléans, France. E-mail address: cem.ertur@univ-orleans.fr

‡*Corresponding author:* Département des Sciences Economiques, Université du Québec à Montréal (UQAM), Canada; and CIRPÉE, Canada. E-mail address: koch.wilfried@uqam.ca

Appendix A. Linearization of the model

The linear approximation of f at $\sigma = 1$ is given by $\ln X_{ij} = f(1) + (\sigma - 1) f'(1)$. Letting $L \equiv \sum_k L_k$, we have:

$$f(1) = \ln Y_j - \ln\left(\sum_k \frac{L_k}{L_i}\right) = \ln Y_j - \ln L + \ln L_i. \tag{A.1}$$

Turning to the derivative, some longer calculations show that

$$\begin{aligned} f'(\sigma) &= \ln Y_j - \ln\left[\sum_k \frac{L_k}{L_i}\left(\frac{\tau_{kj} Y_k}{\tau_{ij} Y_i}\right)^{\frac{1}{\sigma}-1} X_{kj}^{1-\frac{1}{\sigma}}\right] \\ &\quad -\sigma \frac{\sum_k \frac{L_k}{L_i}\left(\frac{1}{\sigma^2}\right)\left[-\left(\frac{\tau_{kj} Y_k}{\tau_{ij} Y_i}\right)^{\frac{1}{\sigma}-1} X_{kj}^{1-\frac{1}{\sigma}} \ln\left(\frac{\tau_{kj} Y_k}{\tau_{ij} Y_i}\right) + \left(\frac{\tau_{kj} Y_k}{\tau_{ij} Y_i}\right)^{\frac{1}{\sigma}-1} X_{kj}^{1-\frac{1}{\sigma}} \ln X_{kj}\right]}{\sum_l \frac{L_l}{L_i} \frac{\tau_{lj} Y_l}{\tau_{ij} Y_i}^{\frac{1}{\sigma}-1} X_{lj}^{1-\frac{1}{\sigma}}}, \end{aligned}$$

thus implying that

$$f'(1) = \ln Y_j - \ln L + \ln L_i + \sum_k \frac{L_k}{L} \ln \frac{\tau_{kj}}{\tau_{ij}} + \sum_k \frac{L_k}{L} \ln \frac{Y_k}{Y_i} - \sum_k \frac{L_k}{L} \ln X_{kj}. \tag{A.2}$$

Using (A.1) and (A.2), the linear approximation of f is given by

$$\begin{aligned} \ln X_{ij} &= \sigma \ln L_i - \sigma \ln L + \sigma \ln Y_j - (\sigma - 1) \ln Y_i - (\sigma - 1) \ln \tau_{ij} \\ &\quad + (\sigma - 1) \sum_k \frac{L_k}{L} \ln \tau_{kj} + (\sigma - 1) \sum_k \frac{L_k}{L} \ln Y_k - (\sigma - 1) \sum_k \frac{L_k}{L} \ln X_{kj}, \end{aligned}$$

which, from the aggregate income constraint $Y_i = w_i L_i$, can be rewritten as

$$\begin{aligned} \ln\left(\frac{X_{ij}}{Y_i Y_j}\right) &= \sigma \sum_k \frac{L_k}{L} \ln \frac{L_k}{L} + (\sigma - 1) \ln Y_j - (\sigma - 1)\left(\ln \tau_{ij} - \sum_k \frac{L_k}{L} \ln \tau_{kj}\right) \\ &\quad -\sigma\left(\ln w_i - \sum_k \frac{L_k}{L} \ln w_k\right) - \sum_k \frac{L_k}{L} \ln Y_k - (\sigma - 1) \sum_k \frac{L_k}{L} \ln X_{kj}. \end{aligned}$$

Making again use of $Y_i = L_i w_i$, and since $\sum_k (L_k / L) = 1$, we then have

$$\begin{aligned} \ln\left(\frac{X_{ij}}{Y_i Y_j}\right) &= \sigma \sum_k \frac{L_k}{L} \ln \frac{L_k}{L} + (\sigma - 1) \sum_k \frac{L_k}{L} \ln Y_j - (\sigma - 1)\left(\ln \tau_{ij} - \sum_k \frac{L_k}{L} \ln \tau_{kj}\right) \\ &\quad -\sigma\left(\ln w_i - \sum_k \frac{L_k}{L} \ln Y_k + \sum_k \frac{L_k}{L} \ln L_k\right) - \sum_k \frac{L_k}{L} \ln Y_k - (\sigma - 1) \sum_k \frac{L_k}{L} \ln X_{kj} \\ &= -\sigma \ln L - (\sigma - 1)\left(\ln \tau_{ij} - \sum_k \frac{L_k}{L} \ln \tau_{kj}\right) - \sigma \ln w_i - (\sigma - 1) \sum_k \frac{L_k}{L} \ln\left(\frac{X_{kj}}{Y_k Y_j}\right). \end{aligned}$$

The foregoing expression has an autoregressive structure with respect to the dependent variable $Z_{ij} \equiv X_{ij}/(Y_i Y_j)$:

$$\ln Z_{ij} = -\sigma \ln L - (\sigma - 1)\left(\ln \tau_{ij} - \sum_k \frac{L_k}{L} \ln \tau_{kj}\right) - \sigma \ln w_i - (\sigma - 1)\sum_k \frac{L_k}{L} \ln Z_{kj}. \tag{A.3}$$

Observe that (A.3) is structurally close to the estimating equations of both Feenstra (2002, 2004) and Anderson and van Wincoop (2003).

Appendix B. Data description

Trade flows betweem U.S. states and Canadian provinces (measured in thousands of U.S. dollars), as well as regional GDPs (also measured in thousands of U.S. dollars) are those used by Anderson and van Wincoop (2003). Their data set can be freely obtained from Robert C. Feenstra's data page (`http://cid.econ.ucdavis.edu/`). The data set also contains bilateral distances (measured in kilometers) between the different states and provinces. They are computed using the great circle formula applied to the state and province capitals' geographic coordinates. Concerning the internal distances of the provinces and states, we compute three different measures. First, following Redding and Venables (2004), we compute the internal distance as two-thirds times the square root of the region's surface divided by π. Regional surface data (in square kilometers) comes from the ArcView database. Alternatively, as a robustness check, we also used half of this surface-based distance measure (i.e., one-third times the square root of the region's surface divided by π). Anderson and van Wincoop's internal distance measure, defined as one-fourth of the minimum distance between an exporter and all the other regions, is readily computed from the distances given in Feenstra's data set.

We augment the above mentionned data set by including population and wage data. First, we use 1988 population figures at the state and province level to compute the regional population shares. U.S. figures come from the U.S. Census Bureau's historical population estimates (`http://www.census.gov/popest/archives/1980s/st8090ts.txt`) and Canadian figures are provided by Statistics Canada (Table 051-0001). Second, the wage data we use is also for 1988 and is the average hourly manufacturing wage at the state and province level. The U.S. data comes from the Bureau of Labor Statistics and is computed for SIC 30 using private sector employment only. The hourly wage is obtained as total annual wages divided by average annual employment, divided by 1,836 hours worked per year. Data is available as a flat file (`ftp://ftp.bls.gov/pub/special.requests/cew/ SIC/history/state/staa7587.zip`). The Canadian average hourly manufacturing wages by province are obtained from Statistics Canada

(Table 281-0008). These wages are converted into U.S. dollars using the 1988 exchange rate obtained from the PennWorld Tables rev6.3.

Appendix C. Border effects

C.1. Homogeneous coefficients. Following Anderson and van Wincoop (2003) we decompose the border effects into two components: the trade-boosting intranational effect and the trade-reducing international effect of the border. To disentangle the two components and to retrieve the full implied border effect (both intranational and international), we proceed as follows. First, we define the border effects as the ratio of trade flows in a world with borders to that which would prevail in a borderless world. Let Z_{ij} denote the former and $\overline{Z}_{ij}$ the latter. Using (10) and (12) in the paper, we then have

$$B_{ij} \equiv \frac{Z_{ij}}{\overline{Z}_{ij}} = \mathrm{e}^{\theta\left[b_{ij} - \sum_k \frac{L_k}{L} b_{kj}\right]} \prod_k \left(\frac{Z_{kj}}{\overline{Z}_{kj}}\right)^{\rho \frac{L_k}{L}}, \tag{C.1}$$

where the term $\mathrm{e}^{\theta\left[b_{ij} - \sum_k \frac{L_k}{L} b_{kj}\right]}$ subsumes the border frictions as a deviation from their population-weighted average. Note that (C.1) defines a log-linear system of all the relative trade flows, which depend on all border effects. Let $\mathbf{B}$ stand for the $n^2 \times 1$ vector of the $\ln(Z_{ij}/\overline{Z}_{ij})$ and let $\mathbf{b}$ stand for the $N^2 \times 1$ vector of the $\left[b_{ij} - \sum_k \frac{L_k}{L} b_{kj}\right]$. The log-linearized version of the system has the following solution, $\mathbf{B} = \theta(\mathrm{I} - \rho\mathbf{W})^{-1}\mathbf{b}$, which allows us to retrieve the border effect as the exponential of the foregoing expression.

Note that (C.1) quite naturally depends upon where regions i and j are located. Four cases may therefore arise with respect to Canada-U.S. trade. Let $\mathrm{popCA} \equiv \sum_{k \in \mathrm{CA}} \frac{L_k}{L}$ (resp., $\mathrm{popUS} \equiv \sum_{k \in \mathrm{US}} \frac{L_k}{L}$) stand for the Canadian (resp., the U.S.) population share. It is readily verified that

$$\theta \left[b_{ij} - \sum_k \frac{L_k}{L} b_{kj}\right] = \begin{cases} -\theta\,\mathrm{popUS} & \text{if} \quad i \in \mathrm{CA}, j \in \mathrm{CA} \\ \theta\,\mathrm{popUS} & \text{if} \quad i \in \mathrm{CA}, j \in \mathrm{US} \\ \theta\,\mathrm{popCA} & \text{if} \quad i \in \mathrm{US}, j \in \mathrm{CA} \\ -\theta\,\mathrm{popCA} & \text{if} \quad i \in \mathrm{US}, j \in \mathrm{US} \end{cases} \tag{C.2}$$

The explicit solution for $\ln B_{ij}$ is then given by

$$\ln B_{ij} = \theta \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\right]_i \mathbf{b}, \tag{C.3}$$

where $[(\mathbf{I} - \rho\mathbf{W})^{-1})]_i$ denotes the i-th line of the matrix. Using (C.2) and (C.3), and the fact that $\mathbf{W}$ is row-standardized and has a special structure which implies that $\mathbf{Wb} = 0$, the border effects are finally given as follows:

$$\ln B_{ij} = \begin{cases} -\theta \,\text{popUS} & \text{if} \quad i \in \text{CA}, j \in \text{CA} \\ \theta \,\text{popUS} & \text{if} \quad i \in \text{CA}, j \in \text{US} \\ \theta \,\text{popCA} & \text{if} \quad i \in \text{US}, j \in \text{CA} \\ -\theta \,\text{popCA} & \text{if} \quad i \in \text{US}, j \in \text{US} \end{cases}$$

These expressions for the border effects reveal several interesting points. First, the expressions for CA-CA and U.S.-U.S. can be interpreted as the *trade-boosting* effect of the international border on flows within each country. Indeed, when ξ is positive and ρ is negative (as implied by our model), the trade flows within each country will be larger in a world with border than in a borderless world. The reason is that the border protects domestic firms from import competition and gives them an advantage in the home market. Second, the expressions for CA-U.S. and U.S.-CA can be interpreted as the *trade-reducing* effect of the international border on flows across countries. When ξ is positive and ρ is negative, the trade flows across countries will be smaller in a world with borders than in a borderless world. Third, as in Anderson and van Wincoop (2003), smaller countries will have larger implied border effects than large countries since their magnitude depends positively on the size of the trading partner, as measured by its population share. The reason is that the border affects the small country more than the large country, as it creates trade frictions for a larger share of the total demand served by its firms. Finally, the full border effect (combining the trade-boosting and trade-reducing effects), is given by $\mathrm{e}^{-2\xi\rho\,\text{popUS}}$ for Canadian provinces and by $\mathrm{e}^{-2\xi\rho\,\text{popCA}}$ for U.S. states.

C.2. Heterogeneous coefficients. In the heterogeneous coefficients model, we can retrieve the region-specific border effects in an analogous way to that presented in the foregoing Appendix C.1. Starting from (C.1), taking logarithms and rearranging, we readily obtain:

$$\ln Z_{ij} - \ln \overline{Z}_{ij} = \underbrace{\frac{\theta}{1-\rho\frac{L_i}{L}}}_{\overline{\theta}_i} \left[b_{ij} - \sum_k \frac{L_k}{L} b_{kj} \right] - \underbrace{\frac{\rho}{1-\rho\frac{L_i}{L}}}_{\overline{\rho}_i} \sum_{k\neq i} \frac{L_k}{L} \left(\ln Z_{kj} - \ln \overline{Z}_{kj} \right). \tag{C.4}$$

Using the expressions established in Appendix C.1. (which remain unchanged in the heterogeneous coefficient case), as well as the same matrix notation, we then obtain:

$$\ln B_{ij} = \overline{\theta}_i \left[\mathbf{I} - \overline{\rho} \otimes \mathbf{W_d} \right]_i^{-1} \mathbf{b}.$$

The only change with respect to the homogeneous coefficient case is that the coefficient $\overline{\theta}_i$ captures the *local border frictions*, whereas $\overline{\rho}$ is a vector of elements accounting for the varying 'thoughness of competition' in the different regional markets.

Appendix D. Log-likelihood and the information matrix

In this technical appendix, we derive the theoretical properties of the heterogeneous coefficients SARMA model with country-specific autoregressive parameters ($\overline{\rho}_j$ and $\overline{\lambda}_j$ for $j = 1, 2$) and region-specific non-autoregressive parameters ($\overline{\beta}_{1i}$, $\overline{\beta}_{2i}$ and $\overline{\theta}_i$ for $i = 1, ..., n$).

D.1. Model. To make notation as compact as possible, let $\mathbf{V}_i$ stand for the diagonal matrix defined by $\mathbf{V}_i \equiv \mathbf{E}_i \otimes \mathbf{I_n}$, where $\mathbf{E}_i = [\ 0 \mid 0 \mid \ \dots e_i \dots \ \mid 0 \mid 0\ ]$ with e_i (the i-th vector of the canonical base of $\mathbb{R}^n$) in position i and zero column vectors elsewhere. The diagonal matrix $\mathbf{V}_i$ is, therefore, a selection matrix with 1 on its main diagonal for the selected variables and 0 otherwise. Note that, by construction, $\sum_{i=1}^n \mathbf{V}_i = \mathbf{I_{n^2}}$. Analogously, let $\mathbf{D}_j$ stand for the diagonal selection matrix with 1 on its main diagonal for selecting canadian provinces or U.S. states, and 0 otherwise. Again, $\sum_{j=1}^2 \mathbf{D}_j = \mathbf{I_{n^2}}$ by construction. Using the definitions of $\mathbf{V}_i$ and $\mathbf{D}_j$, the estimating equation (19) can be rewritten as follows:

$$\begin{aligned} \mathbf{Z} &= \sum_i \mathbf{V}_i \left\{ \overline{\beta}_{1i}\widetilde{\mathbf{d}} + \overline{\beta}_{2i}\mathbf{w} + \overline{\theta}_i \widetilde{\mathbf{b}} \right\} + \sum_j \mathbf{D}_j\ \overline{\rho}_j \mathbf{W_d Z} + \mathbf{u}, \\ &= \ \mathbf{\Gamma}\overline{\beta} + \mathbf{D}\big(\overline{\rho} \otimes \mathbf{W_d}\big)\mathbf{Z} + \mathbf{u}, \end{aligned} \tag{D.1}$$

where

$$\mathbf{u} = \varepsilon + \mathbf{D}(\overline{\lambda} \otimes \mathbf{W_d})\varepsilon. \tag{D.2}$$

In expressions (D.1) and (D.2), $\mathbf{W_d} \equiv \mathbf{W} - \mathbf{W_{diag}}$ denotes the interaction matrix ; $\mathbf{\Gamma} \equiv \mathbf{V}\left(\mathbf{I_n} \otimes \mathbf{M}\right)$ denotes the $n^2 \times 6n$ block diagonal matrix of explanatory variables, with $\mathbf{M} \equiv [\ \widetilde{\mathbf{d}} \mid \widetilde{\mathbf{w}} \mid \widetilde{\mathbf{b}}\ ]$; $\mathbf{V} \equiv [\ \mathbf{V}_1 \mid \mathbf{V}_2\ \dots\ \mathbf{V}_i\ \dots\ \mathbf{V}_n\ ]$ stands for the $n^2 \times n^3$ selection matrix which extracts local subsamples from the full sample; $\overline{\beta}$ is the $6n \times 1$ vector of region-specific parameters; and $\overline{\rho}$ and $\overline{\lambda}$ are the 2×1 vectors of autoregressive interaction coefficients. Expressions (D.1) and (D.2) constitute the most compact and general specification of our model and will be useful for deriving the log-likelihood function and the information matrix.

Note that, in contrast to the SARMA model in the homogeneous case, we need to estimate two autoregressive interaction coefficients associated with different interaction matrices, the sum of which is equal to the interaction matrix that is used in the homogenous case ($\overline{\rho}_j = \rho$ and $\overline{\lambda}_j = \lambda$ for $j = 1, 2$). Letting $\mathbf{S}(\overline{\rho}) = \mathbf{I_{n^2}} - \mathbf{D}(\overline{\rho} \otimes \mathbf{W_d})$ and $\mathbf{S}(\overline{\lambda}) = \mathbf{I_{n^2}} + \mathbf{D}(\overline{\lambda} \otimes \mathbf{W_d})$, the equilibrium vector $\mathbf{Z}$ is as follows:

$$\mathbf{Z} = \mathbf{S}(\overline{\rho})^{-\mathbf{1}}\big[\mathbf{\Gamma}\beta + \mathbf{S}(\overline{\lambda})\varepsilon\big], \tag{D.3}$$

where $\mathbf{S}(\overline{\rho})$ and $\mathbf{S}(\overline{\lambda})$ are both non-singular. We propose to estimate this model by standard maximum likelihood techniques.

D.2. Log-likelihood. Let $\varepsilon(\theta) \equiv \mathbf{S}(\overline{\lambda})^{-\mathbf{1}}\left[\mathbf{S}(\overline{\rho})\mathbf{Z} - \mathbf{\Gamma}\beta\right]$, where $\theta = \left[\ \beta' \mid \overline{\rho}' \mid \overline{\lambda}'\ \right]'$. The log-likelihood of (D.3) is then given by:

$$\ln L(\theta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|\mathbf{S}(\overline{\rho})| - \ln|\mathbf{S}(\overline{\lambda})| - \frac{1}{2\sigma^2}\varepsilon'(\theta)\varepsilon(\theta). \tag{D.4}$$

The Maximum Likelihood Estimators (MLE) $\hat{\theta}_{ML}$ and $\hat{\sigma}^2_{ML}$ are derived from the maximization of equation (D.4). In order to compute these estimators, it is convenient to work with the concentrated log-likelihood.

D.3. Estimators. The first-order conditions yield the following expressions for the estimators as a function of the autoregressive parameters:

$$\hat{\beta}_{ML}(\overline{\rho}, \overline{\lambda}) = \left[\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\right]^{-1}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z} \tag{D.5}$$

$$\hat{\sigma}^2_{ML}(\overline{\rho}, \overline{\lambda}) = \frac{1}{n}\mathbf{Z}'\mathbf{S}'(\overline{\rho})\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{M}(\overline{\lambda})\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z}, \tag{D.6}$$

with $\mathbf{M}(\overline{\lambda}) \equiv \mathbf{I_{n^2}} - \mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\left[\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\right]^{-1}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}$ the $n^2 \times n^2$ projection matrix.

Proof. The first-order condition with respect to β is given by:

$$\nabla_\beta \ln L(\theta, \sigma^2) = 0 \quad \Longleftrightarrow \quad \mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z} = \mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\beta,$$

which directly yields

$$\hat{\beta}_{ML}(\overline{\rho}, \overline{\lambda}) = \left[\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\right]^{-1}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z}.$$

The first-order condition with respect to σ^2 is given by:

$$\nabla_{\sigma^2} \ln L(\theta, \sigma^2) = 0 \quad \Longleftrightarrow \quad -n + \frac{1}{\sigma^2}\varepsilon'(\theta)\varepsilon(\theta) = 0,$$

which directly yields

$$\hat{\sigma}^2_{ML}(\overline{\rho}, \overline{\lambda}) = \frac{1}{n}\varepsilon'(\theta)\varepsilon(\theta).$$

Using the definition of the projection matrix $\mathbf{M}(\overline{\lambda})$ we then obtain:

$$\hat{\sigma}^2_{ML}(\overline{\rho}, \overline{\lambda}) = \frac{1}{n}\mathbf{Z}'\mathbf{S}'(\overline{\rho})\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{M}(\overline{\lambda})\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z},$$

which establishes the result. ■

D.4. Maximization of the concentrated log-likelihood. The concentrated log-likelihood can be rewritten as a function of the vectors $\overline{\rho}$ and $\overline{\lambda}$ as follows:

$$\begin{aligned}\ln L_c(\overline{\rho},\overline{\lambda}) &= -\frac{n}{2}(\ln(2\pi)+1)+\ln|\mathbf{S}(\overline{\rho})|+\ln|\mathbf{S}(\overline{\lambda})| \\ &- \frac{n}{2}\ln\left[\frac{\left(\mathbf{e}_0(\overline{\lambda})-\sum_{i=1}^{n}\rho_i\mathbf{e}_i(\overline{\lambda})\right)'\left(\mathbf{e}_0(\overline{\lambda})-\sum_{i=1}^{n}\rho_i\mathbf{e}_i(\overline{\lambda})\right)}{n}\right],\end{aligned} \tag{D.7}$$

where $\mathbf{e}_0(\overline{\lambda})=\mathbf{M}(\overline{\lambda})\mathbf{S}(\overline{\lambda})^{-1}\mathbf{Z}$, and where $\mathbf{e}_i(\overline{\lambda})=\mathbf{M}(\overline{\lambda})\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\mathbf{Z}$ for $i=1,2$. Put differently, $\mathbf{e}_0(\overline{\lambda})$ is the vector of residuals of a regression of $\mathbf{Z}$ on $\mathbf{\Gamma}$, and $\mathbf{e}_i(\overline{\lambda})$ is the vector of residuals of a regression of $\mathbf{D}_i\mathbf{W_d}\mathbf{Z}$ on $\mathbf{\Gamma}$, for $i=1,2$.

Proof. Note first that, using the expression for $\hat{\sigma}^2_{ML}(\overline{\rho},\overline{\lambda})$, we have the following relation: $\varepsilon'(\theta)\varepsilon(\theta)=n\hat{\sigma}^2_{ML}(\overline{\rho},\overline{\lambda})$. Moreover, using the expression of the projection matrix $\mathbf{M}(\overline{\lambda})$, it is straightforward to obtain the concentrated log-likelihood. ■

The MLEs of $\overline{\rho}$ and $\overline{\lambda}$, denoted respectively by $\widehat{\overline{\rho}}_{ML}$ and $\widehat{\overline{\lambda}}_{ML}$, maximize the concentrated log-likelihood (D.7). The MLEs of β and of σ^2 are then given by $\hat{\beta}_{ML}\equiv\beta_{ML}(\widehat{\overline{\rho}}_{ML},\widehat{\overline{\lambda}}_{ML})$ and by $\hat{\sigma}^2_{ML}\equiv\sigma^2_{ML}(\widehat{\overline{\rho}}_{ML},\widehat{\overline{\lambda}}_{ML})$, respectively.

D.5. Information matrix. The asymptotic covariance matrix of the maximum likelihood estimators is given by the inverse of the information matrix, which is defined as follows:

$$\mathbf{I}(\tilde{\theta})=-E\left[\nabla^2_{\tilde{\theta},\tilde{\theta}'}\ln L(\tilde{\theta})\right] \tag{D.8}$$

with $\tilde{\theta}=(\theta',\sigma^2)'$. We can use the following estimator for this matrix:

$$\left[\hat{\mathbf{I}}(\hat{\tilde{\theta}})\right]^{-1}=\left[-\nabla^2_{\hat{\tilde{\theta}},\hat{\tilde{\theta}}'}\ln L(\hat{\tilde{\theta}})\right]^{-1} \tag{D.9}$$

To obtain this estimate, we need to compute that derivatives of the log-likelihood function.

D.6. First-order derivatives of the log-likelihood. We start with the first-order derivatives. By definition, $\varepsilon(\theta)=\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z}-\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\beta$. Because the transpose of a scalar is that scalar itself, we then obtain:

$$\begin{aligned}\varepsilon'(\theta)\varepsilon(\theta) &= \mathbf{Z}'\mathbf{S}'(\overline{\rho})\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z}-2\beta'\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z} \\ &+ \beta'\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\beta.\end{aligned} \tag{D.10}$$

The derivative of the log-likelihood with respect to β is given by:

$$\begin{aligned}\nabla_\beta\ln L(\theta,\sigma^2) &= -\frac{1}{2\sigma^2}\nabla_\beta\left[\varepsilon'(\theta)\varepsilon(\theta)\right] \\ &= -\frac{1}{2\sigma^2}\left[-2\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{S}(\overline{\rho})\mathbf{Z}+2\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}\beta\right] \\ &= \frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta).\end{aligned} \tag{D.11}$$

The derivative of the log-likelihood with respect to σ^2 is given by:

$$\nabla_{\sigma^2} \ln L(\theta, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{2}{4(\sigma^2)^2}\varepsilon'(\theta)\varepsilon(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\varepsilon'(\theta)\varepsilon(\theta). \tag{D.12}$$

The derivative of the log-likelihood with respect to $\overline{\rho}_i$, for $i = 1, 2$, is given by:

$$\nabla_{\overline{\rho}_i} \ln L(\theta, \sigma^2) = -\mathrm{tr}\left(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W}_{\mathbf{d}}\right) + \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{W}_{\mathbf{d}}'\mathbf{D}_i'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta). \tag{D.13}$$

Proof. To establish the expression for $\nabla_{\overline{\rho}_i} \ln L(\theta, \sigma^2)$, note that

$$\nabla_{\overline{\rho}} \ln L(\theta, \sigma^2) = \nabla_{\overline{\rho}} \ln |\mathbf{S}(\overline{\rho})| - \frac{1}{2\sigma^2}\nabla_{\overline{\rho}}\big(\varepsilon'(\theta)\varepsilon(\theta)\big)$$

Computation of the first term requires to apply the theorem for chain derivation of a matrix expression. Applying it for each element of the vector $\overline{\rho}$, we have:

$$\nabla_{\overline{\rho}_i} \ln |\mathbf{S}(\overline{\rho})| = \mathrm{tr}\left(\nabla_{\mathbf{S}(\overline{\rho})}(\ln |\mathbf{S}(\overline{\rho})|)'\nabla_{\overline{\rho}_i}\mathbf{S}(\overline{\rho})\right),$$

with $\nabla_{\mathbf{S}(\overline{\rho})} \ln |\mathbf{S}(\overline{\rho})| = (\mathbf{S}(\overline{\rho})')^{-1}$, and with

$$\nabla_{\overline{\rho}_i}\mathbf{S}(\overline{\rho}) = -\mathbf{D}\left[(\nabla_{\overline{\rho}_i}\overline{\rho}) \otimes \mathbf{W}_{\mathbf{d}} + \overline{\rho} \otimes (\nabla_{\overline{\rho}_i}\mathbf{W}_{\mathbf{d}})\right] = -\mathbf{D}(\mathbf{e}_i \otimes \mathbf{W}_{\mathbf{d}}) = -\mathbf{D}_i\mathbf{W}_{\mathbf{d}}.$$

As in the foregoing, $\mathbf{e}_i$ denotes the i-th vector of the canonical base, with 1 in position i and 0 otherwise. We then, therefore, obtain:

$$\nabla_{\overline{\rho}_i} \ln |\mathbf{S}(\overline{\rho})| = -\mathrm{tr}\left(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W}_{\mathbf{d}}\right).$$

To compute the second term, note that

$$\begin{aligned}
\nabla_{\overline{\rho}_i}(\varepsilon'(\theta)\varepsilon(\theta)) &= \nabla_{\overline{\rho}_i}(\varepsilon'(\theta)\varepsilon(\theta)) + \varepsilon(\theta)\nabla_{\overline{\rho}_i}\varepsilon(\theta) \\
&= \mathbf{Z}'\nabla_{\overline{\rho}_i}\mathbf{S}'(\overline{\rho})\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta) + \varepsilon'(\theta)\mathbf{S}(\overline{\lambda})^{-1}\nabla_{\overline{\rho}_i}\mathbf{S}(\overline{\rho})\mathbf{Z} \\
&= 2\mathbf{Z}'\nabla_{\overline{\rho}_i}\mathbf{S}'(\overline{\rho})\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta),
\end{aligned}$$

where we use the property that the transpose of a scalar is the scalar itself. We obtain:

$$\nabla_{\overline{\rho}_i}(\varepsilon'(\theta)\varepsilon(\theta)) = -2\mathbf{Z}'(\mathbf{e}_i \otimes \mathbf{W}_{\mathbf{d}})'\mathbf{D}'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta) = -2\mathbf{Z}'\mathbf{W}_{\mathbf{d}}'\mathbf{D}_i'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta)$$

Putting finally the expressions together, we have:

$$\nabla_{\overline{\rho}_i} \ln L(\theta, \sigma^2) = -\mathrm{tr}\left(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W}_{\mathbf{d}}\right) + \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{W}_{\mathbf{d}}'\mathbf{D}_i'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta)$$

for $i = 1, 2$, which establishes the result. ∎

Next, the derivative of the log-likelihood with respect to the vector $\overline{\lambda}$ is given by:

$$\nabla_{\overline{\lambda}_i} \ln L(\theta, \sigma^2) = \mathrm{tr}\left(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right) - \frac{1}{\sigma^2}\varepsilon'(\theta)\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta). \tag{D.14}$$

Proof. To begin with, note that

$$\nabla_{\overline{\lambda}} \ln L(\theta, \sigma^2) = -\nabla_{\overline{\lambda}} \ln |\mathbf{S}(\overline{\lambda})| - \frac{1}{2\sigma^2}\nabla_{\overline{\lambda}}(\varepsilon'(\theta)\varepsilon(\theta))$$

Computation of the first term requires to apply the theorem for chain derivation of a matrix expression. Applying it for each element of the vector $\overline{\lambda}$, we have:

$$\nabla_{\overline{\lambda}_i} \ln |\mathbf{S}(\overline{\lambda})| = \mathrm{tr}\left(\nabla_{\mathbf{S}(\overline{\lambda})} \ln |\mathbf{S}(\overline{\lambda})|' \nabla_{\overline{\lambda}_i}\mathbf{S}(\overline{\lambda})\right),$$

with $\nabla_{\mathbf{S}(\overline{\lambda})} \ln |\mathbf{S}(\overline{\lambda})| = (\mathbf{S}(\overline{\lambda})')^{-1}$, and with

$$\nabla_{\overline{\lambda}_i}\mathbf{S}(\overline{\lambda}) = \mathbf{D}\left[(\nabla_{\overline{\lambda}_i}\overline{\lambda}) \otimes \mathbf{W_d} + \overline{\lambda} \otimes (\nabla_{\overline{\lambda}_i}\mathbf{W_d})\right] = \mathbf{D}(\mathbf{e}_i \otimes \mathbf{W_d}) = \mathbf{D}_i\mathbf{W_d}. \tag{D.15}$$

As in the foregoing, $\mathbf{e}_i$ denotes the i-th vector of the canonical base, with 1 in position i and 0 otherwise. We then, therefore, obtain:

$$\nabla_{\overline{\lambda}_i} \ln |\mathbf{S}(\overline{\lambda})| = \mathrm{tr}\left(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right).$$

To compute the second term, note that

$$\begin{aligned}
\nabla_{\overline{\lambda}_i}(\varepsilon'(\theta)\varepsilon(\theta)) &= \nabla_{\overline{\lambda}_i}\varepsilon'(\theta)\varepsilon(\theta) + \varepsilon'(\theta)\nabla_{\overline{\lambda}_i}\varepsilon(\theta) \\
&= \left[\mathbf{S}(\overline{\rho})\mathbf{Z} - \mathbf{\Gamma}\beta\right]' \nabla_{\overline{\lambda}_i}\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta) + \varepsilon'(\theta)\nabla_{\overline{\lambda}_i}\mathbf{S}(\overline{\lambda})^{-1}\left[\mathbf{S}(\overline{\rho})\mathbf{Z} - \mathbf{\Gamma}\beta\right] \\
&= 2\left[\mathbf{S}(\overline{\rho})\mathbf{Z} - \mathbf{\Gamma}\beta\right]' \nabla_{\overline{\lambda}_i}\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta)
\end{aligned}$$

where we use the property that the transpose of a scalar is the scalar itself. We obtain:

$$\nabla_{\overline{\lambda}_i}\mathbf{S}'(\overline{\lambda})^{-1} = -\mathbf{S}'(\overline{\lambda})^{-1}\nabla_{\overline{\lambda}_i}\mathbf{S}'(\overline{\lambda})\mathbf{S}'(\overline{\lambda})^{-\mathbf{1}} = -\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}.$$

Putting finally the expressions together, we have:

$$\nabla_{\overline{\lambda}_i}(\varepsilon'(\theta)\varepsilon(\theta)) = -2\left[\mathbf{S}(\overline{\rho})\mathbf{Z} - \mathbf{\Gamma}\beta\right]' \mathbf{S}'(\overline{\lambda})^{-1}\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta)$$

and

$$\nabla_{\overline{\lambda}_i} \ln L(\theta, \sigma^2) = -\mathrm{tr}\left(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right) + \frac{1}{\sigma^2}\left[\mathbf{S}(\overline{\rho})\mathbf{Z} - \mathbf{\Gamma}\beta\right]' \mathbf{S}'(\overline{\lambda})^{-1}\mathbf{W}'_{\mathbf{d}}\mathbf{D_i}'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta),$$

for $i = 1, 2$, which establishes the result. ■

D.7. Second-order derivatives of the log-likelihood. We next turn to the second-order derivatives with respect to β. Deriving (D.11) with respect to β, we obtain:

$$\nabla^2_\beta \ln L(\theta,\sigma^2) = \frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\nabla_\beta\varepsilon(\theta) = -\frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{\Gamma}. \tag{D.16}$$

Deriving (D.11) with respect to σ^2 yields:

$$\frac{\partial\big(\nabla_\beta \ln L(\theta,\sigma^2)\big)}{\partial\sigma^2} = -\frac{1}{(\sigma^2)^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta). \tag{D.17}$$

Taking the derivative of (D.11) with respect to $\overline{\rho}_j$ yields:

$$\begin{aligned}\frac{\partial\big(\nabla_\beta \ln L(\theta,\sigma^2)\big)}{\partial\overline{\rho}_j} &= \frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\nabla_{\overline{\rho}_j}\varepsilon(\theta)\\ &= \frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\nabla_{\overline{\rho}_j}\mathbf{S}(\overline{\rho})\mathbf{Z}\\ &= -\frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W}_\mathbf{d}\mathbf{Z}\end{aligned} \tag{D.18}$$

for $j=1,2$. Finally, the derivative of (D.11) with respect to $\overline{\lambda}_j$ is given by:

$$\begin{aligned}\frac{\partial\big(\nabla_\beta \ln L(\theta,\sigma^2)\big)}{\partial\overline{\lambda}_j} &= \frac{1}{\sigma^2}\mathbf{\Gamma}'\left[\frac{\partial\mathbf{S}'(\overline{\lambda})^{-\mathbf{1}}}{\partial\overline{\lambda}_j}\varepsilon(\theta) + \mathbf{S}'(\overline{\lambda})^{-\mathbf{1}}\frac{\partial\varepsilon(\theta)}{\partial\overline{\lambda}_j}\right] \qquad \text{(D.19)}\\ &= -\frac{1}{\sigma^2}\mathbf{\Gamma}'\left[\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{W}_\mathbf{d}'\mathbf{D}_j'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta)\right.\\ &\quad \left.+\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W}_\mathbf{d}\mathbf{S}(\overline{\lambda})^{-1}(\mathbf{S}(\overline{\rho})\mathbf{Z}-\mathbf{\Gamma}\beta)\right]\\ &= -\frac{1}{\sigma^2}\mathbf{\Gamma}'\mathbf{S}'(\overline{\lambda})^{-1}\left[\mathbf{W}_\mathbf{d}'\mathbf{D}_j'\mathbf{S}'(\overline{\lambda})^{-1} + \mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W}_\mathbf{d}\right]\varepsilon(\theta)\end{aligned}$$

for $j=1,2$. We next derive (D.12) with respect to σ^2 to obtain the following second-order derivative:

$$\frac{\partial^2 \ln L(\theta,\sigma^2)}{\partial(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\varepsilon'(\theta)\varepsilon(\theta). \tag{D.20}$$

The derivative of (D.12) with respect to $\overline{\rho}_j$ is computed as follows:

$$\frac{\partial^2 \ln L(\theta,\sigma^2)}{\partial\sigma^2\partial\overline{\rho}_j} = \frac{1}{2(\sigma^2)^2}\nabla_{\overline{\rho}_j}\left(\varepsilon'(\theta)\varepsilon(\theta)\right) = -\frac{1}{(\sigma^2)^2}\mathbf{Z}'\mathbf{W}_\mathbf{d}'\mathbf{D}_j'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta), \tag{D.21}$$

for $j=1,2$. The derivative of (D.12) with respect to λ_j is given by:

$$\frac{\partial^2 \ln L(\theta,\sigma^2)}{\partial\sigma^2\partial\overline{\lambda}_j} = \frac{1}{2(\sigma^2)^2}\nabla_{\overline{\lambda}_j}\left(\varepsilon'(\theta)\varepsilon(\theta)\right) = -\frac{1}{(\sigma^2)^2}\varepsilon'(\theta)\mathbf{W}_\mathbf{d}'\mathbf{D}_j'\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta), \tag{D.22}$$

for $j=1,2$. We next derive (D.13) with respect to $\overline{\rho}_j$:

$$\frac{\partial^2 \ln L(\theta,\sigma^2)}{\partial\overline{\rho}_i\partial\overline{\rho}_j} = -\frac{\partial\big(\mathrm{tr}(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W}_\mathbf{d})\big)}{\partial\overline{\rho}_j} + \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{W}_\mathbf{d}'\mathbf{D}_i'\mathbf{S}'(\overline{\lambda})^{-1}\nabla_{\overline{\rho}_j}\varepsilon(\theta) \tag{D.23}$$

for $j=1,2$. Since

$$\begin{aligned}
\frac{\partial\big(\mathrm{tr}(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W_d})\big)}{\partial\overline{\rho}_j} &= \mathrm{tr}\left(\nabla_{\overline{\rho}_j}\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W_d}\right)\\
&= \mathrm{tr}\left(-\mathbf{S}(\overline{\rho})^{-1}\nabla_{\overline{\rho}_j}\mathbf{S}(\overline{\rho})\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W_d}\right)\\
&= \mathrm{tr}\left(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_j\mathbf{W_d}\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W_d}\right),
\end{aligned}$$

and since

$$\nabla_{\overline{\rho}_j}\varepsilon(\theta) = \mathbf{S}(\overline{\lambda})^{-1}\nabla_{\overline{\rho}_j}\mathbf{S}(\overline{\rho})\mathbf{Z} = -\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W_d}\mathbf{Z},$$

we finally obtain:

$$\frac{\partial^2\ln L(\theta,\sigma^2)}{\partial\overline{\rho}_i\partial\overline{\rho}_j} = -\mathrm{tr}\left(\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_j\mathbf{W_d}\mathbf{S}(\overline{\rho})^{-1}\mathbf{D}_i\mathbf{W_d}\right) - \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W_d}\mathbf{Z}. \quad \text{(D.24)}$$

We next derive (D.13) with respect to $\overline{\lambda}_j$, which yields:

$$\begin{aligned}
\frac{\partial^2\ln L(\theta,\sigma^2)}{\partial\overline{\rho}_i\partial\overline{\lambda}_j} &= \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\left[\nabla_{\overline{\lambda}_j}\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta) + \mathbf{S}'(\overline{\lambda})^{-1}\nabla_{\overline{\lambda}_j}\varepsilon(\theta)\right]\\
&= -\frac{1}{\sigma^2}\mathbf{Z}'\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\left[\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_j\mathbf{S}'(\overline{\lambda})^{-1} + \mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W_d}\right]\varepsilon(\theta), \qquad \text{(D.25)}
\end{aligned}$$

for $j=1,2$. Finally, the derivative of (D.14) with respect to $\overline{\lambda}_j$ is computed as follows:

$$\begin{aligned}
\frac{\partial^2\ln L(\theta,\sigma^2)}{\partial\overline{\lambda}_i\partial\overline{\lambda}_j} &= -\frac{\partial\big(\mathrm{tr}(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d})\big)}{\partial\overline{\lambda}_j} + \frac{1}{\sigma^2}\Big[\nabla_{\overline{\lambda}_j}\varepsilon'(\theta)\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta)\\
&+ \varepsilon'(\theta)\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\left(\nabla_{\overline{\lambda}_j}\mathbf{S}'(\overline{\lambda})^{-1}\varepsilon(\theta) + \mathbf{S}'(\overline{\lambda})^{-1}\nabla_{\overline{\lambda}_j}\varepsilon(\theta)\right)\Big] \qquad \text{(D.26)}
\end{aligned}$$

for $j=1,2$. We have:

$$\begin{aligned}
\frac{\partial\big(\mathrm{tr}(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d})\big)}{\partial\overline{\lambda}_j} &= \mathrm{tr}\left(\nabla_{\overline{\lambda}_j}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right)\\
&= \mathrm{tr}\left(-\mathbf{S}(\overline{\lambda})^{-1}\nabla_{\overline{\lambda}_j}\mathbf{S}(\overline{\lambda})\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right)\\
&= -\mathrm{tr}\left(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W_d}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right)
\end{aligned}$$

so that, using the foregoing results, we obtain:

$$\begin{aligned}
\frac{\partial^2\ln L(\theta,\sigma^2)}{\partial\overline{\lambda}_i\partial\overline{\lambda}_j} &= \mathrm{tr}\left(\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W_d}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_i\mathbf{W_d}\right) - \frac{1}{\sigma^2}\varepsilon'(\theta)\Big[\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_j\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\\
&+ \mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{W}'_{\mathbf{d}}\mathbf{D}'_j\mathbf{S}'(\overline{\lambda})^{-1} + \mathbf{W}'_{\mathbf{d}}\mathbf{D}'_i\mathbf{S}'(\overline{\lambda})^{-1}\mathbf{S}(\overline{\lambda})^{-1}\mathbf{D}_j\mathbf{W_d}\Big]\varepsilon(\theta). \qquad \text{(D.27)}
\end{aligned}$$