ECOM40006/ECOM90013 Econometrics 3
Department of Economics
University of Melbourne

Generalized Least Squares Re-Visited

Semester 1, 2025

Version: March 20, 2025

Contents

1 Introduction

Before moving on to discuss other theories of estimation, it will prove helpful to briefly discuss an important extension to the theory of least squares and that is how to handle relaxation of the assumption of a *scalar covariance structure*. Recall that in the linear regression model

$$y = X\beta + u, \tag{1}$$

we typically assume that X has full column rank, that $\operatorname{plim} n^{-1}X'X = Q > 0$, that $\mathrm{E}[u \mid X] = 0$, and that $\operatorname{Var}[u \mid X] = \sigma^2 I_n$. This last is an assumption that the disturbances have a scalar covariance structure. In this handout we briefly wish to examine how to proceed if, instead, $\operatorname{Var}[u \mid X] = \Omega > 0$.

2 The Gauss-Markov Theorem

As a preliminary consideration, let's remind ourselves of the very important Gauss-Markov Theorem, which states:

Theorem 1. *In the model 1, with its various accompanying assumptions and, specifically, with scalar covariance matrix $\Omega = \sigma^2 I_n$, among the class of linear unbiased estimators for β, OLS is best, where the OLS estimator is $\hat{\beta} = (X'X)^{-1}X'y$.*

Remark 1. To be clear, a linear estimator is of the form $\tilde{\beta} = Ay$, where the elements of A are not functions of the elements of y, so we can either treat them as fixed or, at least, fixed conditional on the values of X. An unbiased estimator satisfies the condition $\mathrm{E}\left[\tilde{\beta} \mid X\right] = \beta$. An estimator that is best has the smallest variance among the class of estimators, in this case the class of linear, unbiased estimators.

There are lots of proofs of this result. Here is one that you mightn't have seen before.

Proof. As $\mathrm{E}\left[\tilde{\beta} \mid X\right] = \beta$ it follows that

$$\mathrm{E}\left[Ay \mid X\right] = A\,\mathrm{E}\left[y \mid X\right] = AX\beta = \beta \implies AX = I_k.$$

Moreover, we see that

$$\begin{aligned}
\mathrm{Var}\left[\tilde{\beta} \mid X\right] &= \mathrm{E}\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' \mid X\right] \\
&= \mathrm{E}\left[(Ay - \beta)(Ay - \beta)' \mid X\right] \\
&= \mathrm{E}\left[(A(X\beta + u) - \beta)(A(X\beta + u) - \beta)' \mid X\right] \\
&= \mathrm{E}\left[Auu'A' \mid X\right] = A\,\mathrm{E}\left[uu' \mid X\right]A' = A(\sigma^2 I_n)A',
\end{aligned}$$

where we have used the facts that $AX = I_k$ and that, conditional on X, we can treat A as fixed. Now, recall that for X of full column rank $I_n = P_X + M_X$, where $P_X = X(X'X)^{-1}X'$ and $M_X = I_n - P_X$ are both positive semi-definite matrices. This implies that, for any non-zero matrix C of dimension $p \times n$, $p \leq n$, $CP_XC' \geq 0$ and $CM_XC' \geq 0$. Therefore,

$$\mathrm{Var}\left[\tilde{\beta} \mid X\right] = A(\sigma^2 I_n)A' = \sigma^2 A(P_X + M_X)A' \geq \sigma^2 AP_XA' = \sigma^2(X'X)^{-1} = \mathrm{Var}\left[\hat{\beta} \mid X\right].$$

In order for $\mathrm{Var}\left[\tilde{\beta} \mid X\right] = \mathrm{Var}\left[\hat{\beta} \mid X\right]$, we require $AM_XA' = 0$, which implies that there must exist some square matrix B such that $A = BX'$. Post-multiplying both sides of this last equality by X we obtain $AX = BX'X \implies I_k = BX'X \implies B = (X'X)^{-1}$. That is, in order for $\tilde{\beta}$ to have a variance as small as that of $\hat{\beta}$, A must equal $(X'X)^{-1}X'$ which, in turn, implies that $\tilde{\beta} = \hat{\beta}$. □

Remark 2. An alternative conclusion to the proof follows on noting that if $A = BX'$ then $AP_XA' = BX'P_XXB' = BX'XB'$, which will only equal $(X'X)^{-1}$ if $B = (X'X)^{-1}$, meaning that $A = (X'X)^{-1}X'$ and hence $\tilde{\beta} = \hat{\beta}$ as before.

Long story short, should you wish to restrict yourself to the class of linear unbiased estimators when estimating β in model (1), then you can't get a smaller variance than that given by OLS. Note: If you are prepared to live with a little bit of bias then you may be able to make sufficiently substantial gains in variance reduction that you get a reduction in mean squared error (MSE) over that of OLS, where

$$\mathrm{MSE} = (\mathrm{Bias})^2 + \mathrm{Variance},$$

but that is a story for another time.

3 Generalized Least Squares

Now suppose that $\mathrm{Var}\left[u \mid X\right] = \Omega > 0$, where Ω is not necessarily a scalar matrix. By way of example, two special cases are that of (i) heteroskedasticity, where $\Omega = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ with $\sigma_j^2 > 0$ for all $j = 1, \ldots, n$, and (ii) autocorrelation. Now there are lots of different models of autocorrelated disturbances that one might consider but simple first order autocorrelation is typically characterized along the following lines. Starting with the linear model

$$y_t = x_t'\beta + u_t, \quad t = 1, \ldots, T, \tag{2}$$

where we assume that $X = [x_1, \ldots, x_T]'$ has full column rank, and that $\operatorname{plim} T^{-1}X'X = Q > 0$. We further assume that the disturbances are correlated with their immediate past according to

$$u_t = \rho u_{t-1} + \epsilon_t, \quad \mathrm{abs}\left(\rho\right) < 1, \tag{3}$$

with $\mathrm{E}\left[\epsilon \mid X\right] = 0$ and $\mathrm{Var}\left[\epsilon \mid X\right] = \sigma_\epsilon^2 I_T$, where $\epsilon = [\epsilon_1, \ldots, \epsilon_T]'$. In this case it is reasonably straightforward to show that

$$\Omega = \frac{\sigma_\epsilon^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \ldots & \rho^{T-2} \\ \rho^2 & \rho & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \rho & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \ldots & \rho^2 & \rho & 1 \end{bmatrix}.$$

Note that I was silent about the range of values for t in (3). To have the equation valid for $t = 1, \ldots, T$, we need to make an assumption about ϵ_0. Alternatively, we might choose to work with an effective sample of size $T - 1$, so that $t = 2, \ldots, T$, only. Not surprisingly, such considerations have generated years of literature — one would think that academics have nothing better to do than write papers — but is irrelevant for the point that I am trying to make here, which is simply that it isn't hard to generate non-scalar covariance matrices for the disturbances.

The first observation to make is that if Ω is known then we could simply pre-multiply our data by $\Omega^{-1/2}$ to obtain

$$y^* = X^*\beta + u^*,$$

where a * superscript indicates that the variable has been pre-multiplied by $\Omega^{-1/2}$; for example, $y^* = \Omega^{-1/2}y$, and so on. In this model, all of the classical assumptions can be written in terms of the starred variables and, if the original variables satisfied the assumptions then so too will the starred variables, except that now $\mathrm{Var}\left[u^* \mid X^*\right] = I_n$, which is a scalar covariance structure. This means that we can apply OLS to the starred model and we will obtain a consistent and asymptotically efficient estimator. The resulting estimator is known as *generalized least squares* (GLS). It takes the form

$$\check{\beta}_n = (X^{*'}X^*)^{-1}X^{*'}y^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \tag{4}$$

To be clear, our model here is

$$y = X\beta + u,$$

and we shall make the following assumptions:

1. X has full column rank,
2. $\mathrm{E}[u \mid X] = 0$ and $\mathrm{Var}[u \mid X] = \mathrm{E}[uu' \mid X] = \Omega > 0$,
3. $\operatorname{plim} n^{-1}X'\Omega^{-1}X = Q_\Omega > 0$

Then,

$$n^{1/2}(\check{\beta}_n - \beta) \xrightarrow{d} \mathrm{N}(0, Q_\Omega^{-1}).$$

By way of contrast, OLS in this non-scalar model, requires both $\operatorname{plim} n^{-1}X'X = Q > 0$ and $\operatorname{plim} n^{-1}X'\Omega X = Q_\Omega > 0$. Then we see that

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathrm{N}(0, Q^{-1}Q_\Omega Q^{-1}).$$

We can compare the two variances and we see that

$$\begin{aligned}
\mathrm{Var}\left[\hat{\beta}_n\right] - \mathrm{Var}\left[\check{\beta}_n\right] &= (X'X)^{-1}X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1} \\
&= (X'X)^{-1}[X'\Omega X - X'X(X'\Omega^{-1}X)^{-1}X'X](X'X)^{-1} \\
&= (X'X)^{-1}X'\Omega^{1/2}[I_n - \Omega^{-1/2}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1/2}]\Omega^{1/2}X(X'X)^{-1} \\
&= (X'X)^{-1}X'\Omega^{1/2}M_{\Omega^{-1/2}X}\Omega^{1/2}X(X'X)^{-1}
\end{aligned}$$

Because $M_{\Omega^{-1/2}X} \geq 0$ so too $AM_{\Omega^{-1/2}X}A' \geq 0$ for any matrix A of full row rank. Setting $A = (X'X)^{-1}X'\Omega^{1/2}$, which clearly has full row rank given our assumptions, we see that

$$\mathrm{Var}\left[\hat{\beta}_n\right] - \mathrm{Var}\left[\check{\beta}_n\right] \geq 0.$$

We conclude that the generalized least squares estimator is relatively efficient when compared with OLS in this model, most of the time. You might wonder why the inequality is not strict. It has to do with whether or not the columns of X span the space spanned by the columns of Ω. This can only happen if Ω has multiple eigenvalues that are equal. The papers by Amemiya (1983) and Balestra (1983) provide a beautiful example of where OLS coincides exactly with GLS, even though $\Omega \neq \sigma^2 I_n$; see also Amemiya (1985, pp.199–200).

The one final thing worth noting at this time is that the implicit optimization problem leading to $\check{\beta}_n$ is

$$\check{\beta}_n = \underset{\beta}{\operatorname{argmin}}(y^* - X^*\beta)'(y^* - X^*\beta) = \underset{\beta}{\operatorname{argmin}}(y - X\beta)'\Omega^{-1}(y - X\beta).$$

We see that it minimizes a different, although clearly related, criterion function to that minimized by OLS. It is sometimes described as minimizing the sum of squared errors *in the metric of* Ω. It is a weighted sum of squares that is minimized and that is sometimes why GLS is sometimes generically referred to as *weighted least squares*. This important notion of minimizing a weighted sum of squares is one that we will encounter again shortly.

4 Feasible Generalized Least Squares

The previous section assumed that Ω was known, which is unrealistic in a situation where you need to estimate β or, indeed, pretty much any other situation. In this section we

explore how one might proceed if Ω were not known. Historically, if $\Omega \neq \sigma^2 I_n$ then there were two approaches to consistent estimation of β: OLS and *feasible GLS* estimators (FGLS). Specifically, in the model (1), the OLS estimator is

$$\hat{\beta}_n = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$$

and, subject to assumptions of the form (i) $\operatorname{plim} n^{-1}X'X = Q > 0$ and (ii) $\operatorname{plim} n^{-1}X'u = 0$, consistency of $\hat{\beta}_n$ for β follows. It is the second of these conditions that is of interest here because our usual approach to establishing this probability limit is to establish mean squared convergence. Now, our standard assumption $\mathrm{E}[u \mid X] = 0$ will get the mean part of this requirement sorted.[1] As for the variance, we see that

$$\operatorname{Var}[X'u] = \mathrm{E}[X'uu'X] = \mathrm{E}_X[X'\,\mathrm{E}[uu' \mid X]\,X] = \mathrm{E}_X[X'\Omega X],$$

and so, we needed to add an assumption of the form: $\operatorname{plim} n^{-1}X'\Omega X = Q_\Omega > 0$; mean squared convergence then follows. And so we see that the problem was not the lack of consistency of the estimator but rather the fact that

$$\lim_{n\to\infty} \operatorname{Var}\left[\hat{\beta}_n\right] = Q^{-1}Q_\Omega Q^{-1}$$

differs from that of OLS in a homoskedastic, serially uncorrelated model. Consequently, the estimated standard errors and t-statistics that typically accompany regression output are wrong if the disturbance covariance matrix is non-scalar. Moreover, correction of these same standard errors will require some way to model $X'\Omega X$.

For a very long time it was thought that modelling $X'\Omega X$ meant being able to model Ω directly. From early papers like Cochrane and Orcutt (1949) right through the 1970s, the literature was full of models for serially correlated disturbances (which impacts upon the off-diagonal elements of Ω), for heteroskedastic disturbances (which impacts upon the off-diagonal elements of Ω), and for combinations of both. The key feature of these models is that whatever was leading to the non-scalar covariance structure could be modelled using a relatively small number of parameters. By way of contrast, if one tried to allow for arbitrary heteroskedasticity, then the combination of k regression coefficients plus a different variance for each observation meant that there were $n + k$ parameters in total, which is more parameters than we had observations and so there is no way of doing that, let alone doing it consistently (which requires an accumulation of information, something that can only happen if you have increasingly more observations than parameters as the sample size grows).[2] If such a model can be postulated then it might be possible to use a multi-step estimator, where the first step was typically used to fit the regression model using an estimator that was consistent for the regression coefficients but is typically inefficient because it is ignoring the covariance structure of the disturbances, such as OLS. Using the OLS residuals, one then fitted the model for the covariance structure. The fitted values from this process are then estimates of the variances for each observation. The next

[1] This is a strong assumption because it precludes the possibility of lagged dependent variables as regressors if the disturbances are serially correlated. Yet, lagged dependent variables are commonplace in dynamic regression models and so this requirement may be unreasonable. We will not have much to say about dynamic models in these notes but they bring with them their own special requirements which demand great care and attention.

[2] Strictly this last parenthetic remark allows for the possibility of the number of parameters growing with the sample size, but at a slower rate than the sample size is growing. This is a nuance that we will concern ourselves with no further.

step is to use these variance estimates to fit a *feasible generalized least squares estimator*, essentially by pre-multiplying (1) by $\widehat{\Omega}^{-1/2}$ to obtain

$$y^\dagger = X^\dagger \beta + u^\dagger, \tag{5}$$

where a † superscript indicates that the variable has been pre-multiplied by $\widehat{\Omega}^{-1/2}$; for example, $y^\dagger = \widehat{\Omega}^{-1/2} y$, and so on. If you feel more comfortable with sigma notation then all that we are doing is estimating the following model by OLS:

$$\frac{y_i}{\hat{\sigma}_i} = \frac{x_i'}{\hat{\sigma}_i}\beta + \frac{u_i}{\hat{\sigma}_i}.$$

We see that, under suitable conditions $\operatorname{plim} n^{-1} X^{\dagger\prime} u^\dagger = \operatorname{plim} n^{-1} X' \widehat{\Omega}^{-1} u = 0$ and so OLS will be both consistent and efficient in the transformed model. Unravelling this transformations brings us back to the FGLS, which is 'feasible' because we can actually do it, even if we don't know the true value of Ω. This estimator is then

$$\beta_n^\dagger = (X^{\dagger\prime} X^\dagger)^{-1} X^{\dagger\prime} y^\dagger = (X' \widehat{\Omega}^{-1} X)^{-1} X' \widehat{\Omega}^{-1} y.$$

There are two important points to be aware of here. First, the procedure described above is one that lends itself to iteration. For example, why not adopt the following procedure?

1. Use the residuals from (5) to construct a new estimate of Ω, $\Omega^\dagger$ say.
2. Use $\Omega^\dagger$ to construct a new FGLS estimate of β.
3. Create a new set of residuals based on the new estimate of β. Return to 1.

Estimators so formed are known as *iterated FGLS estimators*.[3] There is some evidence that iterating for one step does have some benefits. The second point is that because we are using $\widehat{\Omega}$ instead of Ω in a GLS estimator, the estimators finite sample performance has scope to vary along with our ability to estimate Ω. So, although asymptotically, we may be able to generate FGLS estimators that are the same as GLS, in finite samples their performance may be quite different. Indeed, we are unable to prove that, in general, FGLS estimators will dominate OLS on any reasonable grounds. While they frequently do, sadly there is evidence to the contrary in certain cases.

At the start of the 1980s much work focussed on the benefits of using OLS, but with corrected standard errors. In particular, it was found that even though Ω could not be estimated directly without some parametric model that avoided the so-called *curse of dimensionality*,[4] $X'\Omega X$ could be. For example, as shown by White (1980),[5] if confronted

[3] Iterating is a common practice once one starts working with non-linear estimators and you should be aware that the FGLS estimators are inherently non-linear functions of the data, y, which is in sharp contrast to OLS and GLS, which remain linear functions of y.

[4] The curse of dimensionality arises when you have more parameters than observations. The case of arbitrary heteroskedasticity in the linear regression model is a perfect statistical example. It arises in a variety of other computational fields as well. See, for example, the discussion at https://en.wikipedia.org/wiki/Curse_of_dimensionality.

[5] It should be noted that Eicker (1963, 1967) and Hinkley (1977) are among those who had discussed this idea before White (1980) had introduced it into the econometrics literature. For this reason, the estimator of the covariance matrix of the OLS estimator in linear models with arbitrary heteroskedasticity is sometimes known as the Eicker-White heteroskedasticity-consistent covariance matrix estimator, rather than just Whites estimator. Use of this estimators is sometimes referred to in econometrics as White-washing the standard errors.

by arbitrary heteroskedasticity of an unspecified form, so that $\Omega = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, then a consistent estimator for

$$\operatorname*{plim}_{n\to\infty} n^{-1} X'\Omega X = \operatorname*{plim}_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sigma_i^2 x_i x_i' = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathrm{E}_X \left[u_i^2 x_i x_i' \right]$$

is given by

$$\operatorname*{plim} n^{-1} X'\widehat{\Omega} X = \operatorname*{plim}_{n\to\infty} n^{-1} \sum_{i=1}^{n} e_i^2 x_i x_i',$$

where $\sigma_i^2 = \mathrm{E}\left[u_i^2 \mid X\right]$, $\widehat{\Omega} = \text{diag}(e_1^2, \ldots, e_n^2)$, with e_i being the OLS residual for the i-th observation from (1), and x_i' denotes the i-th row of X. This approach works because when we think about $X'\Omega X = \sum_{i=1}^{n} \mathrm{E}_X \left[u_i^2 x_i x_i'\right]$, we have a sum of $k \times k$ matrices (things whose dimensions don't change as the sample size changes) that we are able to accumulate information on as the sample size changes.

For completeness we note that we have shown that

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathrm{N}\left(0, Q^{-1} Q_\Omega Q^{-1}\right) \tag{6}$$

and that an asymptotic approximation to the true sampling distribution of $\hat{\beta}_n$ is

$$\hat{\beta}_n \underset{a}{\sim} \mathrm{N}\left(\beta, (X'X)^{-1} X'\widehat{\Omega} X (X'X)^{-1}\right) \tag{7}$$

The heteroskedastic-robust standard errors produced using this formula have proved to be of considerable practical importance. We note that the form of the covariance matrix is described to be that of a sandwich estimator: two copies of a matrix $(X'X)^{-1}$ wrapped around either side of another matrix $(X'\widehat{\Omega}X)$. Sandwich estimators of covariance matrices are symptomatic of the the coefficient estimator, $\hat{\beta}_n$ in this case, being inefficient. We shall encounter sandwich estimators frequently in what follows.

As a second example, it is possible to construct an FGLS estimator that will be consistent in the simultaneous presence of both heteroskedasticity and autocorrelation. Again, the key idea is that we make no attempt to estimate Ω directly, rather $n^{-1}X'\widehat{\Omega}X$ is the quantity that will converge to what we need. For suitably chosen $\widehat{\Omega}$ this is known as a *heteroskedasticity and autocovariance consistent* (HAC) covariance estimator. The specific details of how best to construct such an estimator are beyond the scope of what we are aiming for. A useful place to learn more about them is Davidson and MacKinnon (2004, Section 9.3) and the references cited therein, although there the discussion is couched in terms of GMM estimation, which we consider shortly. To give a flavour of the key result, the idea lies in thinking of

$$\frac{1}{n} X'\Omega X = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}\left[u_i u_j x_i x_j'\right],$$

where x_k', $k = 1, \ldots, n$ denotes the k-th row of X. We shall break Ω up into its diagonal elements, or variance terms, which we shall denote by $\Gamma(0)$ and then the off-diagonal elements denoted by

$$\sum_{j=1}^{n-1} (\Gamma(j) + \Gamma(j)'),$$

where $\Gamma(j)'$ denotes the transpose of $\Gamma(j)$ which is defined according to

$$\Gamma(j) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} \mathrm{E}\left[u_t u_{t-j} x_t x_{t-j}'\right], & \text{for } j \geq 0, \\ \frac{1}{n}\sum_{t=-j+1}^{n} \mathrm{E}\left[u_{t+j} u_t x_{t+j} x_t'\right], & \text{for } j < 0. \end{cases}$$

This is all made feasible through the use of of OLS residuals, $\hat{u}_t$, which come from a consistently although inefficiently estimated model. For example,

$$\widehat{\Gamma}(j) = \frac{1}{n}\sum_{t=j+1}^{n} \hat{u}_t \hat{u}_{t-j} x_t x_{t-j}'$$

for any $j \geq 0$. Where this seemingly simple idea breaks down is that $\widehat{\Gamma}(j)$ is not a consistent estimator for $\Gamma(j)$ for arbitrary j. To see this note that for j very close to n the sums have very few terms in them, too few for a law of large numbers to be applicable. The way around this is to truncate the sum, so that j no longer runs to $n-1$, but that brings with it further complications that we will not pursue here. Suffice to say, these complications have been sorted out and HAC estimators are in widespread use; see, for example, Newey and West (1987).

References

Amemiya, T. (1983). Partially generalized least squares and two-stage least squares estimators. *Journal of Econometrics* 23(2), 275–283, doi:https://doi.org/10.1016/0304-4076(93)90082-G. 4

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts. 4

Balestra, P. (1983). A note on amemiya's partially generalized least squares. *Journal of Econometrics* 23(2), 285 – 290, doi:https://doi.org/10.1016/0304-4076(93)90083-H. 4

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association* 44(245), 32–61. 5

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford University Press, New York. 7

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *The Annals of Mathematical Statistics* 34(2), 447–456. 6

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. Le Cam and J. Neyman, editors, volume 1, 59–82, University of California Press, Berkeley, CA. 6

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* 19(3), 285–292, ISSN 00401706. 6

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708. 8

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838. 6