Asymptotic Expansions and Approximate Moments for Non-Linear Panel Data Models With Separable Errors

Paul Rilstone

March 2015

York University

Preliminary and Incomplete

Abstract: Third-order stochastic expansions are derived for the fixed and random effects versions of non-linear panel data models with separable errors. These are used to derive the approximate bias, mean-squared error, skewness and kurtosis of the estimators. In turn, these moments are used to derive Edgeworth, saddlepoint and related expansions for these estimators. Illustrations are provided. Simulations are reported indicating the confidence intervals based on these expansions have substantially better coverage properties than those based on standard asymptotic results and somewhat better than those provided by the bootstrap.

1. Introduction

A substantial amount of recent research has dealt with the asymptotic properties of non-linear estimators based on panel data. Much of this research is focused on bias-reduction in transformation models in which the estimating equations cannot be written as additive functions of, say, individual (constant over time) effects and idiosyncratic effects. Notable examples are probit and logit models and certain Poisson count models in which individual effect enter into an underlying latent model. In these cases, the incidental parameter problem results in what may be seen as a first-order bias. See, e.g., Phillips and Moon (1999), Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Hahn and Kuersteiner (2011), Hahn and Newey (2004), Carro (2007), Fernandez-Val (2009), and Fernandez-Val and Vella (2011) Fernandez-Val and Martin Weidner (2013). First order bias can also occur in linear models as per Nickell (1981) with correlation between residuals and lagged values of the dependent variable used as explanatory variables although this is not the focus here.

Although non-linear panel data models with non-additive errors is a more general model an additivity assumption allows for simpler results and fairly straightforward derivation of moments beyond the first and second and, consequently, derivation of Edgeworth and related expansions of similar order. Such derivations do not seem to have been attempted. We also note that many of the recent contributions for non-additive errors depend on an assumption of increasing T. This may not be plausible in some instances, or if it is, the assumption that N is increasing may not be that credible.

In the sense of the class of models we address, then, our goals are relatively modest. On the other hand, for the class of models we address, we provide substantially fuller results. For the class of models we look at, we derive third order asymptotic expansions and use these to derive the approximate first through fourth moments of the estimators. This permits us to do bias corrections and second order mean-squared error (MSE) calculations as well as provide the basis for higher order analytical analysis say through the Edgeworth and saddlepoint approximations.

Higher-order approximations to the sampling distributions of estimators is a central feature of econometrics. There are a number of approaches for doing so. Stochastic expansions and the derivation of approximate moments of estimators based on the these expansions are perhaps the most common method to capture certain features of most sampling distribution of an estimator not captured by standard asymptotic theory. Relevant monographs are Ullah (2004) and Butler (2007). To name a few more closely related papers, Lahiri (1992) and Taniguchi and Puri (1996) derived Edgeworths for M-estimators of linear regression models. Rilstone, et al (1996, henceforth RSU) derived third–order stochastic expansions and the second order bias and mean squared error of nonlinear estimators. Newey and Smith (2004) compared the second order bias and variance of Generalized Empirical Likelihood (GEL) and Generalized Method of Moments (GMM) estimators. Hengartner and Wegkamp (2004) examined single parameter M-estimators. Edgeworth expansions for Wald statistics for nonlinear restrictions were first developed by Phillips and Park (1988).

The results derived below reflect common knowledge with respect to both basic panel data models and the higher-order asymptotics literature. With respect to fixed effects models, the moments and expansions for within estimators resemble those for nonlinear models with the rejoinder that the usual terms in these expansions such as residuals, gradients and higher order derivatives, are all filtered by the usual differencing from means matrix. With respect to random effects models, at least for the class considered here in which the variances disturbance are estimated by sample analogs, simultaneously with the parameters of interest, the higher order moments of the estimators of the parameters of interest are effected by the estimation of the variance of individual-specific errors.

The benefits and costs, at least from the perspective of standard first-order asymptotic theory are well documented and discussed in text book treatments. We will not dwell on these here save to note the usual points that random effects estimators are efficient, but inconsistent, if the individual effects are correlated with the regressors. Also, as is usual in fixed effects cases, estimation is of those parameters in which the gradient varies over time and not swept out by mean-differencing.

A number of results emerge. The basic results are extensions of existing results found e.g. in Rilstone et al (1996) to the panel data case. In the random effects case, estimation of the disturbances' covariance matrix leads to additional higher-order terms in the stochastic expansions.

The discussion proceeds as follows. In the next section we discuss the basic class of models covered. In section 3 we state some basic results regarding stochastic expansions and approximate moments for non-linear estimators. In Section 4 we apply this to random effects models. Section 5 does this for fixed effects models. Section 6 derives comparable results for the Poisson model with individual effects. Section 7 reports on a simple Monte Carlo experiment which highlights the analytical results. Section 7 concludes.

2. Non-linear Panel Data Models

The basic linear regression panel data models are readily extended to a non-linear setting with with

$$\begin{aligned} y_{it} &= g(X_{it};\beta) + u_{it}, \qquad t = 1, \ldots, T, \quad i = 1, \ldots, N \\ u_{it} &= v_i + e_{it} \end{aligned} \tag{2.1}$$

or

$$y_i = g_i(\beta) + v_i \iota_T + e_i$$

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad g_i(\theta) = \begin{pmatrix} g(X_{i1};\beta) \\ g(X_{i2};\beta) \\ \vdots \\ g(X_{iT};\beta) \end{pmatrix}, \quad e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{pmatrix}$$

As a unifying framework, it is useful to put the various estimators of these models into a method of moments framework so that the models of interest can be seen as the solution to a set of k_θ moment or estimating equations such as $\psi_N(\widehat{\theta}) = 0$ where

$$\psi_N(\theta) = \frac{1}{N}\sum q_i(\theta) \tag{2.2}$$

where $q_i(\theta)$ is a $k_\theta \times 1$ vector, potentially depending on the k_β parameter of interest, β and perhaps additional parameters. ι_T denotes a $T \times 1$ vector of ones. This implies as many moments as parameters, with slight modification many if not most commonly used estimators can be put into this framework by judicious definition of the moments.

We can generalize this readily to situations with endogenous regressors and allow for instruments. We indicate below how this can be done. We do not allow here for cases is which the instruments may depend on an infinite dimensional parameter. The model can be generalized to allow for estimation of additional nuisance parameters such as in IV estimation or estimation of heteroskedasticity.

We only consider unobserved individual specific and not time-specific errors. These are small T models and asymptotics are as N grows large.

For the most part, we follow conventional practice and differentiate between fixed effects models which are characterized by non-independence (or at least non-orthogonality) between the unobserved and observed covariates and random effects models which are characterized by independence or at least orthogonality between the observed and unobserved covariates.

3. Stochastic Expansions and Approximate Moments for Nonlinear Models

Stochastic expansions and approximate moments for various nonlinear estimators commonly used in econometrics have been obtained by various authors such as Rilstone et al (1996) and Newey and Smith (2004). The class of models used in Rilstone et al can be adapted to the present case. In that paper stochastic expansions to third order are obtained for estimators of a parameter θ which solve a vector of k_θ moment conditions $\psi_N(\widehat{\theta}) = 0$ as in 2.1 with the identifying condition that $\mathbb{E}[q_i(\theta)] = 0$ only at $\theta = \theta_0$ where θ_0 is thus implicitly defined as the “true” value of the parameter.

The stochastic expansion in Rilstone et al (1996) is written

$$\widehat{\theta} - \theta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + o_p\left(N^{-3/2}\right) \tag{3.1}$$

where the $a_{-j/2}$ terms are j'th order matrix polynomials in random averages in mean zero random variables.

We use a number of conventions from matrix calculus which will (in as much as this is possible) streamline the presentation. As in Rilstone et al (1996) and also discussed in Turkington we use Kronecker matrix differentiation. One of the principal benefits of these is that it allows derivation of higher order Taylor series for vector-valued functions in a manner which is analogous to the univariate single function case. In the course of the presentation we will cite some results in this regard. From McCrae (1974), if X is $n \times m$ and A is $p \times q$, then we define

$$\bigtriangledown A(X) = A(X) \otimes \frac{d}{dX},$$

a $pn \times qm$ matrix with the same coordinate system as Kronecker products. Some standard results which we will use frequently are the following. If X is $m \times n$ then $\bigtriangledown_X AB = \bigtriangledown_X A(B \otimes I_n) + (A \otimes I_m) \bigtriangledown_X B$ and if A is $s \times t$ and B is $p \times q$ then

$$\bigtriangledown_X (A \otimes B) = A \otimes \bigtriangledown B + (K_{ps} \otimes I_m)(B \otimes \bigtriangledown_A A)(K_{tq} \otimes I_n)$$

where K_{ps} is the usual permutation matrix which is most expediently defined as that $ps \times ps$ matrix such that for any $p \times p$ matrix A, $\mathrm{Vec}[A] = K_{ps}\mathrm{Vec}[A^\top]$. In most of our calculations we have simplifications resulting from m and/or n being 1 in which case K_{ps} is the identity matrix. Higher order derivatives are defined recursively. We put $A^{(j)}(X) = \bigtriangledown^j_{X^\top} A(X)$. One useful feature of this approach is that for A and θ vectors we have

$$A(\theta) = \sum_{j=1}^{s} \frac{1}{j!} A^{(j)}(\theta_0)(\theta - \theta_0)^{\otimes j} + o(\|\theta - \theta_0\|^s)$$

We use the convention that when the argument of a function is omitted, the function is understood to be evaluated at its true value. Also, $\widetilde{A}_i = A_i - \mathbb{E}[A_i]$, $\overrightarrow{A} = \mathrm{Vec}[A]$,

In Rilstone (2015) it is shown that an equivalent way to write the terms in the expansion in equation 3.2 is by defining

$$\begin{aligned} a_{-1/2} &= \frac{1}{N} \sum_{i_1} d_{i_1} \\ a_{-1} &= \frac{1}{N^2} \sum_{i_1} \sum_{i_2} d_{i_1 i_2} \\ a_{-3/2} &= \frac{1}{N^3} \sum_{i_1} \sum_{i_2} \sum_{i_3} d_{i_1 i_2 i_3} \end{aligned}$$

where

$$\begin{aligned} d_{i_1} &= (-\mathbb{E}[q_i(\theta_0)])^{-1} q_{i_1} \\ d_{i_1 i_2} &= d^{(1)}_{i_1} d_{i_2} + \frac{1}{2} \overline{d^{(2)}_1} (d_{i_1} \otimes d_{i_2}) \\ d_{i_1 i_2 i_3} &= \frac{1}{2} \overline{d^{(2)}_1} (d_{i_1} \otimes d_{i_2 i_3} + d_{i_1 i_2} \otimes d_{i_3}) + \frac{1}{6} \overline{d^{(3)}_1} (d_{i_1} \otimes d_{i_2} \otimes d_{i_3}) + \widetilde{d^{(1)}_{i_1}} (d_{i_2 i_3}) + \frac{1}{2} \widetilde{d^{(2)}_{i_1}} (d_{i_2} \otimes d_{i_3}) \end{aligned}$$

The vector d_i is commonly referred to as the influence function for the estimator $\widehat{\theta}$ in that it describes the influence, to first order, of the data on the estimator. Under regularity conditions, we have $\sqrt{N}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, V_1)$ where $V_1 = \mathbb{E}[d_1 d_1^\top]$. There are several benefits to using this notation. First is that the expansions are all written in terms of the influence function and its derivatives. This simplifies the interpretation of the results. Second, as will be shown below, it simplifies calculation of the approximate moments, which in turn simplifies derivation of approximations to the distribution of the estimators, namely Edgeworth, Saddlepoint and related expansions.

We say that $A_m = A(\Lambda_1, \Lambda_2, \ldots, \Lambda_m)$ is the (generic) matrix product of Λ_j, $j = 1, \ldots, m$ if an arbitrary element of A_m, say a_m can be written as $a_m = \sum_{j=1}^{n} a_{mj}$ with $a_{mj} = \prod_{l=1}^{m} c_{j,l}^{\top} \overrightarrow{\Lambda_l}$.

The definitions to follow sum over sets of indices of the form $\{j_1, \ldots, j_m\}!$ which denote all unique combinations of the given indices. Specifically, put:

$$\sum_{\{1,1,2,2\}!} \mathbb{E}\left[U_{i_1 \cdots i_4}\right] = \mathbb{E}[U_{1122} + U_{1212} + U_{1221}]$$

$$\sum_{\{1,1,1,2,2\}!} \mathbb{E}[U_{i_1 \cdots i_5}] = \mathbb{E}\left[U_{11122} + U_{11212} + U_{11221} + U_{12112} + U_{12121} + U_{12211}\right] + \mathbb{E}\left[U_{11222} + U_{12122} + U_{12212} + U_{12221}\right]$$

$$\begin{aligned}\sum_{\{1,1,2,2,3,3\}!} \mathbb{E}[U_{i_1 \cdots i_6}] = {} & \mathbb{E}[U_{112233} + U_{112323} + U_{112332}] + \mathbb{E}[U_{121233} + U_{121323} + U_{121332}] \\ & + \mathbb{E}[U_{122133} + U_{123123} + U_{123132}] + \mathbb{E}[U_{123213} + U_{122313} + U_{123312}] \\ & + \mathbb{E}[U_{123231} + U_{123321} + U_{122331}]\end{aligned}$$

Proposition: Let $A_m = A(\Lambda_1, \Lambda_2, \ldots, \Lambda_m)$ denote the generic matrix product of $\Lambda_j = \frac{1}{N}\sum \Lambda_{ji}$, $\Lambda_j = \frac{1}{N}\sum \Lambda_{ji}$, $j = 1, \ldots, m$ and $U_{i_1 \cdots i_m} = A(\Lambda_{1i_1}, \ldots, \Lambda_{mi_m})$. Assume $\mathbb{E}[\Lambda_{j1}] = 0$, $\mathbb{E}\|\Lambda_{j1}\|^m < \infty$, $j = 1, \ldots, m$. Then,

$$\mathbb{E}[A_m] = \begin{cases} 0 & m = 1 \\ \frac{1}{N^{m-1}}\mathbb{E}[U_{1 \cdots 1}] & m = 2, 3 \\ \frac{1}{N^2}\sum_{\{1,1,2,2\}!} \mathbb{E}\left[U_{i_1 \cdots i_4}\right] + O(N^{-3}) & m = 4 \\ \frac{1}{N^3}\sum_{\{1,1,1,2,2\}!} \mathbb{E}\left[U_{i_1 \cdots i_5}\right] + O(N^{-4}) & m = 5 \\ \frac{1}{N^3}\sum_{\{1,1,2,2,3,3\}!} \mathbb{E}[U_{i_1 \cdots i_6}] + O(N^{-4}) & m = 6. \end{cases}$$

Proof: See Rilstone (2015).

In the discussion below we use this result to derive approximate moments of $\widehat{\theta}$ setting the $U.$ terms into various products of the $d.$ terms. These moments can then be used to derive Edgeworth and Saddlepoint approximations to the distribution of the estimators.

4. Additive Fixed Effects Models

Estimation of linear and nonlinear models with fixed effects is often conducted with a "within" estimator such that the data is transformed by differencing individual sample averages.

Slightly more generally, we could transform the data using Arellano's transformation matrix. However, given that we have no dynamic component here, there is no need for that additional generality.

Each of the estimators we consider for the additive case has the same formal structure with some caveats. In particular, each can be seen as solving, perhapsin conjunction with additional parameters, of estimating equations of the form:

$$\psi_N(\beta; W) = \frac{1}{N} \sum q_{Wi}(\beta)$$

$$q_{Wi}(\beta) = G_i^\top(\beta) W u_i(\beta)$$

where $G_i(\beta) = \bigtriangledown_\beta g_i(\beta)$ is the $T \times k_\beta$ Jacobian of $g_i(\beta)$. The least squares estimator sets $W = I_T$, fixed effects sets W equal to the usual mean-differencing matrix, defined below, and random effects sets W equal to the covariance matrix of the disturbances, perhaps estimated.

These kind of estimating equations also correspond to first order equations form estimators minimizing, say, $S_N(\beta; W) = \sum u_i(\beta)^\top W u_i(\beta)$.

As such it is useful to have a generic representation for the matrix derivatives of q_{Wi}.

The fixed effects estimators we consider then can be seen as solving first order conditions of the form above. with

$$q_i(\beta) = G_i^\top(\beta) Q u_i(\beta)$$

$$Q = I_T - P, \qquad P = \iota_T(\iota_T^\top \iota_T)^{-1} \iota_T^\top$$

Defined this way $\psi_N(\beta)$ represents generalized estimating equations, say either from minimizing squared residuals with Q as a weighting matrix or similarly the partial score from a ML problem. The usual caveats for estimation apply here implicitly here such the elements of G_i cannot contain any constants. (The model is thus implicitly written so that intercepts are subsumed into the v_i's.)

We could allow for more general forms of models by defining G differently and perhaps augmenting the parameter space. This will not change the form of the results and we find the current notation more transparent. We assume that $\mathbb{E}[q_i(\beta) = 0]$ only at $\beta = \beta_0$ and that $\mathbb{E}[q_i(\beta_0) q_i(\beta_0)^\top] = \Sigma$ is positive definite, that $\mathbb{E}[q_i^{(1)}(\beta)]$ is non-singular in a neighbourhood of β_0 and $\widehat{\beta} - \beta_0 = o_p(1)$.

$$\sqrt{N}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, V_1)$$

$$V_1 = \mathbb{E}[d_1 d_1^\top]$$

$$d_i = -\left(\mathbb{E}[q_1^{(1)}(\beta_0)]\right)^{-1} q_i(\beta_0)$$

The expansions for many of the estimators share a common element in that each depend on derivatives of quadratics in $G_i(\beta)$ such as

$$V_{Wi}(\beta) = G_i(\beta)^\top W G_i(\beta) \tag{4.1}$$

We will make repeated use of terms such as such as $\overline{V}_W = \mathbb{E}[V_{Wi}]$

The estimating equations we use can be written in whole or in part as

$$q_{Wi}(\beta) = G_i^\top(\beta) W u_i(\beta)$$

where W is a weighting matrix.

The following is thus useful

$$\begin{aligned}
V_{Wi}^{(1)} &= G_i^\top W G_i^{(1)} + G_i^{\top(1)}(WG_i \otimes I_k) \\
V_{Wi}^{(2)} &= G_i^\top W G_i^{(2)} + G_i^{\top(1)}(WG_i^{(1)} \otimes I_k) \\
&\quad + G_i^{\top(1)} K_{Tk}(I_k \otimes (WG_i^{(1)})(K_{kk} \otimes I_k) + G_i^{\top(2)}(WG_i \otimes I_{k^2})
\end{aligned}$$

Lemma 4.1. Let $g_i(\beta)$ have continuous third derivatives with finite moments. Then

$$\begin{aligned}
q_{Wi}^{(1)} &= -G_i^\top W G_i + G_i^{\top(1)}((Wu_i) \otimes I_k) \\
\mathbb{E}[q_{Wi}^{(1)}] &= -\mathbb{E}[G_i^\top W G_i]
\end{aligned}$$

$$\begin{aligned}
q_{Wi}^{(2)} &= -V_{Wi}^{(1)} - G_i^{\top(1)} K_{kT}(I_k \otimes (WG_i)) + G_i^{\top(2)}((Wu_i) \otimes I_{k^2}) \\
\mathbb{E}[q_{Wi}^{(2)}] &= -\mathbb{E}\left[V_{Wi}^{(1)}\right] - \mathbb{E}\left[G_i^{\top(1)} K_{kT}(I_k \otimes (WG_i))\right]
\end{aligned}$$

$$\begin{aligned}
q_{Wi}^{(3)} &= -V_{Wi}^{(2)} \\
&\quad - G_i^{\top(1)} K_{kT}(I_k \otimes (WG_i^{(1)}) - G_i^{\top(2)}\,((K_{kT}(I_k \otimes (WG_i) \otimes I_k) \\
&\quad + G_i^{\top(2)}((Wu_i \otimes I_{k^2}) - G_i^{\top(3)} K_{k^3T}(I_{k^3} \otimes WG_i) \\
\mathbb{E}[q_{Wi}^{(3)}] &= -\mathbb{E}[V_{Wi}^{(2)}] \\
&\quad - \mathbb{E}[G_i^{\top(1)} K_{kT}(I_k \otimes (WG_i^{(1)})] - \mathbb{E}[G_i^{\top(2)}\,((K_{kT}(I_k \otimes (WG_i) \otimes I_k)] \\
&\quad - \mathbb{E}[G_i^{\top(3)} K_{k^3T}(I_{k^3} \otimes WG_i)]
\end{aligned}$$

In the case $k = 1$, it may be clarifying to note that

$$\begin{aligned}
q_i^{(1)} &= -G_i^\top W G_i + G_i^{\top(1)} W u_i \\
\mathbb{E}[q_i^{(1)}] &= -\mathbb{E}[G_i^\top W G_i]
\end{aligned}$$

$$\begin{aligned}
q_i^{(2)} &= -3G_i^\top W G_i^{(1)} + G_i^{\top(2)} W u_i \\
\mathbb{E}[q_i^{(2)}] &= -3\mathbb{E}[G_i^\top W G_i^{(1)}]
\end{aligned}$$

$$\begin{aligned}
q_i^{(3)} &= -3G_i^\top W G_i^{(2)} - 3G_i^{\top(1)} W G_i^{(1)} + G_i^{\top(2)} W u_i - G_i^{\top(3)} W G_i \\
\mathbb{E}[q_i^{(3)}] &= -3\mathbb{E}[G_i^\top W G_i^{(2)}] - 3\mathbb{E}[G_i^{\top(1)} W G_i^{(1)}] - \mathbb{E}[G_i^{\top(3)} W G_i]
\end{aligned}$$

For the fixed effects case we thus see that the influence function is given by

$$d_i = (\mathbb{E}[G_i^\top Q G_i])^{-1} G_i^\top Q u_i = \overline{V}_Q^{-1} G_i Q u_i$$

with the first-order covariance matrix given by $V_1 = \mathbb{E}[d_1 d_1^\top] = \sigma_e^2 \overline{V}_Q^{-1}$ The rest of the terms needed for second-order analysis follow by substutiting Q for W into the expressions for the $q_{Wi}^{(j)}$ and in turn into the $d.$ terms.

The second and third order influence functions given by

$$
\begin{aligned}
d_{i_1 i_2} &= \overline{V}_Q^{-1}\left(-G_{i_1}^\top Q G_{i_1} + G_{i_1}^{\top(1)}((Qu_{i_1})\otimes I_k)\right) d_{i_2} \\
&\quad + \frac{1}{2}\overline{V}_Q^{-1}\left(-\mathbb{E}\left[V_{Qi}^{(1)}\right] - \mathbb{E}\left[G_{i_1}^{\top(1)} K_{kT}(I_k \otimes (QG_i))\right]\right)(d_{i_1}\otimes d_{i_2}) \\
&= \overline{V}_Q^{-1}\left(-G_{i_1}^\top Q G_{i_1} + G_{i}^{\top(1)}((Qu_{i_1})\otimes I_k)\right) d_{i_2} + \frac{1}{2}\overline{V}_Q^{-1}\bar{H}\,(d_{i_1}\otimes d_{i_2})
\end{aligned}
$$

$$
\bar{H} = \left(-\mathbb{E}\left[V_{Qi}^{(1)}\right] - \mathbb{E}\left[G_{i1}^{\top(1)} K_{kT}(I_k \otimes (QG_i))\right]\right)
$$

$$
d_{i_1 i_2 i_3} = \frac{1}{2}\overline{d_1^{(2)}}\,(d_{i_1}\otimes d_{i_2 i_3} + d_{i_1 i_2}\otimes d_{i_3}) + \frac{1}{6}\overline{d_1^{(3)}}\,(d_{i_1}\otimes d_{i_2}\otimes d_{i_3}) + \widetilde{d_{i_1}^{(1)}}\,(d_{i_2 i_3}) + \frac{1}{2}\widetilde{d_{i_1}^{(2)}}\,(d_{i_2}\otimes d_{i_3})
$$

Following Rilstone et al (1996) we derive the first order bias as follow. First note that

$$
\begin{aligned}
((Qu_i)\otimes I_k)\overline{V}_Q^{-1} G_i^\top (Qu_i) &= (Qu_i)\otimes \overline{V}_Q^{-1} G_i^\top (Qu_i) \\
&= \left(I_T \otimes \overline{V}_Q^{-1} G_i^\top\right)((Qu_i)\otimes (Qu_i))
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left((Qu_i)\otimes (Qu_i)\right) &= \mathrm{Vec}[Q\mathbb{E}[e_i e_i^\top]Q] \\
&= \sigma_e^2 \mathrm{Vec}[Q]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[G_i^{\top(1)}((Qu_i)\otimes I_k)\overline{V}_Q^{-1} G_i^\top (Qu_i)] &= \sigma_e^2 \mathbb{E}[G_i^{\top(1)}\left(I_T \otimes \overline{V}_Q^{-1} G_i^\top\right)]\mathrm{Vec}[Q] \\
&= \sigma_e^2 \mathbb{E}[G_i^{\top(1)} \mathrm{Vec}[\overline{V}_Q^{-1} G_i^\top Q]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[V_{Qi}^{(1)}\right]\overrightarrow{V_1} &= \mathbb{E}\left[G_i^\top Q G_i^{(1)} + G_i^{\top(1)}(QG_i \otimes I_k)\right]\overrightarrow{V_1} \\
&= \mathbb{E}\left[G_i^\top Q G_i^{(1)}\right]\overrightarrow{V_1} + \mathbb{E}\left[G_i^{\top(1)}\overrightarrow{V_1 G_i^\top Q}\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[G_i^{\top(1)} K_{kT}(I_k \otimes (QG_i))\right]\overrightarrow{V_1} &= \mathbb{E}\left[G_i^{\top(1)} K_{kT}\overrightarrow{(QG_i)V_1}\right] \\
&= \mathbb{E}\left[G_i^{\top(1)}\overrightarrow{V_1 G_i^\top Q}\right]
\end{aligned}
$$

$$
\begin{aligned}
\text{ABIAS}[\widehat{\beta}] &= \frac{1}{N}\mathbb{E}[d_{11}] \\
&= \frac{1}{N}\left(\overline{V}_Q^{-1}\mathbb{E}[G_i^{\top(1)}((Qu_i)\otimes I_k)\overline{V}_Q^{-1}G_i^{\top}(Qu_i)]\right) \\
&\quad + \frac{1}{2N}\overline{V}_Q^{-1}\left(-\mathbb{E}\left[V_{Qi}^{(1)}\right] - \mathbb{E}\left[G_i^{\top(1)}K_{kT}(I_k\otimes(QG_i))\right]\right)\overrightarrow{V_1} \\
&= \frac{1}{N}\left(\overline{V}_Q^{-1}\sigma_e^2\mathbb{E}[G_i^{\top(1)}\text{Vec}[\overline{V}_Q^{-1}G_i^{\top}Q]\right) \\
&\quad - \frac{1}{2N}\overline{V}_Q^{-1}\left(\mathbb{E}\left[G_i^{\top}QG_i^{(1)}\right]\overrightarrow{V_1} + \mathbb{E}\left[G_i^{\top(1)}\overrightarrow{V_1G_i^{\top}Q}\right]\right) \\
&\quad - \frac{1}{2N}\overline{V}_Q^{-1}\left(\mathbb{E}\left[G_i^{\top(1)}K_{kT}\overrightarrow{(QG_i)V_1}\right]\right) \\
&= \frac{1}{N}\left(\overline{V}_Q^{-1}\sigma_e^2\mathbb{E}[G_i^{\top(1)}\text{Vec}[\overline{V}_Q^{-1}G_i^{\top}Q]\right) \\
&\quad - \frac{1}{2N}\overline{V}_Q^{-1}\left(\mathbb{E}\left[G_i^{\top}QG_i^{(1)}\right]\overrightarrow{V_1}\right) - \frac{1}{N}\overline{V}_Q^{-1}\mathbb{E}\left[G_i^{\top(1)}\overrightarrow{V_1G_i^{\top}Q}\right] \\
&= -\frac{1}{2N}\overline{V}_Q^{-1}\left(\mathbb{E}\left[G_i^{\top}QG_i^{(1)}\right]\overrightarrow{V_1}\right) \\
&= -\sigma_e^2\frac{1}{2N}\overline{V}_Q^{-1}\left(\mathbb{E}\left[G_i^{\top}QG_i^{(1)}\right]\overrightarrow{\overline{V}_Q^{-1}}\right)
\end{aligned}
$$

The second-order Mean Squared Error (AMSE) is derived as follows. Put

$$V_2 = \left(V_{21} + V_{21}^{\top}\right) + \left(V_{22} + V_{22}^{\top}\right) + V_{23}$$

$$V_{21} = \mathbb{E}\left[d_1 d_{11}^{\top}\right], \qquad V_{22} = \sum_{\{1,1,2,2\}!}\mathbb{E}\left[d_{i_1}d_{i_2i_3i_4}^{\top}\right] \qquad V_{23} = \sum_{\{1,1,2,2\}!}\mathbb{E}\left[d_{i_1i_2}d_{i_3i_4}^{\top}\right]$$

and

$$\text{AMSE}\left[\widehat{\theta}\right] = \frac{1}{N}V_1 + \frac{1}{N^2}V_2$$

From symmetry,

$$
\begin{aligned}
\mathbb{E}\left[d_{11}d_1^{\top}\right] &= \left(\overline{V}_Q^{-1}\left(-G_1^{\top}QG_1 + G_1^{\top(1)}((Qu_1)\otimes I_k)\right)d_1d_1^{\top}\right) \\
&\quad + \overline{V}_Q^{-1}\left(-\mathbb{E}\left[V_{Q1}^{(1)}\right] - \mathbb{E}\left[G_1^{\top(1)}K_{kT}(I_k\otimes(QG_1))\right]\right)d_1^{\otimes 2}d_1^{\top} \\
&= -\overline{V}_Q^{-1}\mathbb{E}\left[\left(G_1^{\top}QG_1\right)d_1d_1^{\top}\right] \\
&= -\overline{V}_Q^{-1}\mathbb{E}\left[G_1^{\top}QG_1\overline{V}_Q^{-1}G_1^{\top}QG_1\right]\overline{V}_Q^{-1}
\end{aligned}
$$

The expressions for V_{22} and V_{23} are given in the Appendix where we note $\mathbb{E}[d_{11}d_{22}^{\top}] = N\,\text{ABIAS}[\widehat{\beta}]\text{ABIAS}[\widehat{\beta}]^{\top}$

5. Additive Random Effects Models

Referring to the notation in equation (2.3) we make standard assumptions regarding the disturbances. We adapt the standard convention of assuming that the individual effects are uncorrelated with each other and independent of the regressors. (This could be relaxed to uncorrelated with the regressors but would needlessly complicate the presentation.)

Thus we have $\mathbb{E}[v_i] = \mathbb{E}[e_{it}] = \mathbb{E}[v_i e_{it}] = 0$, $\mathbb{E}[v_i^2] = \sigma_v^2$, $\mathbb{E}[e_{it}^2] = 0$, for all i, t and $\mathbb{E}[e_{it}e_{is}] = 0$, $s \neq t$. As a result we have

$$\mathbb{E}[(v_i\iota_T + e_i)(v_i\iota_T + e_i)^\top] \equiv \Omega = \sigma_v^2\iota_T\iota_T^\top + \sigma_e^2 I_T$$

For completeness we state the following regarding the usual least squares estimator of $\widehat{\beta}$. The inefficient estimator solves $\psi_N(\widehat{\beta}) = 0$ where

$$q_i(\beta) = G_i^\top(\beta)u_i(\beta)$$

where in the nonlinear regression case $G_i(\beta) = -\bigtriangledown u_i(\beta)$ and in the linear regression case

$$q_i(\beta) = X_i^\top(y_i - X_i\beta)$$

with

$$q_i^{(1)}(\beta) = -G_i^\top(\beta)G_i(\beta) + G_i^{\top(1)}(\beta)(u_i(\beta) \otimes I_{k_\beta})$$

in the linear regression case

$$q_i^{(1)}(\beta) = -X_i^\top X_i$$

Here the influence function is

$$d_i = -(\mathbb{E}[G_1^\top G_1])^{-1}G_i u_i$$

and in the linear regression case

$$d_i = -(\mathbb{E}[X_i^\top X_i])^{-1}X_i^\top(y_i - X_i\beta)$$

and

$$\sqrt{N}(\widehat{\beta} - \beta_0) \to N(0, V_1)$$

$$V_1 = (\mathbb{E}[G_1^\top G_1])^{-1}\mathbb{E}[G_1^\top\Omega G_1](\mathbb{E}[G_1^\top G_1])^{-1}$$

and in the linear regression case

$$V_1 = (\mathbb{E}[X_1^\top X_1])^{-1}\mathbb{E}[X_1^\top\Omega X_1](\mathbb{E}[X_1^\top X_1])^{-1}$$

The GLS estimator solves $\psi_N(\widehat{\beta}) = 0$ where

$$q_i(\theta) = G_i^\top(\beta)\Omega^{-1}u_i(\beta) \tag{3.1a}$$

and in the linear regression case

$$q_i(\theta) = X_i^\top\Omega^{-1}(y_i - X_i\beta) \tag{3.1b}$$

is of course efficient, but infeasible. There are various feasible GLS estimators available, largely differing in the method of estimation of σ_v^2 and σ_e^2 and whether they are one step or two step estimators. Often, when estimators are iterated upon they are higher order equivalent.

A straightforward efficient method of moments estimator combines the estimating equation in (3.1) with the observation that $\mathbb{E}[u_i(\beta_0)^\top u_i(\beta_0)] = T(\sigma_v^2 + \sigma_e^2)$ and that by mean differencing the observations we have $\mathbb{E}[u_i(\beta_0)^\top Q u_i(\beta_0)] = T\sigma_e^2$ where Q as defined above is the usual mean differencing matrix used in panel models. We implicitly and repeatedly use the relations $QQ = Q$, $PP = P$ and $QP = 0$. We suppose that the estimator jointly solves moment equations consistent with the assumptions. In this case we stack estimating equations, defining:

$$q_i(\theta) = \begin{pmatrix} q_{1i}(\theta) \\ q_{2i}(\theta) \\ q_{3i}(\theta) \end{pmatrix} = \begin{pmatrix} G_i^\top(\beta)\Omega^{-1}u_i(\beta) \\ u_i(\beta)^\top u_i(\beta) - T(\sigma_v^2 + \sigma_e^2) \\ u_i(\beta)^\top Q u_i(\beta) - T\sigma_e^2 \end{pmatrix}$$

and in the linear regression case

$$q_i(\theta) = \begin{pmatrix} q_{1i}(\theta) \\ q_{2i}(\theta) \\ q_{3i}(\theta) \end{pmatrix} = \begin{pmatrix} X_i^\top \Omega^{-1}(y_i - X_i\beta) \\ (y_i - X_i\beta)^\top (y_i - X_i\beta) - T(\sigma_v^2 + \sigma_e^2) \\ (y_i - X_i\beta)^\top Q (y_i - X_i\beta) - T\sigma_e^2 \end{pmatrix}$$

$$\theta = \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} \sigma_v^2 \\ \sigma_e^2 \end{pmatrix}$$

Putting $u_i = v_i\iota_T + e_i$, $\sigma_\eta^2 = T\sigma_v^2 + \sigma_e^2$, we note the standard result that $\Omega = \sigma_\eta^2 P + \sigma_e^2 Q$ and $\Omega^{-1} = \frac{1}{\sigma_\eta^2}P + \frac{1}{\sigma_e^2}Q$ which allows for simple derivation of a few results. In this regard it is useful to denote

$$\begin{aligned} \Omega_v^{-1} &= \frac{\partial\Omega^{-1}}{\partial\sigma_v^2} = -P\frac{T}{(\sigma_v^2)^2} \\ \Omega_{vv}^{-1} &= \frac{\partial^2\Omega^{-1}}{\partial^2\sigma_v^2} = P\frac{T^2}{(\sigma_\eta^2)^3} \\ \Omega_e^{-1} &= \frac{\partial\Omega^{-1}}{\partial\sigma_e^2} = -P\frac{1}{(\sigma_\eta^2)^2} - Q\frac{1}{(\sigma_e^2)^2} \\ \Omega_{ee}^{-1} &= \frac{\partial^2\Omega^{-1}}{\partial\sigma_e^2} = P\frac{1}{(\sigma_\eta^2)^3} + Q\frac{1}{(\sigma_e^2)^3} \\ \Omega_{ve}^{-1} &= \frac{\partial^2\Omega^{-1}}{\partial\sigma_v^2\partial\sigma_e^2} = P\frac{T}{(\sigma_\eta^2)^3} \end{aligned}$$

Also put $V_{G\Omega} = \mathbb{E}[G_1^\top\Omega^{-1}G_1]$, $V_{Gv} = \mathbb{E}[G_1^\top\Omega_v^{-1}G_1]$, $V_{Ge} = \mathbb{E}[G_1^\top\Omega_e^{-1}G_1]$, $V_{Wvv} = \mathbb{E}[G_1^\top\Omega_{vv}^{-1}\rho_1^{(1)}]$ $V_{X\Omega} = \mathbb{E}[X_1^\top\Omega^{-1}X_1]$, $V_{Xv} = \mathbb{E}[X_1^\top\Omega_v^{-1}X_1]$, $V_{Xe} = \mathbb{E}[X_1^\top\Omega_e^{-1}X_1]$, $V_{Xvv} = \mathbb{E}[X_1^\top\Omega_{vv}^{-1}X_1]$

To derive the approximate moments we need to first derive the matrix derivatives of q_i and their mean values. In this regard we have the following. In some cases it will be useful to partition these derivatives in an obvious manner. The steps are included in the Appendix.

$$q_i^{(1)} = \begin{pmatrix} q_{1i}^{(\beta)} & q_{1i}^{(\sigma_v^2)} & q_{1i}^{(\sigma_e^2)} \\ q_{2i}^{(\beta)} & q_{2i}^{(\sigma_v^2)} & q_{2i}^{(\sigma_e^2)} \\ q_{3i}^{(\beta)} & q_{3i}^{(\sigma_v^2)} & q_{3i}^{(\sigma_e^2)} \end{pmatrix}$$

$$q_i^{(1)} = \begin{pmatrix} -G_i^\top(\beta)\Omega^{-1}G_i(\beta) + G_i^{\top(1)}(\beta)(\Omega^{-1}u_i(\beta) \otimes I_{k_\beta}) & G_i^\top(\beta)\Omega_v^{-1}u_i & G_i^\top(\beta)\Omega_e^{-1}u_i \\ 2u_i^\top G_i(\beta) & -T & -T \\ 2u_i^\top Q G_i(\beta) & 0 & -T \end{pmatrix}$$

$$\mathbb{E}\left[q_i^{(1)}\right] = \begin{pmatrix} -\mathbb{E}\left[G_1^\top\Omega^{-1}G_1\right] & 0 & 0 \\ 0 & -T & -T \\ 0 & 0 & -T \end{pmatrix}$$

With linearity

$$q_i^{(1)} = \begin{pmatrix} -X_i^\top\Omega^{-1}X_i & X_i^\top\Omega_v^{-1}u_i & X_i^\top\Omega_e^{-1}u_i \\ -2u_i^\top X_i & -T & -T \\ -2u_i^\top Q X_i & 0 & -T \end{pmatrix}$$

$$\mathbb{E}\left[q_i^{(1)}\right] = -\begin{pmatrix} V_{X\Omega} & 0 & 0 \\ 0 & T & T \\ 0 & 0 & T \end{pmatrix}$$

$$\left(\mathbb{E}\left[q_i^{(1)}\right]\right)^{-1} = -\begin{pmatrix} V_{X\Omega}^{-1} & 0 & 0 \\ 0 & T^{-1} & -T^{-1} \\ 0 & 0 & T^{-1} \end{pmatrix}$$

The influence function is given by

$$\begin{aligned} d_i &= \left(\mathbb{E}[q_i^{(1)}]\right)^{-1} q_i \\ &= -\begin{pmatrix} -V_{G\Omega}^{-1} & 0 & 0 \\ 0 & T^{-1} & -T^{-1} \\ 0 & 0 & T^{-1} \end{pmatrix} \begin{pmatrix} G_i^\top\Omega^{-1}u_i \\ u_i^\top u_i - T(\sigma_v^2+\sigma_e^2) \\ e_i^\top e_i - T\sigma_e^2 \end{pmatrix} \\ &= -\begin{pmatrix} V_{G\Omega}^{-1}X_i^\top\Omega^{-1}u_i \\ T^{-1}(u_i^\top u_i - e_i^\top e_i) - \sigma_v^2 \\ T^{-1}e_i^\top e_i - \sigma_e^2 \end{pmatrix} \end{aligned}$$

With linearity

$$d_i = -\begin{pmatrix} V_{X\Omega}^{-1}X_i^\top\Omega^{-1}u_i \\ T^{-1}(u_i^\top u_i - e_i^\top e_i) - \sigma_v^2 \\ T^{-1}e_i^\top e_i - \sigma_e^2 \end{pmatrix}$$

$$\mathbb{E}[q_i q_i^\top] = \begin{pmatrix} \mathbb{E}\left[G_1^\top \Omega^{-1} G_1\right] & 0 & 0 \\ 0 & \mathrm{Var}[u_i^\top u_i] & 0 \\ 0 & 0 & \mathrm{Var}[e_i^\top e_i] \end{pmatrix}$$

$$\mathbb{E}[q_i q_i^\top] = \begin{pmatrix} V_{X\Omega} & 0 & 0 \\ 0 & \mathrm{Var}[u_i^\top u_i] & 0 \\ 0 & 0 & \mathrm{Var}[e_i^\top e_i] \end{pmatrix}$$

Expressions for $\mathrm{Var}[u_i^\top u_i]$ and $\mathrm{Var}[e_i^\top e_i]$ are given in the Appendix. The asymptotic variance of $\widehat{\theta}$ is given by

$$V_1 = \begin{pmatrix} V_{1\beta} & 0 & 0 \\ 0 & T^{-2}\mathrm{Var}[u_i^\top u_i] & -T^{-2}\mathrm{Var}[e_i^\top e_i] \\ 0 & -T^{-2}\mathrm{Var}[e_i^\top e_i] & T^{-2}\mathrm{Var}[e_i^\top e_i] \end{pmatrix}$$

where $V_{1\beta} = \mathbb{E}\left[G_1^\top \Omega^{-1} G_1\right]^{-1} \mathbb{E}\left[G_1^\top \Omega^{-1} G_1\right] \mathbb{E}\left[G_1^{(1)\top} \Omega^{-1} W_1^\top\right]^{-1}$. For regression models $V_{1\beta} = \mathbb{E}\left[\rho_1^{(1)\top} \Omega^{-1} \rho_1^{(1)}\right]^{-1}$ and for linear regression: $V_{1\beta} = \mathbb{E}\left[X_1^\top \Omega^{-1} X_1\right]^{-1}$. These latter are known as the usual variance of efficient estimators in these cases.

$$q_i^{(2)} = \begin{pmatrix} q_{1i}^{(\beta\theta)} & q_{1i}^{(\sigma_v^2\theta)} & q_{1i}^{(\sigma_e^2\theta)} \\ & q_{2i}^{(2)} & \\ & q_{3i}^{(2)} & \end{pmatrix}$$

The following selection matrices[1] will be useful.

$$\begin{aligned} S_{2j} &= I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \iota_2^{j\top} \end{pmatrix} \\ S_{4j} &= I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta^2+4k_\beta} & \iota_4^{j\top} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} q_{1i}^{(\beta\theta)} &= -X_i^\top \Omega_v^{-1} X_i S_{21} - X_i^\top \Omega_e^{-1} X_i S_{22} \\ q_{1i}^{(\sigma_v^2\theta)} &= \begin{pmatrix} -X_i^\top \Omega_v^{-1} X_i & 2X_i^\top \Omega_{vv}^{-1} u_i & 2X_i^\top \Omega_{ve}^{-1} u_i \end{pmatrix} \\ q_{1i}^{(\sigma_e^2\theta)} &= \begin{pmatrix} -X_i^\top \Omega_e^{-1} X_i & 2X_i^\top \Omega_{ve}^{-1} u_i & 2X_i^\top \Omega_{ee}^{-1} u_i \end{pmatrix} \\ q_{2i}^{(2)}(\theta) &= \begin{pmatrix} 2\mathrm{Vec}[X_i^\top X_i]^\top & 0_{1\times k_\theta^2-\theta_\beta^2} \end{pmatrix} \\ q_{3i}^{(2)}(\theta) &= \begin{pmatrix} 2\mathrm{Vec}[X_i^\top Q X_i]^\top & 0_{1\times k_\theta^2-\theta_\beta^2} \end{pmatrix} \end{aligned}$$

$$\mathbb{E}[q_i^{(2)}] = \begin{pmatrix} -V_{Xv}S_{21} - V_{Xe}S_{22} & \begin{pmatrix} -V_{Xv} & 0_{k_\beta\times 2} \end{pmatrix} & \begin{pmatrix} -V_{Xe} & 0_{k_\beta\times 2} \end{pmatrix} \\ & \begin{pmatrix} 2\mathrm{Vec}[V_X]^\top & 0_{1\times k_\theta^2-\theta_\beta^2} \end{pmatrix} & \\ & \begin{pmatrix} 2\mathrm{Vec}[V_{XQ}]^\top & 0_{1\times k_\theta^2-\theta_\beta^2} \end{pmatrix} & \end{pmatrix}$$

[1] In this preliminary version, the remaining derivations in this section are for the linear regression model. Nonlinearity introduces additional terms in the top left sub-matrices of partial derivatives.

The derivation of the approximate moments of the estimators can be done in a variety of ways. Our interest is $\widehat{\beta}-\beta_0$. To derive the first two moments we will do this by examining the elements of the first k_β elemnts of $\widehat{\theta}$. We will be making substantial use of selection matrices to so. For the singling out $\widehat{\beta}$ here define $\tau^\top = \begin{pmatrix} I_{k_\beta} & 0_2 \end{pmatrix}$. We see that

$$\tau^\tau(\mathbb{E}q_i^{(1)}(\theta_0))^{-1} = -\left((V_{X\Omega}^{-1} \quad 0 \quad 0\right)$$

and that $\tau_i^{(j)} = \tau^\tau d_i^{(j)} \equiv d_{1i}^{(j)}$ returns the first k_β rows of $d_i^{(j)}$ and $\tau^\tau d_{i_1i_2} = d_{1,i_1i_2}$ and $\tau_{i_1i_2i_3} = \tau^\tau d_{i_1i_2i_3} = d_{1,i_1i_2i_3}$ are the first k_β rows of $d_{i_1i_2}$ and $d_{i_1i_2i_3}$

The approximate bias of $\widehat{\beta}$ is given by

$$\begin{aligned}
\text{ABIAS}\left[\widehat{\beta}\right] &= \mathbb{E}[\tau_{11}] \\
&= \mathbb{E}[\tau_1^{(1)}d_1] + \frac{1}{2}\overline{\tau_1^{(2)}}\mathbb{E}\left(d_1 \otimes d_1\right)
\end{aligned}$$

Note that the models we are examining are all first-order non-biased. By inspection there are a few sources of bias. For linear models, the second term is zero, although for nonlinear models it will in general not be. The first term will be potentially nonzero either for linear models which have some form of endogeneity and/or nonlinear models such that variability in the estimation of the instrument correlates with the errors in the model.

In the linear case we see

$$\begin{aligned}
\tau_i^{(1)} &= \tau^\top d_i^{(1)} \\
&= -V_{X\Omega}^{-1}\begin{pmatrix} -X_i^\top \Omega^{-1} X_i & X_i^\top \Omega_v^{-1} u_i & X_i^\top \Omega_e^{-1} u_i \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\tau_i^{(1)} d_i &= V_{X\Omega}^{-1}\begin{pmatrix} -X_i^\top \Omega^{-1} X_i & X_i^\top \Omega_v^{-1} u_i & X_i^\top \Omega_e^{-1} u_i \end{pmatrix}\begin{pmatrix} V_{X\Omega}^{-1} X_i^\top \Omega^{-1} u_i \\ T^{-1}(u_i^\top u_i - e_i^\top e_i) - \sigma_v^2 \\ T^{-1} e_i^\top e_i - \sigma_e^2 \end{pmatrix} \\
&= -V_{X\Omega}^{-1} X_i^\top \Omega^{-1} X_i V_{X\Omega}^{-1} X_i^\top \Omega^{-1} u_i + V_{X\Omega}^{-1} X_i^\top \Omega_v^{-1} u_i (T^{-1}(u_i^\top u_i - e_i^\top e_i) - \sigma_v^2) \\
&\quad + V_{X\Omega}^{-1} X_i^\top \Omega_e^{-1} u_i (T^{-1} e_i^\top e_i - \sigma_e^2)
\end{aligned}$$

so that in the case of linear regression by inspection $\mathbb{E}[q_{1i}^{(2)}]\text{Vec}[V_1] = 0$ and $\mathbb{E}[\tau_i^{(1)} d_i] = 0$ if the errors have a symmetric distribution. If not the expectation of either of the last two terms can be non-zero. Note that the models we are examining are all first-order non-biased. For the general case the first term will be potentially nonzero.

$$\mathbb{E}[q_{1i}^{(3)}] = \begin{pmatrix} \mathbb{E}[q_{1i}^{(\beta\theta\theta)}] \quad \mathbb{E}[q_{1i}^{(\sigma_v^2\theta\theta)}] \quad \mathbb{E}[q_{1i}^{(\sigma_e^2\theta\theta)}] \\ 0 \\ 0 \end{pmatrix}$$

$$\mathbb{E}[q_{1i}^{(\beta\theta\theta)}] = -V_{Xvv}S_{41} - V_{Xve}S_{43} - V_{Xve}S_{42} - V_{Xee}S_{44}$$

$$\mathbb{E}[q_{1i}^{(\sigma_v^2\theta\theta)}] = \begin{pmatrix} -V_{X_v v}S_{21} - V_{Xve}S_{22} & \begin{pmatrix} -2V_{Xvv} & 0_{k_\beta \times 2} \end{pmatrix} & \begin{pmatrix} -2V_{Xve} & 0_{k_\beta \times 2} \end{pmatrix} \end{pmatrix}$$

$$\mathbb{E}[q_{1i}^{(\sigma_e^2\theta\theta)}] = \begin{pmatrix} -V_{X_e v}S_{21} - V_{Xee}S_{22} & \begin{pmatrix} -2V_{Xev} & 0_{k_\beta \times 2} \end{pmatrix} & \begin{pmatrix} -2V_{Xee} & 0_{k_\beta \times 2} \end{pmatrix} \end{pmatrix}$$

Put

$$V_2 = \left(V_{21} + V_{21}^\top\right) + \left(V_{22} + V_{22}^\top\right) + V_{23}$$

$$V_{21} = \mathbb{E}\left[d_1 d_{1,1}^\top\right], \qquad V_{22} = \sum_{\{1,1,2,2\}!} \mathbb{E}\left[d_{i_1} d_{i_2 i_3 i_4}^\top\right] \qquad V_{23} = \sum_{\{1,1,2,2\}!} \mathbb{E}\left[d_{i_1 i_2} d_{i_3 i_4}^\top\right]$$

and

$$\text{AMSE}\left[\widehat{\theta}\right] = \frac{1}{N}V_1 + \frac{1}{N^2}V_2$$

let $\widehat{\theta}_3 - \theta_0 = a_{-1} + a_{-1/2} + a_{-3/2}$, denote the third order stochastic expansion of the estimator.

Proposition (Approximate MSE):

$$\mathbb{E}\left[\left(\widehat{\theta}_3 - \theta_0\right)\left(\widehat{\theta}_3 - \theta_0\right)^\top\right] = \text{AMSE}\left[\widehat{\theta}\right] + O\left(N^{-3}\right)$$

Proof: See Appendix . We have put the proof of the proposition in the Appendix. It is straightforward, but a little lengthy. A number of observations can be made regarding this term.

The interpretation of these terms is straightforward. The approximate MSE is the expectation of the outer product of $\widehat{\theta}_3 - \theta_0 = a_{-1/2} + a_{-1} + a_{-3/2}$, which in principal results in nine terms. The expectation of $a_{-1/2}a_{-1/2}^\top$ is the usual asymptotic variance of $\widehat{\theta}$: V_1/N. Of the remaining eight terms, the expectations of three: $a_{-1}a_{-3/2}^\top$, its transpose and $a_{-3/2}a_{-3/2}^\top$ are, as per Lemma of $O(N^{-3})$ or smaller and dropped in the statement of the second-order MSE. The additional terms we see in the variance, V_{21} , V_{22} and V_{23} are the expected values, up to the scaling factor N^{-2}, of, respectively, $a_{-1/2}a_{-1}^\top$, $a_{-1/2}a_{-3/2}^\top$ and $a_{-1}a_{-1}^\top$. Note the caveat "up to the scaling factor"; in the evaluation of these expectations, we have dropped terms which are of smaller order.

This expression is equivalent to those derived in, say Rilstone et al (1996) but is substantially more compact than their expression.

Derivation of the third approximate moment (and consequently measures of skewness) is as follows. Let τ now denote a $1 \times k_\theta$ vector of known constants. Put

$$\text{ASKEW}\left[\tau^\top\widehat{\theta}\right] = \frac{1}{N^2}\mathbb{E}\left[\tau_1^3\right] + 3\frac{1}{N^2}\sum_{\{1,1,2,2\}!} \mathbb{E}\left[\tau_{i_1}\tau_{i_2}\tau_{i_3 i_4}\right]$$

Proposition:

$$\mathbb{E}\left[\left(\tau^{\top}(\widehat{\theta}_3-\theta_0)\right)^3\right]=\text{ASKEW}\left[\tau^{\top}\widehat{\theta}\right]+O\left(N^{-3}\right)$$

Proposition:

$$\begin{aligned}\mathbb{E}\left[\left(\tau^{\top}\left(\widehat{\theta}_3-\theta_0\right)\right)^4\right]&=\frac{1}{N^3}\mathbb{E}\left[\tau_1^4\right]+3\frac{N-1}{N^3}\left(\mathbb{E}[\tau_1^2]\right)^2\\&+4\frac{1}{N^3}\sum_{\{1,1,1,2,2\}!}\mathbb{E}[\tau_{i_1}\tau_{i_2}\tau_{i_3}\tau_{i_4i_5}]+4\frac{1}{N^3}\sum_{\{1,1,2,2,3,3\}!}\mathbb{E}[\tau_{i_1}\tau_{i_2}\tau_{i_3}\tau_{i_4i_5i_6}]\\&+6\frac{1}{N^3}\sum_{\{1,1,2,2,3,3\}!}\mathbb{E}[\tau_{i_1}\tau_{i_2}\tau_{i_3i_4}\tau_{i_5i_6}]+O(N^{-4})\end{aligned}$$

Proof: See the Appendix.

A number of remarks are worth making regarding this moment. First we note that this approximation is accurate up to $O\left(N^{-4}\right)$. The main reason for obtaining this level of accuracy is that, when obtaining Edgeworth expansions below, the expansions will be constructed with respect to $\sqrt{N}(\widehat{\theta}_3-\theta_0)$, not $(\widehat{\theta}_3-\theta_0)$ so that the j'th approximate moment of $\sqrt{N}(\widehat{\theta}_3-\theta_0)$, say, will be $N^{j/2}$ times the j'th moment of $(\widehat{\theta}_3-\theta_0)$.

The second term in the expression for $\mathbb{E}[(\tau^{\top}(\widehat{\theta}_3-\theta_0))^4]$ may seem odd as it is $O(N^{-2})$ and will dominate the other terms in the fourth moment. However, a more useful measure of the fourth moment is the approximate kurtosis which subtracts from this expression three times the approximate MSE squared. If we define

$$\begin{aligned}\text{AKURT}\left[\tau^{\top}\widehat{\theta}_3\right]&=\frac{1}{N^3}\left(\mathbb{E}\left[\tau_1^4\right]-3\left(\mathbb{E}[\tau_1^2]\right)^2\right)+4\frac{1}{N^3}\sum_{\{1,1,1,2,2\}!}\mathbb{E}[\tau_{i_1}\tau_{i_2}\tau_{i_3}\tau_{i_4i_5}]\\&+4\frac{1}{N^3}\sum_{\{1,1,2,2,3,3\}!}\mathbb{E}[\tau_{i_1}\tau_{i_2}\tau_{i_3}\tau_{i_4i_5i_6}]\\&+6\frac{1}{N^3}\sum_{\{1,1,2,2,3,3\}!}\mathbb{E}[\tau_{i_1}\tau_{i_2}\tau_{i_3i_4}\tau_{i_5i_6}]\end{aligned}$$

we see that

$$\mathbb{E}\left[\left(\tau^{\top}\left(\widehat{\theta}_3-\theta_0\right)\right)^4\right]-3\mathbb{E}\left[\left(\tau^{\top}\left(\widehat{\theta}_3-\theta_0\right)\right)^2\right]^2=\text{AKURT}\left[\tau^{\top}\widehat{\theta}_3\right]+O\left(N^{-4}\right)$$

6. Multiplicative Effects: Poisson Models

The results in the previous two sections apply to models in which the unobservable individual effects are additive. This assumption may is untenable in many cases and the results will not hold. One situation in which we can apply the results is that of standard Poisson panel data model with an unobservable individual effect. There we have a count variable, y_{it} and with probability mass function conditional on observable covariates x_{it} and an unobservable individual effect, c_i

$$f(y_{it}|x_{it}, c_i) = \frac{\exp\{-c_i\lambda_{it}(\beta)\}(c_i\lambda_{it}(\beta))^{y_{it}}}{y_{it}!}$$

where, for convenience we use the almost ubiquitous parameterization: $\lambda_{it}(\beta) = \exp(x'_{it}\beta)$ where β has true value β_0. Note that $\mathbb{E}[y_{it}|x_{it}, c_i] = c_i\lambda_{it}$. Treating the c_i as parameters as per Cameron and Trivedi (1998) the concentrated single observation likelihood can be written as

$$L_i(\beta) = \exp\left\{-\sum_{t=1}^{T} y_{it}\right\}\prod_{t=1}^{T}\left(\frac{\bar{y}_i}{\bar{\lambda}_i(\beta)}\right)^{y_{it}}\left(\prod_{t=1}^{T}\lambda_{it}(\beta)\right)^{y_{it}} / \left(\prod_{t=1}^{T} y_{it}!\right)$$

Putting

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \qquad x_i = \begin{pmatrix} x_{i1}^\top \\ \vdots \\ x_{iT}^\top \end{pmatrix}, \qquad \lambda_i = \begin{pmatrix} \lambda_{i1} \\ \vdots \\ \lambda_{iT} \end{pmatrix}$$

$\bar{\lambda}_i = \frac{1}{T}\sum\lambda_{it}$, $\bar{\Lambda}_i = \bigtriangledown_\beta\bar{\lambda}_i$ The concentrated conditional score function can be written in the form above with

$$\begin{aligned} q_i(\beta) &= \sum_{t=1}^{T} x_{it}\left(y_{it} - \lambda_{it}(\beta)\frac{\bar{y}_i}{\bar{\lambda}_i(\beta)}\right) \\ &= x_i^\top\left(y_i - \lambda_i(\beta)\frac{\bar{y}_i}{\bar{\lambda}_i(\beta)}\right) \\ &= T\left(\bar{xy}_i - \bar{y}_i\frac{\bar{\Lambda}_i}{\bar{\lambda}_i}\right) \end{aligned}$$

Differentiating:

$$q_i^{(1)}(\beta) = -T\overline{y}_i\left(\frac{\bar{\Lambda}_i^{(1)}}{\bar{\lambda}_i} - \frac{\bar{\Lambda}_i\lambda_i^{(1)}}{\bar{\lambda}_i^2}\right)$$

We show in the appendix that

$$q_i^{(2)}(\beta) = -T\overline{y}_i\left(\frac{A_{1i}}{\bar{\lambda}_i} + \frac{A_{2i}}{\bar{\lambda}_i^2} + \frac{A_{3i}}{\bar{\lambda}_i^3}\right)$$

$$\begin{aligned} A_{1i} &= \bar{\Lambda}_i^{(2)} \\ A_{2i} &= -\bar{\Lambda}_i\bar{\lambda}_i^{(2)} - 2\,\bar{\Lambda}_i^{(1)} \otimes \bar{\lambda}_i^{(1)} \\ A_{3i} &= 2\left(\bar{\Lambda}_i\bar{\lambda}_i^{(1)}\right) \otimes \bar{\lambda}_i^{(1)} \end{aligned}$$

and

$$q_i^{(3)}(\beta) = -T\overline{y}_i \left(\frac{B_{1i}}{\overline{\lambda}_i} + \frac{B_{2i} + B_{3i}}{\overline{\lambda}_i^2} + \frac{B_{4i}}{\overline{\lambda}_i^3} + \frac{B_{5i}}{\overline{\lambda}_i^4} \right)$$

$$\begin{aligned}
B_{1i} &= \bar{\Lambda}_i^{(3)} \\
B_{2i} &= -\bar{\Lambda}_i^{(1)}(\bar{\lambda}_i^{(2)} \otimes I_k) - 2\,\bar{\Lambda}_i^{(1)} \otimes \bar{\lambda}_i^{(2)} \\
B_{3i} &= -\bar{\Lambda}_i \bar{\lambda}_i^{(3)} - 2(\bar{\lambda}_i \otimes \bar{\Lambda}_i^{(2)})(K_{kk} \otimes I_k) \\
B_{4i} &= 4\,\bar{\Lambda}_i^{(1)} \otimes \bar{\lambda}_i^{(1)\otimes 2} + 2\,(\bar{\lambda}_i^{(1)} \bar{\Lambda}_i) \otimes \bar{\lambda}_i^{(2)} \\
&\quad + 2\left[\bar{\lambda}_i^{(1)} \otimes ((\bar{\Lambda}_i^{(1)}(\bar{\lambda}_i^{(1)} \otimes I_k)) + \bar{\lambda}_i^{(1)} \otimes (\bar{\Lambda}_i \bar{\lambda}_i^{(2)})\right](K_{kk} \otimes I_k) \\
B_{5i} &= -6\,(\bar{\Lambda}_i \bar{\lambda}_i^{(1)}) \otimes \bar{\lambda}_i^{(1)\otimes 2}
\end{aligned}$$

Note that the j derivatives of λ_{it} are simply polynomials in the X_{it}'s. To gleam so clarity into these expressions, its useful to consider the case of a single regressor in which case:

$$q_i(\beta) = T\left(\overline{xy}_i + \overline{y}_i \frac{\overline{\nabla \lambda}_i}{\overline{\lambda}_i} \right)$$

$$q_i^{(1)}(\beta) = T\overline{y}_i \left(\frac{\overline{\lambda X^2}_i}{\overline{\lambda}_i} - \frac{\overline{\lambda X}_i^2}{\overline{\lambda}_i^2} \right)$$

$$q_i^{(2)}(\beta) = T\overline{y}_i \left(\frac{A_{1i}}{\overline{\lambda}_i} + \frac{A_{2i}}{\overline{\lambda}_i^2} + \frac{A_{3i}}{\overline{\lambda}_i^3} \right)$$

$$\begin{aligned}
A_{1i} &= \overline{\lambda X^3}_i \\
A_{2i} &= -3(\overline{\lambda X}_i)(\overline{\lambda X^2}_i) \\
A_{3i} &= 2\,\overline{\lambda X}_i^3
\end{aligned}$$

and

$$q_i^{(3)}(\beta) = T\overline{y}_i \left(\frac{B_{1i}}{\overline{\lambda}_i} + \frac{B_{2i} + B_{3i}}{\overline{\lambda}_i^2} + \frac{B_{4i}}{\overline{\lambda}_i^3} + \frac{B_{5i}}{\overline{\lambda}_i^4} \right)$$

$$\begin{aligned}
B_{1i} &= \overline{\lambda X^4}_i \\
B_{2i} &= -3\overline{\lambda X^2}_i^2 \\
B_{3i} &= -3\overline{\lambda X}_i \overline{\lambda X^3}_i \\
B_{4i} &= 10\overline{\lambda X}_i^2 \overline{\lambda X^2}_i \\
B_{5i} &= -6\,\overline{\lambda X}_i^4
\end{aligned}$$

Appendix
Proof to Lemma 4.1

$$q_{Wi}^{(1)}(\beta) = -G_i^\top(\beta) W G_i(\beta) + G_i^{\top(1)}(\beta)((W(y_i - g_i(\beta)) \otimes I_k)$$

$$q_i^{(1)} = -G_i^\top W G_i + G_i^{\top(1)}((W u_i) \otimes I_k)$$

$$\mathbb{E}[q_i^{(1)}] = -\mathbb{E}[G_i^\top W G_i]$$

$$\begin{aligned}
q_i^{(2)}(\beta) &= -G_i^\top(\beta) W G_i^{(1)}(\beta) - G_i^{\top(1)}(\beta)(W G_i(\beta) \otimes I_k) \\
&\quad + G_i^{\top(1)}(\beta) \bigtriangledown_{\beta^\top} ((W(y_i - g_i(\beta)) \otimes I_k) + G_i^{\top(2)}(\beta)((W(y_i - g_i(\beta)) \otimes I_{k^2}) \\
&= -G_i^\top(\beta) W G_i^{(1)}(\beta) - G_i^{\top(1)}(\beta)(W G_i(\beta) \otimes I_k) \\
&\quad - G_i^{\top(1)}(\beta)(K_{kT} \otimes I_1)(I_k \otimes (W G_i(\beta))(K_{1k} \otimes I_k) + G_i^{\top(2)}(\beta)((W(y_i - g_i(\beta)) \otimes I_{k^2}) \\
&= -G_i^\top(\beta) W G_i^{(1)}(\beta) - G_i^{\top(1)}(\beta)(W G_i(\beta) \otimes I_k) \\
&\quad - G_i^{\top(1)}(\beta) K_{kT}(I_k \otimes (W G_i(\beta)) + G_i^{\top(2)}(\beta)((W(y_i - g_i(\beta)) \otimes I_{k^2})
\end{aligned}$$

$$\begin{aligned}
q_i^{(2)} &= -G_i^\top W G_i^{(1)} - G_i^{\top(1)}(W G_i \otimes I_k) \\
&\quad - G_i^{\top(1)} K_{kT}(I_k \otimes (W G_i)) + G_i^{\top(2)}((W u_i) \otimes I_{k^2})
\end{aligned}$$

$$\mathbb{E}[q_i^{(2)}] = -\mathbb{E}[G_i^\top W G_i^{(1)}] - \mathbb{E}[G_i^{\top(1)}(W G_i \otimes I_k)] - \mathbb{E}[G_i^{\top(1)} K_{kT}(I_k \otimes (W G_i))]$$

$$
\begin{aligned}
q_i^{(3)}(\beta) &= -G_i^{\top}(\beta)WG_i^{(2)}(\beta) - G_i^{\top(1)}(\beta)(WG_i^{(1)}(\beta)\otimes I_k)\\
&\quad - G_i^{\top(1)}(\beta)\bigtriangledown_{\beta^{\top}}(WG_i(\beta)\otimes I_k) - G_i^{\top(2)}(\beta)(WG_i(\beta)\otimes I_{k^2})\\
&\quad - G_i^{\top(1)}(\beta)K_{kT}\bigtriangledown_{\beta^{\top}}(I_k\otimes(WG_i(\beta)) - G_i^{\top(2)}(\beta)\left((K_{kT}(I_k\otimes(WG_i(\beta)))\otimes I_k\right)\\
&\quad + G_i^{\top(2)}(\beta)((W(y_i-g_i(\beta))\otimes I_{k^2}) + G_i^{\top(3)}(\beta)\bigtriangledown_{\beta^{\top}}((W(y_i-g_i(\beta))\otimes I_{k^3})\\
&= -G_i^{\top}(\beta)WG_i^{(2)}(\beta) - G_i^{\top(1)}(\beta)(WG_i^{(1)}(\beta)\otimes I_k)\\
&\quad - G_i^{\top(1)}(\beta)(K_{Tk}\otimes I_1)(I_k\otimes(WG_i^{(1)}(\beta))(K_{kk}\otimes I_k) - G_i^{\top(2)}(\beta)(WG_i(\beta)\otimes I_{k^2})\\
&\quad - G_i^{\top(1)}(\beta)K_{kT}(I_k\otimes(WG_i^{(1)}(\beta)) - G_i^{\top(2)}(\beta)\left((K_{kT}(I_k\otimes(WG_i(\beta)))\otimes I_k\right)\\
&\quad + G_i^{\top(2)}(\beta)((W(y_i-g_i(\beta))\otimes I_{k^2})\\
&\quad - G_i^{\top(3)}(\beta)(K_{k^3T}\otimes I_1)(I_{k^3}\otimes WG_i(\beta))(K_{1k^3}\otimes I_k)\\
&= -G_i^{\top}(\beta)WG_i^{(2)}(\beta) - G_i^{\top(1)}(\beta)(WG_i^{(1)}(\beta)\otimes I_k)\\
&\quad - G_i^{\top(1)}(\beta)K_{Tk}(I_k\otimes(WG_i^{(1)}(\beta))(K_{kk}\otimes I_k) - G_i^{\top(2)}(\beta)(WG_i(\beta)\otimes I_{k^2})\\
&\quad - G_i^{\top(1)}(\beta)K_{kT}(I_k\otimes(WG_i^{(1)}(\beta)) - G_i^{\top(2)}(\beta)\left((K_{kT}(I_k\otimes(WG_i(\beta)))\otimes I_k\right)\\
&\quad + G_i^{\top(2)}(\beta)((W(y_i-g_i(\beta))\otimes I_{k^2})\\
&\quad - G_i^{\top(3)}(\beta)K_{k^3T}(I_{k^3}\otimes WG_i(\beta))
\end{aligned}
$$

$$
\begin{aligned}
q_i^{(3)} &= -G_i^{\top}WG_i^{(2)} - G_i^{\top(1)}(WG_i^{(1)}\otimes I_k)\\
&\quad - G_i^{\top(1)}K_{Tk}(I_k\otimes(WG_i^{(1)})(K_{kk}\otimes I_k) - G_i^{\top(2)}(WG_i\otimes I_{k^2})\\
&\quad - G_i^{\top(1)}K_{kT}(I_k\otimes(WG_i^{(1)}) - G_i^{\top(2)}\left((K_{kT}(I_k\otimes(WG_i)\otimes I_k\right)\\
&\quad + G_i^{\top(2)}((Wu_i\otimes I_{k^2}) - G_i^{\top(3)}K_{k^3T}(I_{k^3}\otimes WG_i)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[q_i^{(3)}] &= -\mathbb{E}[G_i^{\top}WG_i^{(2)}] - \mathbb{E}[G_i^{\top(1)}(WG_i^{(1)}\otimes I_k)]\\
&\quad - \mathbb{E}[G_i^{\top(1)}K_{Tk}(I_k\otimes(WG_i^{(1)})(K_{kk}\otimes I_k)] - \mathbb{E}[G_i^{\top(2)}(WG_i\otimes I_{k^2})]\\
&\quad - \mathbb{E}[G_i^{\top(1)}K_{kT}(I_k\otimes(WG_i^{(1)})] - \mathbb{E}[G_i^{\top(2)}\left((K_{kT}(I_k\otimes(WG_i)\otimes I_k\right)]\\
&\quad - \mathbb{E}[G_i^{\top(3)}K_{k^3T}(I_{k^3}\otimes WG_i)]
\end{aligned}
$$

Derivations for fixed regressor case.

Put $A_{i_1i_2i_3i_4} = \overline{V}_Q d_{i_1i_2} d_{i_3i_4}^{\top}\overline{V}_Q$

$$
\begin{aligned}
A_{i_1i_2i_3i_4} &= \left(-G_{i_1}^{\top}QG_{i_1}d_{i_2} + G_{i_1}^{\top(1)}((Qu_{i_1})\otimes d_{i_2}) + \frac{1}{2}\mathbb{E}[q_{Qi}^{(2)}]\,(d_{i_1}\otimes d_{i_2})\right)\\
&\quad\times\left(-G_{i_3}^{\top}QG_{i_3}d_{i_4} + G_{i_3}^{\top(1)}((Qu_{i_3})\otimes d_{i_4}) + \frac{1}{2}\mathbb{E}[q_{Q1}^{(2)}]\,(d_{i_3}\otimes d_{i_4})\right)^{\top}
\end{aligned}
$$

$$
\begin{aligned}
A_{1212} = &\left(-G_1^\top Q G_1 d_2 + G_1^{\top(1)}((Qu_1)\otimes d_2) + \frac{1}{2}\mathbb{E}[q_{Qi}^{(2)}]\,(d_1 \otimes d_2)\right) \\
&\quad \times \left(-G_1^\top Q G_1 d_2 + G_1^{\top(1)}((Qu_1)\otimes d_2) + \frac{1}{2}\mathbb{E}[q_{Q1}^{(2)}]\,(d_1 \otimes d_2)\right)^\top \\
= &\ \sigma_e^2 \mathbb{E}[G_1^\top Q G_1 V_Q G_1^\top Q G_1] + A_1 + A_2 + A_2^\top + A_3
\end{aligned}
$$

$$
\begin{aligned}
A_1 &= \sigma_e^4 \mathbb{E}[G_1^{\top(1)} K_{Tk}(V_Q \otimes Q) G_1^{\top(1)\top}] \\
A_2 &= \sigma_e^4 \frac{1}{2}\mathbb{E}[q_{Qi}^{(2)}]\mathbb{E}\left[(V_Q^{-1} G_1^\top Q \otimes V_Q K_{kT} G_1^{\top(1)\top}\right] \\
A_3 &= \sigma_e^4 \frac{1}{4}\mathbb{E}[q_{Q1}^{(2)}] V_Q^{-1\otimes 2}\mathbb{E}[q_{Q1}^{(2)\top}]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}((Qu_1)\otimes d_2)(d_2^\top \otimes (u_1^\top Q) &= \mathbb{E}((Qu_1)\otimes V_1 \otimes (u_1^\top Q) \\
&= \mathbb{E} K_{Tk}(V_1 \otimes ((Qu_1)\otimes (u_1^\top Q)) \\
&= \sigma_e^4 K_{Tk}(V_Q \otimes Q)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[(d_1 \otimes d_2)(d_2^\top \otimes (u_1^\top Q) G_1^{\top(1)\top}\right] &= \sigma_e^2 \mathbb{E}\left[(V_Q^{-1} G_1^\top Q u_1 u_1^\top Q \otimes V_Q K_{kT} G_1^{\top(1)\top}\right] \\
&= \sigma_e^4 \mathbb{E}\left[(V_Q^{-1} G_1^\top Q \otimes V_Q K_{kT} G_1^{\top(1)\top}\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[(d_1 \otimes\ d_2)\,(d_1 \otimes d_2)^\top\right] &= \mathbb{E}[d_1 d_1^\top \otimes d_2 d_2^\top] \\
&= \sigma_e^4 V_Q^{-1\otimes 2}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}A_{1221} = &\ \mathbb{E}\left(-G_1^\top Q G_1 d_2 + G_1^{\top(1)}((Qu_1)\otimes d_2) + \frac{1}{2}\mathbb{E}[q_{Qi}^{(2)}]\,(d_1 \otimes d_2)\right) \\
&\quad \times \left(-G_2^\top Q G_2 d_1 + G_2^{\top(1)}((Qu_2)\otimes d_1) + \frac{1}{2}\mathbb{E}[q_{Q1}^{(2)}]\,(d_2 \otimes d_1)\right)^\top \\
= &\ B_1 + B_2 + B_2^\top + B_3
\end{aligned}
$$

$$
\begin{aligned}
B_1 &= \sigma_e^4 \mathbb{E}\left[G_1^{\top(1)}(Q G_1 \otimes G_2^\top Q) G_2^{\top(1)\top}\right] \\
B_2 &= \sigma_e^4 \frac{1}{2}\mathbb{E}[q_{Qi}^{(2)}]\mathbb{E}\mathbb{E}\left[(V_Q^{-1} \otimes V_Q^{-1} G_2^\top Q) G_2^{\top(1)\top}\right] \\
B_3 &= \sigma_e^4 \frac{1}{4}\mathbb{E}[q_{Q1}^{(2)}] V_Q^{-1\otimes 2}\mathbb{E}[q_{Q1}^{(2)\top}]
\end{aligned}
$$

$$\mathbb{E}\left[G_1^{\top(1)}(Qu_1)\otimes d_2(d_1^\top\otimes(u_2^\top Q)G_2^{\top(1)\top}\right]=\mathbb{E}\left[G_1^{\top(1)}(Qu_1)d_1^\top\otimes d_2((u_2^\top Q)G_2^{\top(1)\top}\right]$$
$$=\mathbb{E}\left[G_1^{\top(1)}(Qu_1u_1^\top QG_1)\otimes G_2^\top Qu_2((u_2^\top Q)G_2^{\top(1)\top}\right]$$
$$=\sigma_e^4\mathbb{E}\left[G_1^{\top(1)}(QG_1\otimes G_2^\top Q)G_2^{\top(1)\top}\right]$$

$$\mathbb{E}\left[(d_1\otimes d_2)(d_1^\top\otimes(u_2^\top Q)G_2^{\top(1)\top}\right]=\sigma_e^4\mathbb{E}\left[(V_Q^{-1}\otimes V_Q^{-1}G_2^\top Q)G_2^{\top(1)\top}\right]$$

$$\mathbb{E}\left[(d_1\otimes\ d_2)\left(d_2\otimes d_1\right)^\top\right]=\mathbb{E}\left[(d_1\otimes\ d_2)\left(d_1\otimes d_2\right)^\top K_{kk}\right]$$
$$=\sigma_e^4\mathbb{E}\left[V_Q^{-1\otimes 2}\top K_{kk}\right]$$

Put $A_{i_1i_2i_3i_4}=\overline{V}_Qd_{i_1i_2i_3}d_{i_4}^\top\overline{V}_Q$ so that

$$\begin{aligned}
A_{i_1i_2i_3i_4}&=\frac{1}{2}\left(\overline{q_1^{(2)}}\left(d_{i_1}\otimes d_{i_2i_3}\right)\right)q_{i_4}^\top\\
&\quad+\left(d_{i_1i_2}\otimes d_{i_3}\right)q_{i_4}^\top\\
&\quad+\frac{1}{6}\left(\overline{q_1^{(3)}}\left(d_{i_1}\otimes d_{i_2}\otimes d_{i_3}\right)\right)q_{i_4}^\top\\
&\quad+\left(\widetilde{q_{i_1}^{(1)}}\left(d_{i_2i_3}\right)\right)q_{i_4}^\top\\
&\quad+\frac{1}{2}\left(\widetilde{q_{i_1}^{(2)}}\left(d_{i_2}\otimes d_{i_3}\right)\right)q_{i_4}^\top\\
&=\frac{1}{2}\left(\overline{q_1^{(2)}}\left(d_{i_1}\otimes\left(\overline{V}_Q^{-1}\left(-G_{i_2}^\top QG_{i_2}+G_{i_2}^{\top(1)}((Qu_{i_2})\otimes I_k)\right)d_{i_3}+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_{i_2}\otimes d_{i_3}\right)\right)\right)\right)q_{i_4}^\top\\
&\quad+\left(\left(\overline{V}_Q^{-1}\left(-G_{i_1}^\top QG_{i_1}+G_{i_1}^{\top(1)}((Qu_{i_1})\otimes I_k)\right)d_{i_2}+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_{i_1}\otimes d_{i_2}\right)\right)\otimes d_{i_3}\right)q_{i_4}^\top\\
&\quad+\frac{1}{6}\left(\overline{q_1^{(3)}}\left(d_{i_1}\otimes d_{i_2}\otimes d_{i_3}\right)\right)q_{i_4}^\top\\
&\quad+\left(\widetilde{q_{i_1}^{(1)}}\left(\overline{V}_Q^{-1}\left(-G_{i_2}^\top QG_{i_2}+G_{i_2}^{\top(1)}((Qu_{i_2})\otimes I_k)\right)d_{i_3}+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_{i_2}\otimes d_{i_3}\right)\right)\right)q_{i_4}^\top\\
&\quad+\frac{1}{2}\left(\widetilde{q_{i_1}^{(2)}}\left(d_{i_2}\otimes d_{i_3}\right)\right)q_{i_4}^\top
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}A_{1122} = {}& \sigma_e^2\frac{1}{2}\mathbb{E}\left(\overline{q_1^{(2)}}\left(d_1\otimes\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}((Qu_1)\otimes I_k)\right)I_k+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_1\otimes I_k\right)\right)\right)\right)\\
&+\sigma_e^2\mathbb{E}\left(\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}((Qu_1)\otimes I_k)\right)d_1+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_1\otimes d_1\right)\right)\otimes I_k\right)\\
&+\sigma_e^4\frac{1}{6}\mathbb{E}\left(\overline{q_1^{(3)}}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\otimes I_k\right)\right)\\
&+\sigma_e^2\mathbb{E}\left(G_1^{\top(1)}(Qu_1\otimes I_k)\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}((Qu_1)\otimes I_k)\right)I_k+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_1\otimes I_k\right)\right)\right)\\
&+\sigma_e^2\frac{1}{2}\mathbb{E}\left(G_1^{\top(2)}(Qu_1\otimes I_{k^2})\left(d_1\otimes I_k\right)\right)\\
={}& \sigma_e^2\frac{1}{2}\mathbb{E}\left(\overline{q_1^{(2)}}\left((I_k\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(d_1\otimes(Qu_1)\otimes I_k)I_k+\frac{1}{2}(I_k\otimes\overline{V}_Q^{-1}\bar{H})\left(d_1\otimes d_1\otimes I_k\right)\right)\right)\\
&+\sigma_e^2\mathbb{E}\left(\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}((Qu_1)\otimes I_k)\right)d_1+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(\mathrm{Vec}[\overline{V}_1^{-1}]\right)\right)\otimes I_k\right)\\
&+\sigma_e^4\frac{1}{6}\mathbb{E}\left(\overline{q_1^{(3)}}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\otimes I_k\right)\right)\\
&+\sigma_e^2\mathbb{E}\left(G_1^{\top(1)}(Qu_1\otimes I_k)\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}((Qu_1)\otimes I_k)\right)I_k+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_1\otimes I_k\right)\right)\right)\\
&+\sigma_e^2\frac{1}{2}\mathbb{E}\left(G_1^{\top(2)}(Qu_1\otimes I_{k^2})\left(d_1\otimes I_k\right)\right)\\
={}& \sigma_e^4\frac{1}{2}\mathbb{E}\left(\overline{q_1^{(2)}}\left((I_k\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)I_k+\frac{1}{2}(I_k\otimes\overline{V}_Q^{-1}\bar{H})\left(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k\right)\right)\right)\\
&+\sigma_e^4\mathbb{E}\left(\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}])\right)+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\right)\right)\otimes I_k\right)\\
&+\sigma_e^4\frac{1}{6}\mathbb{E}\left(\overline{q_1^{(3)}}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\otimes I_k\right)\right)\\
&+\sigma_e^2\mathbb{E}\left(G_1^{\top(1)}(Qu_1\otimes I_k)\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}((Qu_1)\otimes I_k)\right)+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(d_1\otimes I_k\right)\right)\right)\\
&+\sigma_e^4\frac{1}{2}\mathbb{E}\left(G_1^{\top(2)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)\right)\\
={}& \sigma_e^4\frac{1}{2}\mathbb{E}\left(\overline{q_1^{(2)}}\left((I_k\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)+\frac{1}{2}(I_k\otimes\overline{V}_Q^{-1}\bar{H})\left(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k\right)\right)\right)\\
&+\sigma_e^4\mathbb{E}\left(\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}])\right)+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\right)\right)\otimes I_k\right)\\
&+\sigma_e^4\frac{1}{6}\mathbb{E}\left(\overline{q_1^{(3)}}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\otimes I_k\right)\right)\\
&+\sigma_e^2\mathbb{E}\left(G_1^{\top(1)}(Qu_1\otimes\overline{V}_Q^{-1}G_1^{\top(1)})((Qu_1)\otimes I_k)+\frac{1}{2}G_1^{\top(1)}(Qu_1\otimes\overline{V}_Q^{-1}\bar{H})\left(d_1\otimes I_k\right)\right)\\
&+\sigma_e^4\frac{1}{2}\mathbb{E}\left(G_1^{\top(2)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}A_{1122} &= \sigma_e^4\frac{1}{2}\mathbb{E}\left(\overline{q_1^{(2)}}\left((I_k\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)+\frac{1}{2}(I_k\otimes\overline{V}_Q^{-1}\bar{H})\left(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k\right)\right)\right)\\
&\quad+\sigma_e^4\mathbb{E}\left(\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}])\right)+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\right)\right)\otimes I_k\right)\\
&\quad+\sigma_e^4\frac{1}{6}\mathbb{E}\left(\overline{q_1^{(3)}}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\otimes I_k\right)\right)\\
&\quad+\sigma_e^2\mathbb{E}\left(G_1^{\top(1)}(I_T\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(Qu_1\otimes Qu_1\otimes I_k)+\frac{1}{2}G_1^{\top(1)}(I_T\otimes\overline{V}_Q^{-1}\bar{H})(Qu_1\otimes d_1\otimes I_k)\right)\\
&\quad+\sigma_e^4\frac{1}{2}\mathbb{E}\left(G_1^{\top(2)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)\right)\\
&= \sigma_e^4\frac{1}{2}\mathbb{E}\left(\overline{q_1^{(2)}}\left((I_k\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)+\frac{1}{2}(I_k\otimes\overline{V}_Q^{-1}\bar{H})\left(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k\right)\right)\right)\\
&\quad+\sigma_e^4\mathbb{E}\left(\left(\overline{V}_Q^{-1}\left(G_1^{\top(1)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}])\right)+\frac{1}{2}\overline{V}_Q^{-1}\bar{H}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\right)\right)\otimes I_k\right)\\
&\quad+\sigma_e^4\frac{1}{6}\mathbb{E}\left(\overline{q_1^{(3)}}\left(\mathrm{Vec}[\overline{V}_Q^{-1}]\otimes I_k\right)\right)\\
&\quad+\sigma_e^4\left(\mathbb{E}G_1^{\top(1)}(I_T\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(\mathrm{Vec}[Q^{\otimes 2}]\otimes I_k)+\frac{1}{2}\mathbb{E}G_1^{\top(1)}(I_T\otimes\overline{V}_Q^{-1}\bar{H})(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)\right)\\
&\quad+\sigma_e^4\frac{1}{2}\mathbb{E}\left(G_1^{\top(2)}(\mathrm{Vec}[QG_1\bar{V}_Q^{-1}]\otimes I_k)\right)
\end{aligned}
$$

$$
\begin{aligned}
(Qu_1\otimes\overline{V}_Q^{-1}G_1^{\top(1)})((Qu_1)\otimes I_k) &= (I_TQu_1\otimes\overline{V}_Q^{-1}G_1^{\top(1)}I_{Tk})((Qu_1)\otimes I_k)\\
&= (I_T\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(Qu_1\otimes I_{Tk})((Qu_1)\otimes I_k)\\
&= (I_T\otimes\overline{V}_Q^{-1}G_1^{\top(1)})(Qu_1\otimes Qu_1\otimes I_k)
\end{aligned}
$$

$$
\begin{aligned}
(Qu_1\otimes\overline{V}_Q^{-1}\bar{H})\,(d_1\otimes I_k) &= (I_TQu_1\otimes\overline{V}_Q^{-1}\bar{H}I_{k^2})\,(d_1\otimes I_k)\\
&= (I_T\otimes\overline{V}_Q^{-1}\bar{H})(Qu_1\otimes I_{k^2})\,(d_1\otimes I_k)\\
&= (I_T\otimes\overline{V}_Q^{-1}\bar{H})(Qu_1\otimes d_1\otimes I_k)
\end{aligned}
$$

We first provide expressions for $\mathrm{Var}[(u_i^\top u_i)]$and $\mathrm{Var}[(e_1^\top e_1)]$. Noting

$$
\begin{aligned}
\mathbb{E}\left[(e_i^\top e_i)^2\right] &= \mathbb{E}\left[\sum_{t=1}^T e_{it}^4\right]+\mathbb{E}\left[\sum_{t\neq\tau}e_{it}^2e_{i\tau}^2\right]\\
&= T\mathbb{E}\left[e_{11}^4\right]+T(T-1)(\sigma_e^2)^2
\end{aligned}
$$

so

$$
\begin{aligned}
\mathrm{Var}[e_1^\top e_1] &= \mathbb{E}\left[(e_i^\top e_i)^2\right]-T^2(\sigma_e^2)^2\\
&= T\mathbb{E}\left[e_{11}^4\right]+T(T-1)(\sigma_e^2)^2-T^2(\sigma_e^2)^2\\
&= T\left(\mathbb{E}\left[e_{11}^4\right]-(\sigma_e^2)^2\right)
\end{aligned}
$$

To derive $\mathrm{Var}[u_i^\top u_i]$, note first that,

$$u_{it}^4 = (v_i + e_{it})^4 = \sum_{j=0}^{4} \binom{4}{j} v_i^j e_{it}^{4-j}$$

so that with symmetric errors

$$\mathbb{E}[u_{it}^4] = \mathbb{E}[v_i^4] + 6\sigma_v^2\sigma_e^2 + \mathbb{E}[e_{11}^4]$$

and

$$\begin{aligned} u_{it}^2 u_{i\tau}^2 &= (v_i + e_{it})^2 (v_i + e_{it})^2 \\ &= (v_i^2 + 2v_i e_{it} + e_{it}^2)(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) \\ &= v_i^2(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) + 2v_i e_{it}(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) + e_{it}^2(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) \end{aligned}$$

and thus for $t \neq \tau$

$$\begin{aligned} \mathbb{E}[u_{it}^2 u_{i\tau}^2] &= \mathbb{E}v_i^2(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) + 2\mathbb{E}v_i e_{it}(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) + \mathbb{E}e_{it}^2(v_i^2 + 2v_i e_{i\tau} + e_{i\tau}^2) \\ &= \mathbb{E}[v_i^4] + \sigma_v^2\sigma_e^2 + 0 + \sigma_v^2\sigma_e^2 + (\sigma_e^2)^2 \\ &= \mathbb{E}[v_i^4] + 2\sigma_v^2\sigma_e^2 + (\sigma_e^2)^2 \end{aligned}$$

We have then

$$\begin{aligned} \mathbb{E}\left[(u_i^\top u_i)^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T} u_{it}^4\right] + \mathbb{E}\left[\sum_{t\neq\tau} u_{it}^2 u_{i\tau}^2\right] \\ &= T\left(\mathbb{E}[v_i^4] + 6\sigma_v^2\sigma_e^2 + \mathbb{E}[e_{11}^4]\right) + T(T-1)\left(\mathbb{E}[v_i^4] + 2\sigma_v^2\sigma_e^2 + (\sigma_e^2)^2\right) \end{aligned}$$

so

$$\begin{aligned} \mathrm{Var}[(u_i^\top u_i)] &= \mathbb{E}\left[(u_i^\top u_i)^2\right] - T^2(\sigma_v^2 + \sigma_e^2)^2 \\ &= T\left(\mathbb{E}[v_i^4] + 6\sigma_v^2\sigma_e^2 + \mathbb{E}[e_{11}^4]\right) + T(T-1)\left(\mathbb{E}[v_i^4] + 2\sigma_v^2\sigma_e^2 + (\sigma_e^2)^2\right) \\ &\quad - T^2(\sigma_v^2 + \sigma_e^2)^2 \end{aligned}$$

When calculating higher order derivatives it is useful to use a result for partitioned matrices. Specifically, if $A = A(\theta)$, an $n \times p$ matrix and we partition $\theta = \begin{pmatrix} \beta \\ \mu \end{pmatrix}$ where β and μ and θ are $k_\beta \times l$, $k_\mu \times 1$ and $k_\theta \times 1$, $k_\theta = k_\beta + k_\mu$ vectors, respectively.

$$S_\mu = \begin{pmatrix} 0_{k_\mu \times k_\beta} & I_{k_\mu} \end{pmatrix}, \qquad S_\beta = \begin{pmatrix} I_{k_\beta} & 0_{k_\beta \times k_\mu} \end{pmatrix}$$

Then:

$$\triangledown_{\theta^\top} A(\theta) = \triangledown_{\beta^\top} A(\theta) \left(I_p \otimes S_\beta\right) + \triangledown_{\mu^\top} A(\theta) \left(I_p \otimes S_\mu\right)$$

Put $\mu = \begin{pmatrix} \sigma_v^2 & \sigma_e^2 \end{pmatrix}$

$$q_i^{(2)}(\theta) = \begin{pmatrix} q_{1i}^{(\beta\theta)}(\theta) & q_{1i}^{(\sigma_v^2\theta)}(\theta) & q_{1i}^{(\sigma_e^2\theta)}(\theta) \\ q_{2i}^{(\beta\theta)}(\theta) & q_{2i}^{(\sigma_v^2\theta)}(\theta) & q_{2i}^{(\sigma_e^2\theta)}(\theta) \\ q_{3i}^{(\beta\theta)}(\theta) & q_{3i}^{(\sigma_v^2\theta)}(\theta) & q_{3i}^{(\sigma_e^2\theta)}(\theta) \end{pmatrix}$$

$$q_{1i}^{(\beta\theta)}(\theta) = -\bigtriangledown_\mu \left(X_i^\top \Omega^{-1} X_i\right) \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix}\right)$$

$$\bigtriangledown_{\mu^\top} (X_i^\top \Omega^{-1} X_i) = X_i^\top \Omega_v^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}\right) + X_i^\top \Omega_e^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix}\right)$$

so

$$\begin{aligned} q_{1i}^{(\beta\theta)}(\theta) &= -\left\{X_i^\top \Omega_v^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}\right) + X_i^\top \Omega_e^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix}\right)\right\} \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix}\right) \\ &= -X_i^\top \Omega_v^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \begin{pmatrix} 1 & 0 \end{pmatrix} \end{pmatrix}\right) - X_i^\top \Omega_e^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \begin{pmatrix} 0 & 1 \end{pmatrix} \end{pmatrix}\right) \\ &= -X_i^\top \Omega_v^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \iota_2^{1\top} \end{pmatrix}\right) - X_i^\top \Omega_e^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \iota_2^{2\top} \end{pmatrix}\right) \end{aligned}$$

$$\mathbb{E}[q_{1i}^{(\beta\theta)}] = -\mathbb{E}[X_i^\top \Omega_v^{-1} X_i] \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \iota_2^{1\top} \end{pmatrix}\right) - \mathbb{E}[X_i^\top \Omega_e^{-1} X_i] \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times k_\beta} & \iota_2^{2\top} \end{pmatrix}\right)$$

$$\begin{aligned} \mathbb{E} \bigtriangledown_{\sigma^2} (X_i^\top \Omega^{-1} X_i) = \mathbb{E}[X_i^\top \Omega_v^{-1} X_i] \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}\right) \\ + \mathbb{E}[X_i^\top \Omega_e^{-1} X_i] \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix}\right) \end{aligned}$$

$$q_{1i}^{(\sigma_v^2\theta)}(\theta) = \begin{pmatrix} -X_i^\top \Omega_v^{-1} X_i & 2X_i^\top \Omega_{vv}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{ve}^{-1}(y_i - X_i\beta) \end{pmatrix}$$

$$q_{1i}^{(\sigma_e^2\theta)}(\theta) = \begin{pmatrix} -X_i^\top \Omega_e^{-1} X_i & 2X_i^\top \Omega_{ve}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{ee}^{-1}(y_i - X_i\beta) \end{pmatrix}$$

To derive $q_{2i}^{(\beta\theta)}$ and $q_{3i}^{(\beta\theta)}$ note that

$$\begin{aligned} \bigtriangledown_{\beta^\top}(\beta^\top X_i^\top X_i) &= (\bigtriangledown_{\beta^\top} \beta^\top)(I_{k_\beta} \otimes X_i^\top X_i) \\ &= (\mathrm{Vec}[I_{k_\beta}])^\top (I_{k_\beta} \otimes X_i^\top X_i) \\ &= (\mathrm{Vec}[X_i^\top X_i])^\top \end{aligned}$$

and consequently we have

$$\begin{pmatrix} q_{2i}^{(2)}(\theta) \\ q_{3i}^{(2)}(\theta) \end{pmatrix} = \begin{pmatrix} 2\mathrm{Vec}[X_i^\top X_i]^\top & 0_{1\times k_\theta^2 - \theta_\beta^2} \\ 2\mathrm{Vec}[X_i^\top Q X_i]^\top & 0_{1\times k_\theta^2 - \theta_\beta^2} \end{pmatrix}$$

Note that

$$\mathbb{E}[q_{1i}^{(\sigma_v^2\theta)}] = \begin{pmatrix} -\mathbb{E}[X_i^\top \Omega_v^{-1} X_i] & 0_{k_\beta \times 1} & 0_{k_\beta \times 1} \end{pmatrix}$$

$$\mathbb{E}[q_{1i}^{(\sigma_e^2\theta)}] = \begin{pmatrix} -\mathbb{E}[X_i^\top \Omega_e^{-1} X_i] & 0_{k_\beta \times 1} & 0_{k_\beta \times 1} \end{pmatrix}$$

$$\mathbb{E}[q_{1i}^{(2)}] = \Big(\mathbb{E}q_{1i}^{(\beta\theta)}(\theta) \quad -\mathbb{E}[X_i^\top \Omega_v^{-1} X_i] \quad 0_{k_\beta \times 2} \quad -\mathbb{E}[X_i^\top \Omega_e^{-1} X_i] \quad 0_{k_\beta \times 2} \Big)$$

The derivation of the approximate moments of the estimators can be done in a variety of ways. Our interest is $\widehat{\beta} - \beta_0$. To derive the first two moments we will do this by examining the elements of the first k_β elements of $\widehat{\theta}$. We make substantial use of selection matrices to so. For singling out $\widehat{\beta}$ here define $\tau^\top = \begin{pmatrix} I_{k_\beta} & 0_2 \end{pmatrix}$. We see that

$$\tau^\tau (\mathbb{E}q_i^{(1)}(\theta_0))^{-1} = -\left((\mathbb{E}[X_i^\top \Omega^{-1} X_i])^{-1} \quad 0 \quad 0 \right)$$

and that $\tau_i^{(j)} = \tau^\tau d_i^{(j)} \equiv d_{1i}^{(j)}$ returns the first k_β rows of $d_i^{(j)}$ and $\tau^\tau d_{i_1 i_2} = d_{1,i_1 i_2}$ and $\tau_{i_1 i_2 i_3} = \tau^\tau d_{i_1 i_2 i_3} = d_{1,i_1 i_2 i_3}$ are the first k_β rows of $d_{i_1 i_2}$ and $d_{i_1 i_2 i_3}$

The approximate bias of $\widehat{\beta}$ is given by

$$\begin{aligned} \text{ABIAS}\left[\widehat{\beta}\right] &= \mathbb{E}[\tau_{11}] \\ &= \mathbb{E}[\tau_1^{(1)} d_1] + \frac{1}{2}\overline{\tau_1^{(2)}}\mathbb{E}\left(d_1 \otimes d_1\right) \end{aligned}$$

$$\begin{aligned} q_{1i}^{(\beta\theta\theta)}(\theta) &= -\bigtriangledown_{\theta^\top} \left\{ (\bigtriangledown_{\sigma^{2\top}}(X_i^\top \Omega^{-1} X_i) \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix} \right) \right\} \\ &= -\left\{ \bigtriangledown_{\theta^\top}(\bigtriangledown_{\sigma^{2\top}}(X_i^\top \Omega^{-1} X_i) \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix} \right) \otimes I_{k_\theta} \right\} \\ &= -\left\{ \bigtriangledown_{\sigma^{2\top}\sigma^{2\top}}(X_i^\top \Omega^{-1} X_i) \right\} \left\{ (I_{k_\theta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix} \right\} \left\{ (I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix}) \otimes I_{k_\theta} \right\} \\ &= -\left\{ \bigtriangledown_{\sigma^{2\top}\sigma^{2\top}}(X_i^\top \Omega^{-1} X_i) \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix} \right\} \\ &= -\left\{ \bigtriangledown_{\sigma^{2\top}\sigma^{2\top}}(X_i^\top \Omega^{-1} X_i) \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4 \end{pmatrix} \right\} \end{aligned}$$

$$
\begin{aligned}
\bigtriangledown_{\sigma^{2\top}\sigma^{2\top}} X_i^\top \Omega^{-1} X_i &= \bigtriangledown_{\sigma^{2\top}} \left(X_i^\top \Omega_v^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right) + X_i^\top \Omega_e^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right) \right) \\
&= \left\{ \bigtriangledown_{\sigma^{2\top}} X_i^\top \Omega_v^{-1} X_i \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \otimes I_2 \right\} \\
&\qquad + \left\{ \bigtriangledown_{\sigma^{2\top}} X_i^\top \Omega_e^{-1} X_i \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right) \otimes I_2 \right\} \\
&= \left\{ X_i^\top \Omega_{vv}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \otimes I_2 \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ve}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \otimes I_2 \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ve}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right) \otimes I_2 \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ee}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right\} \left\{ \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right) \otimes I_2 \right\} \\
&= \left\{ X_i^\top \Omega_{vv}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ve}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ve}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ee}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right\} \\
&= \left\{ X_i^\top \Omega_{vv}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ve}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ve}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \right\} \\
&\quad + \left\{ X_i^\top \Omega_{ee}^{-1} X_i \right\} \left\{ I_{k_\beta} \otimes \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \right\}
\end{aligned}
$$

$$
\begin{aligned}
-q_{1i}^{(\beta\theta\theta)}(\theta) &= \left\{X_i^\top \Omega_{vv}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ee}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&= \left\{X_i^\top \Omega_{vv}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}\begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}\begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ee}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}0_{4\times k_\beta^2} & 0_{4\times 2k_\beta} & 0_{4\times 2k_\beta} & I_4\end{pmatrix}\right\} \\
&= \left\{X_i^\top \Omega_{vv}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2} & 0_{1\times 2k_\beta} & 0_{1\times 2k_\beta} & \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2} & 0_{1\times 2k_\beta} & 0_{1\times 2k_\beta} & \begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2} & 0_{1\times 2k_\beta} & 0_{1\times 2k_\beta} & \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ee}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2} & 0_{1\times 2k_\beta} & 0_{1\times 2k_\beta} & \begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}\end{pmatrix}\right\} \\
&= \left\{X_i^\top \Omega_{vv}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{1\top}\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{3\top}\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ve}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{2\top}\end{pmatrix}\right\} \\
&\quad + \left\{X_i^\top \Omega_{ee}^{-1} X_i\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{4\top}\end{pmatrix}\right\}
\end{aligned}
$$

$$
\mathbb{E}q_{1i}^{(\beta\theta\theta)} = -V_{Xvv}S_{11} - V_{Xve}S_{12} - V_{Xve}S_{21} - V_{Xee}S_{22}
$$

$$
\begin{aligned}
S_{11} &= \left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{1\top}\end{pmatrix}\right\}, \quad S_{12} = \left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{3\top}\end{pmatrix}\right\} \\
S_{21} &= \left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{2\top}\end{pmatrix}\right\}, \quad S_{22} = \left\{I_{k_\beta} \otimes \begin{pmatrix}0_{1\times k_\beta^2+4k_\beta} & \iota_4^{4\top}\end{pmatrix}\right\}
\end{aligned}
$$

$$
\begin{aligned}
q_{1i}^{(\sigma_v^2\theta)}(\theta) &= \begin{pmatrix}-X_i^\top \Omega_v^{-1} X_i & 2X_i^\top \Omega_{vv}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{ve}^{-1}(y_i - X_i\beta)\end{pmatrix} \\
&= \begin{pmatrix}q_{1i}^{(\sigma_v^2\beta)}(\theta) & q_{1i}^{(\sigma_v^2\sigma_v^2)}(\theta) & q_{1i}^{(\sigma_v^2\sigma_e^2)}(\theta)\end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
q_{1i}^{(\sigma_v^2\beta\theta)}(\theta) &= -\bigtriangledown_{\sigma^2}\left(X_i^\top \Omega^{-1} X_i\right)\left(I_{k_\beta} \otimes \begin{pmatrix}0_{2\times k_\beta} & I_2\end{pmatrix}\right) \\
&= -\left\{X_i^\top \Omega_{vv}^{-1} X_i\left(I_{k_\beta} \otimes \begin{pmatrix}1 & 0\end{pmatrix}\right) + X_i^\top \Omega_{ve}^{-1} X_i\left(I_{k_\beta} \otimes \begin{pmatrix}0 & 1\end{pmatrix}\right)\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix}0_{2\times k_\beta} & I_2\end{pmatrix}\right\} \\
&= -X_i^\top \Omega_{vv}^{-1} X_i\left(I_{k_\beta} \otimes \begin{pmatrix}0_{1\times 2} & \iota_2^{1\top}\end{pmatrix}\right) - X_i^\top \Omega_{ve}^{-1} X_i\left(I_{k_\beta} \otimes \begin{pmatrix}0_{1\times 2} & \iota_2^{2\top}\end{pmatrix}\right)
\end{aligned}
$$

$$q_{1i}^{(\sigma_v^2\sigma_v^2\theta)}(\theta) = \begin{pmatrix} -2X_i^\top \Omega_{vv}^{-1} X_i & 2X_i^\top \Omega_{vvv}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{vve}^{-1}(y_i - X_i\beta) \end{pmatrix}$$

$$\mathbb{E}q_{1i}^{(\sigma_v^2\sigma_v^2\theta)} = \begin{pmatrix} -2V_{Xvv} & 0_{k_\beta \times 2} \end{pmatrix}$$

$$V_{Xvv} = \mathbb{E}[X_1^\top \Omega_{vv}^{-1} X_1]$$

$$q_{1i}^{(\sigma_v^2\sigma_v^2\theta)}(\theta) = \begin{pmatrix} -2X_i^\top \Omega_{vv}^{-1} X_i & 2X_i^\top \Omega_{vvv}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{vve}^{-1}(y_i - X_i\beta) \end{pmatrix}$$

$$\mathbb{E}q_{1i}^{(\sigma_v^2\sigma_e^2\theta)} = \begin{pmatrix} -2V_{Xve} & 0_{k_\beta \times 2} \end{pmatrix}$$

$$V_{Xve} = \mathbb{E}[X_1^\top \Omega_{ve}^{-1} X_1]$$

$$\begin{aligned} q_{1i}^{(\sigma_e^2\theta)}(\theta) &= \begin{pmatrix} -X_i^\top \Omega_e^{-1} X_i & 2X_i^\top \Omega_{ve}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{ee}^{-1}(y_i - X_i\beta) \end{pmatrix} \\ &= \begin{pmatrix} q_{1i}^{(\sigma_e^2\beta)}(\theta) & q_{1i}^{(\sigma_e^2\sigma_v^2)}(\theta) & q_{1i}^{(\sigma_e^2\sigma_e^2)}(\theta) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} q_{1i}^{(\sigma_e^2\beta\theta)}(\theta) &= -\bigtriangledown_{\sigma^2}\left(X_i^\top \Omega^{-1} X_i\right)\left(I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix}\right) \\ &= -\left\{X_i^\top \Omega_{ev}^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}\right) + X_i^\top \Omega_{ee}^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix}\right)\right\}\left\{I_{k_\beta} \otimes \begin{pmatrix} 0_{2\times k_\beta} & I_2 \end{pmatrix}\right\} \\ &= -X_i^\top \Omega_{ev}^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times 2} & \iota_2^{1\top} \end{pmatrix}\right) - X_i^\top \Omega_{ee}^{-1} X_i \left(I_{k_\beta} \otimes \begin{pmatrix} 0_{1\times 2} & \iota_2^{2\top} \end{pmatrix}\right) \end{aligned}$$

$$q_{1i}^{(\sigma_e^2\sigma_v^2\theta)}(\theta) = \begin{pmatrix} -2X_i^\top \Omega_{ev}^{-1} X_i & 2X_i^\top \Omega_{evv}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{eve}^{-1}(y_i - X_i\beta) \end{pmatrix}$$

$$\mathbb{E}q_{1i}^{(\sigma_e^2\sigma_v^2\theta)} = \begin{pmatrix} -2V_{Xev} & 0_{k_\beta \times 2} \end{pmatrix}$$

$$q_{1i}^{(\sigma_e^2\sigma_e^2\theta)}(\theta) = \begin{pmatrix} -2X_i^\top \Omega_{ee}^{-1} X_i & 2X_i^\top \Omega_{eev}^{-1}(y_i - X_i\beta) & 2X_i^\top \Omega_{eee}^{-1}(y_i - X_i\beta) \end{pmatrix}$$

$$\mathbb{E}q_{1i}^{(\sigma_e^2\sigma_e^2\theta)} = \begin{pmatrix} -2V_{Xee} & 0_{k_\beta \times 2} \end{pmatrix}$$

Poisson model

Derivations for the Poisson model are simplified noting

$$\begin{aligned}\nabla_{\beta^\top}\frac{B_i}{\overline{\lambda}_i^j} &= \nabla_{\beta^\top}(B_i \otimes \overline{\lambda}_i^{-j})\\ &= \frac{\nabla_{\beta^\top}B_i}{\overline{\lambda}_i^j} + B_i \otimes \nabla_{\beta^\top}\overline{\lambda}_i^{-j}\\ &= \frac{\nabla_{\beta^\top}B_i}{\overline{\lambda}_i^j} - j\frac{B_i \otimes \nabla_{\beta^\top}\overline{\lambda}_i}{\overline{\lambda}_i^{j+1}}\end{aligned}$$

$$\lambda_{it}(\beta)^{(j)} = (-1)^j \lambda_{it}(\beta) x_{it}^{\otimes j\top}$$

and

$$\overline{\lambda_i(\beta)}^{(j)} = (-1)^j \frac{1}{T}\sum_{t=1}^{T} \lambda_{it}(\beta) x_{it}^{\otimes j\top} \equiv (-1)^j \overline{\lambda(\beta) x^{\otimes j}}_i^\top$$

$$\begin{aligned}
q_i^{(2)}(\beta) = {}& \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{\nabla_{\beta^\top}(\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top}\lambda}_i - \overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i)}{\overline{\lambda}_i^4}\right)\\
& - 2\overline{\lambda}_i T\overline{y}_i \left(\frac{\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top}\lambda}_i - \overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i\\
= {}& \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{\nabla_{\beta^\top}(\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top}\lambda}_i)}{\overline{\lambda}_i^4}\right) - \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{\nabla_{\beta^\top}(\overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i)}{\overline{\lambda}_i^4}\right)\\
& - 2\overline{\lambda}_i T\overline{y}_i \left(\frac{\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top}\lambda}_i - \overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i\\
= {}& \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{(\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top\beta^\top}\lambda}_i) + (\overline{\nabla_{\beta\beta^\top}\lambda}_i)(\overline{\nabla_{\beta^\top}\lambda}_i \otimes I_k))}{\overline{\lambda}_i^4}\right)\\
& - \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{(\overline{\nabla_{\beta\beta^\top}\lambda}_i)(\overline{\nabla_{\beta^\top}\lambda}_i \otimes I_k)) + (\overline{\nabla_\beta \lambda}_i)(\overline{\nabla_{\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^4}\right)\\
& - 2\overline{\lambda}_i T\overline{y}_i \left(\frac{\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top}\lambda}_i - \overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i\\
= {}& \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{(\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^4}\right) - \overline{\lambda}_i^2 T\overline{y}_i \left(\frac{(\overline{\nabla_\beta \lambda}_i)(\overline{\nabla_{\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^4}\right)\\
& - 2\overline{\lambda}_i T\overline{y}_i \left(\frac{\overline{\lambda}_i \overline{\nabla_{\beta\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i + 2\overline{\lambda}_i T\overline{y}_i \left(\frac{\overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i\\
= {}& T\overline{y}_i \left(\frac{\overline{\nabla_{\beta\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i}\right) - T\overline{y}_i \left(\frac{(\overline{\nabla_\beta \lambda}_i)(\overline{\nabla_{\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^2}\right)\\
& - 2T\overline{y}_i \left(\frac{\overline{\nabla_{\beta\beta^\top}\lambda}_i}{\overline{\lambda}_i^2}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i + 2T\overline{y}_i \left(\frac{\overline{\nabla_\beta \lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^3}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i
\end{aligned}$$

$$
\begin{aligned}
\frac{q_i^{(3)}(\beta)}{T\overline{y}_i} = &\frac{\overline{\nabla_{\beta\beta^\top\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i} - \frac{\overline{\nabla_{\beta\beta^\top\beta^\top}\lambda}_i \otimes \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^2} \\
&- \frac{\overline{\nabla_{\beta^\top\beta^\top}\lambda}_i \otimes \overline{\nabla_{\beta\beta^\top}\lambda}_i}{\overline{\lambda}_i^2} - \frac{\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i^2} \\
&+ \frac{(\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top\beta^\top}\lambda}_i) \otimes \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^3} \\
&- 2\frac{\overline{\nabla_{\beta\beta^\top}\lambda}_i \otimes \overline{\nabla_{\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i^2} - 2\frac{(\overline{\nabla_{\beta^\top}\lambda}_i \otimes \overline{\nabla_{\beta\beta^\top\beta^\top}\lambda}_i)(K_{kk} \otimes I_k)}{\overline{\lambda}_i^2} \\
&+ 4\frac{\overline{\nabla_{\beta\beta^\top}\lambda}_i \otimes \left(\overline{\nabla_{\beta^\top}\lambda}_i\right)^{\otimes 2}}{\overline{\lambda}_i^3} \\
&+ 2\frac{(\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i) \otimes \overline{\nabla_{\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i^3} \\
&+ 2\frac{\overline{\nabla_{\beta^\top}\lambda}_i^{\otimes 2} \otimes \overline{\nabla_{\beta\beta^\top}\lambda}_i}{\overline{\lambda}_i^3} \\
&+ 2\frac{\left[\overline{\nabla_{\beta^\top}\lambda}_i \otimes (\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top\beta^\top}\lambda}_i)\right](K_{kk} \otimes I_k)}{\overline{\lambda}_i^3} \\
&- 6\frac{(\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i) \otimes \left(\overline{\nabla_{\beta^\top}\lambda}_i\right)^{\otimes 2}}{\overline{\lambda}_i^4}
\end{aligned}
$$

$$
\begin{aligned}
\frac{q_i^{(3)}(\beta)}{T\overline{y}_i} = &\left(\frac{\overline{\nabla_{\beta\beta^\top\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i}\right) - \left(\frac{\overline{\nabla_{\beta\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i^2}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i \\
&- \left(\frac{\nabla_{\beta^\top}\left((\overline{\nabla_{\beta}\lambda}_i)(\overline{\nabla_{\beta^\top\beta^\top}\lambda}_i)\right)}{\overline{\lambda}_i^2}\right) \\
&+ \left(\frac{(\overline{\nabla_{\beta}\lambda}_i)(\overline{\nabla_{\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^3}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i \\
&- 2\left(\frac{\nabla_{\beta^\top}\left((\overline{\nabla_{\beta\beta^\top}\lambda}_i) \otimes \overline{\nabla_{\beta^\top}\lambda}_i\right)}{\overline{\lambda}_i^2}\right) \\
&+ 4\left(\frac{(\overline{\nabla_{\beta\beta^\top}\lambda}_i) \otimes \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^3}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i \\
&+ 2\left(\frac{\nabla_{\beta^\top}\left((\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i) \otimes \overline{\nabla_{\beta^\top}\lambda}_i\right)}{\overline{\lambda}_i^3}\right) \\
&- 6\left(\frac{(\overline{\nabla_{\beta}\lambda}_i \overline{\nabla_{\beta^\top}\lambda}_i) \otimes \overline{\nabla_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right) \otimes \overline{\nabla_{\beta^\top}\lambda}_i
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{q_i^{(3)}(\beta)}{T\overline{y}_i} = {} & \frac{\overline{\bigtriangledown_{\beta\beta^\top\beta^\top\beta^\top}\lambda}_i}{\overline{\lambda}_i} - \left(\frac{(\overline{\bigtriangledown_{\beta\beta^\top\beta^\top}\lambda}_i) \otimes \overline{\bigtriangledown_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^2}\right) \\
& - \left(\frac{(\overline{\bigtriangledown_{\beta\beta^\top}\lambda}_i)(\overline{\bigtriangledown_{\beta^\top\beta^\top}\lambda}_i \otimes I_k)}{\overline{\lambda}_i^2}\right) - \left(\frac{(\overline{\bigtriangledown_{\beta}\lambda}_i)(\overline{\bigtriangledown_{\beta^\top\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^2}\right) \\
& + \left(\frac{(\overline{\bigtriangledown_{\beta}\lambda}_i)(\overline{\bigtriangledown_{\beta^\top\beta^\top}\lambda}_i) \otimes \overline{\bigtriangledown_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^3}\right) \\
& - 2\left(\frac{((\overline{\bigtriangledown_{\beta\beta^\top}\lambda}_i) \otimes \overline{\bigtriangledown_{\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^2}\right) - 2\left(\frac{(\overline{\bigtriangledown_{\beta^\top}\lambda}_i \otimes \overline{\bigtriangledown_{\beta\beta^\top\beta^\top}\lambda}_i)(K_{kk} \otimes I_k)}{\overline{\lambda}_i^2}\right) \\
& + 4\left(\frac{(\overline{\bigtriangledown_{\beta\beta^\top}\lambda}_i) \otimes \overline{\bigtriangledown_{\beta^\top}\lambda}_i \otimes \overline{\bigtriangledown_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^3}\right) \\
& + 2\left(\frac{((\overline{\bigtriangledown_{\beta}\lambda}_i\overline{\bigtriangledown_{\beta^\top}\lambda}_i) \otimes \overline{\bigtriangledown_{\beta^\top\beta^\top}\lambda}_i)}{\overline{\lambda}_i^3}\right) + 2\left(\frac{\overline{\bigtriangledown_{\beta^\top}\lambda}_i \otimes \bigtriangledown_{\beta^\top}((\overline{\bigtriangledown_{\beta}\lambda}_i\overline{\bigtriangledown_{\beta^\top}\lambda}_i)(K_{kk} \otimes I_k)}{\overline{\lambda}_i^3}\right) \\
& - 6\left(\frac{\overline{(\bigtriangledown_{\beta}\lambda}_i\overline{\bigtriangledown_{\beta^\top}\lambda}_i) \otimes \overline{\bigtriangledown_{\beta^\top}\lambda}_i \otimes \overline{\bigtriangledown_{\beta^\top}\lambda}_i}{\overline{\lambda}_i^4}\right)
\end{aligned}
$$

References

Armstrong. M. and Hillier G., (1999), The Density of the Maximum Likelihood Estimator, Econometrica, 67(6), 1459 – 1470.

Bhattacharya R.N. and Ghosh J.K. (1978) On the Validity of the Formal Edgeworth Expansion, The Annals of Statistics, 6 (2), 434–451.

Butler R.W. (2007) Saddlepoint Approximations with Applications, Cambridge University Press, Cambridge, New York.

Daniels, H. E. (1954) Saddlepoint approximations in Statistics, Annals of Mathematical Statistics, 25, 631–650.

Daniels, H. E. (1983) Saddlepoint approximations for Estimating Equations, 70, 89–96.

Davidson. R, and MacKinnon, J.G. (2000) Bootstrap tests: how many bootstraps? Econometric Reviews 19, 5568.

Easton, G.S. and E. Ronchetti (1986) General saddlepoint approximations with applications to L statistics, Journal of the American Statistical Association, 81, 420–430.

Feuerverger A. (1989) On The Empirical Saddlepoint Approximation, Biometrika 76(3), 457–464.

Hall, Peter (1992) The Bootstrap and the Edgeworth Expansion, Springer–Verlag, New York, NY.

Hansen, B. E. (2000) Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions, Manuscript, University of Wisconsin.

Hengartner, Nicolas W. and Marten H. Wegkamp (2004) Second order asymptotics for M-estimators under non-standard conditions, in Javier Rojo and Victor Prez-Abreu, eds., IMS Lecture Notes, 107-124.

Holly, A. and P.C.B. Phillips, P.C.B. (1979) A Saddlepoint Approximation to the Distribution of the k-Class Estimator of a Coefficient in a Simultaneous System" (with A.), Econometrica 47(6), 1527–1548.

Kundhi, G. and Rilstone Paul (2008) The Third Order Bias of Nonlinear Estimators, Communications in Statistics 37(16), 2617–2633.

Lahiri, S.N. (1992) Two term Edgeworth Expansions for M-Estimators of a Regression Coefficient, Journal of Multivariate Analysis, 43, 125–132.

Lahiri, S.N. (1994) Edgeworth expansions for M-Estimators and bootstrap approximation for multivariate studentized M-estimators. Sankhya, Series A 56, 201–226.

Lieberman, Offer (1994) On the Approximation of Saddlepoint Expansions in Statistics, Econometric Theory, 10, 900–916.

Lugannani, R. and S.O. Rice (1980) Saddlepoint approximations for the distributions of the sum of random variables, Advances in Applied Probability 12, 475–490.

MacRae, E.C. (1974) Matrix derivatives with an application to an adaptive linear decision problem, Annals of Statistics, 2, 337–346.

McCullagh, P. (1987) Tensor Methods in Statistics, Chapman and Hall, New York, NY.

Monti, A.C. and E. Ronchetti (1993) On the relationship between empirical likelihood and empirical approximations for multivariate M-estimators, Biometrika 80, 329–338.

Newey, Whitney K. and Richard Smith (2004) Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, Econometrica 72(1) 219–255.

Ohman–Strickland, Pamela and George Casella (2002) Approximate and estimated saddlepoint approximations, Canadian Journal of Statistics 30(1), 97–108.

Petrov, V.V. (1975) Sums of Independent Random Variables, Springer–Verlag, NY.

Phillips, P.C.B. (1978) Edgeworth and Saddlepoint Approximations in a First Order Non-Circular Autoregression, Biometrika 65(1) 91–98.

Phillips, P.C.B. and Park J.Y. (1988) On the Formulation of Wald Tests of Nonlinear Restrictions, Econometrica 56(5) 1065–1083.

Rilstone, Paul (2015) New methods for deriving the approximate moments of non-linear estimators, working paper, York University.

Rilstone, Paul, V.K. Srivastava and Aman Ullah (1996) The second–order bias and mean squared error of nonlinear estimators, Journal of Econometrics 75(2), 369–395.

Rilstone, Paul and Aman Ullah (2005) Corrigendum to: The second–order bias and mean squared error of non-linear estimators, Journal of Econometrics 124(1), 203–204.

Ronchetti, E. and A.H. Welsh (1994) Empirical saddlepoint approximations for multivariate M–estimators, Journal of the Royal Statistical Society series B, 56, 313–326.

Spady, Richard H. (1991) Saddlepoint approximations for regression models, Biometrika 78(4), 879–889.

Sowell, F. (2007) The Empirical Saddlepoint Approximation for GMM Estimators, MRPA paper no. 3356.

Taniguchi, Masanobu and Madan L. Puri (1996) Valid Edgeworth Expansions of M-Estimators in Regression Models with Weakly Dependent Residuals, Econometric Theory, 12, 331–346.

Ullah, Aman (2004) Finite Sample Econometrics, Oxford University Press, NY.

Hengartner, Nicolas W. and Marten H. Wegkamp (2004) Second order asymptotics for M-estimators under non-standard conditions, in Javier Rojo and Victor Prez-Abreu, eds., The First Erich L. Lehmann Symposium, Beachwood, OH: Institute of Mathematical Statistics, 107–124.