Solutions to Chapter 12 Problems

12.1. a. Take the conditional expectation of equation (12.4) with respect to $\mathbf{x}$, and use $\mathrm{E}(u|\mathbf{x}) = 0$:

$$\begin{aligned}\mathrm{E}\{[y - m(\mathbf{x},\boldsymbol{\theta})]^2|\mathbf{x}\} &= \mathrm{E}(u^2|\mathbf{x}) + 2[m(\mathbf{x},\boldsymbol{\theta}_o) - m(\mathbf{x},\boldsymbol{\theta})]\mathrm{E}(u|\mathbf{x}) + \mathrm{E}\{[m(\mathbf{x},\boldsymbol{\theta}_o) - m(\mathbf{x},\boldsymbol{\theta})]^2|\mathbf{x}\} \\ &= \mathrm{E}(u^2|\mathbf{x}) + 0 + [m(\mathbf{x},\boldsymbol{\theta}_o) - m(\mathbf{x},\boldsymbol{\theta})]^2 \\ &= \mathrm{E}(u^2|\mathbf{x}) + [m(\mathbf{x},\boldsymbol{\theta}_o) - m(\mathbf{x},\boldsymbol{\theta})]^2.\end{aligned}$$

The first term does not depend on $\boldsymbol{\theta}$ and the second term is clearly minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ for any $\mathbf{x}$. Therefore, the parameters of a correctly specified conditional mean function minimize the squared error conditional on any value of $\mathbf{x}$.

b. Part a shows that

$$\mathrm{E}\{[y - m(\mathbf{x},\boldsymbol{\theta}_o)]^2|\mathbf{x}\} \leq \mathrm{E}\{[y - m(\mathbf{x},\boldsymbol{\theta})]^2|\mathbf{x}\}, \text{ all } \boldsymbol{\theta} \in \boldsymbol{\Theta}, \text{ all } \mathbf{x} \in \mathcal{X}.$$

If we take the expected value of both sides – with respect the the distribution of $\mathbf{x}$, of course – an apply iterated expectations, we conclude

$$\mathrm{E}\{[y - m(\mathbf{x},\boldsymbol{\theta}_o)]^2\} \leq \mathrm{E}\{[y - m(\mathbf{x},\boldsymbol{\theta})]^2\}, \text{ all } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

In other words, if we know $\boldsymbol{\theta}_o$ solves the population minimization problem conditional on any $\mathbf{x}$, then it also solves the unconditional population problem. Of course, conditional on a particular value of $\mathbf{x}$, $\boldsymbol{\theta}_o$ would usually not be the unique solution. (For example, in the linear case $m(\mathbf{x},\boldsymbol{\theta}) = \mathbf{x}\boldsymbol{\theta}$, any $\boldsymbol{\theta}$ such as that $\mathbf{x}(\boldsymbol{\theta}_o - \boldsymbol{\theta}) = 0$ sets $m(\mathbf{x},\boldsymbol{\theta}_o) - m(\mathbf{x},\boldsymbol{\theta})$ to zero.) Uniqueness of $\boldsymbol{\theta}_o$ as a population minimizer is realistic only after we integrate out $\mathbf{x}$ to obtain $\mathrm{E}\{[y - m(\mathbf{x},\boldsymbol{\theta})]^2\}$.

12.2. a. Since $u = y - \mathrm{E}(y|\mathbf{x})$,

$$\mathrm{Var}(y|\mathbf{x}) = \mathrm{Var}(u|\mathbf{x}) = \mathrm{E}(u^2|\mathbf{x})$$

because $\mathrm{E}(u|\mathbf{x}) = 0$. So $\mathrm{E}(u^2|\mathbf{x}) = \exp(\alpha_o + \mathbf{x}\boldsymbol{\gamma}_o)$.

b. If we knew the $u_i = y_i - m(\mathbf{x}_i, \boldsymbol{\theta}_o)$, then we could do a nonlinear regression of u_i^2 on $\exp(\alpha + \mathbf{x}\boldsymbol{\gamma})$ and just use the asymptotic theory for nonlinear regression. The NLS estimators of α and $\boldsymbol{\gamma}$ would then solve

$$\min_{\alpha,\boldsymbol{\gamma}} \sum_{i=1}^{N} [u_i^2 - \exp(\alpha + \mathbf{x}_i\boldsymbol{\gamma})]^2.$$

The problem is that $\boldsymbol{\theta}_o$ is unknown. When we replace $\boldsymbol{\theta}_o$ with its NLS estimator, $\hat{\boldsymbol{\theta}}$ – that is we replace u_i^2 with $\hat{u}_i^2$, the squared NLS residuals – we are solving the problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \sum_{i=1}^{N} \{[y_i - m(\mathbf{x}_i, \hat{\boldsymbol{\theta}})]^2 - \exp(\alpha + \mathbf{x}_i\boldsymbol{\gamma})\}^2.$$

This objective function has the form of a two-step M-estimator in Section 12.4. Since $\hat{\boldsymbol{\theta}}$ is generally consistent for $\boldsymbol{\theta}_o$, the two-step M-estimator is generally consistent for α_o and $\boldsymbol{\gamma}_o$ (under weak regularity and identification conditions). In fact, $\sqrt{N}$-consistency of $\hat{\alpha}$ and $\hat{\boldsymbol{\gamma}}$ holds very generally.

c. We now estimate $\boldsymbol{\theta}_o$ by solving

$$\min_{\theta} \sum_{i=1}^{N} [y_i - m(\mathbf{x}_i, \boldsymbol{\theta})]^2 / \exp(\hat{\alpha} + \mathbf{x}_i\hat{\boldsymbol{\gamma}}),$$

where $\hat{\alpha}$ and $\hat{\boldsymbol{\gamma}}$ are from part b. The general theory of WNLS under WNLS.1 to WNLS.3 can be applied.

d. Using the definition of v, write $u^2 = \exp(\alpha_o + \mathbf{x}\gamma_o)v^2$. Taking logs gives $\log(u^2) = \alpha_o + \mathbf{x}\boldsymbol{\gamma}_o + \log(v^2)$. Now, if v is independent of $\mathbf{x}$, so is $\log(v^2)$. Therefore, $\mathrm{E}[\log(u^2)|\mathbf{x}] = \alpha_o + \mathbf{x}\boldsymbol{\gamma}_o + \mathrm{E}[\log(v^2)|\mathbf{x}] = \alpha_o + \mathbf{x}\boldsymbol{\gamma}_o + \kappa_o$, where $\kappa_o \equiv \mathrm{E}[\log(v^2)]$. So, if we could observe the u_i, and OLS regression of $\log(u_i^2)$ on $1, \mathbf{x}_i$ would be consistent for $(\alpha_o + \kappa_o, \boldsymbol{\gamma}_o)$; in fact, it would be unbiased. By two-step estimation theory, consistency still holds if u_i is replaced with $\hat{u}_i$, by essentially the same argument in part b. So, if $m(\mathbf{x}, \boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, we can carry out a weighted NLS procedure without ever doing nonlinear estimation.

e. If we have misspecified the variance function – or, for example, we use the approach in part d but v is not independent of $\mathbf{x}$ – then we should use a fully robust variance-covariance matrix in equation (12.60) with $\hat{h}_i = \exp(\hat{\alpha} + \mathbf{x}_i\hat{\boldsymbol{\gamma}})$.

12.3. a. The approximate elasticity is $\partial \log[\hat{\mathrm{E}}(y|\mathbf{z})]/\partial \log(z_1) = \partial[\hat{\theta}_1 + \hat{\theta}_2 \log(z_1) + \hat{\theta}_3 z_2]/\partial \log(z_1) = \hat{\theta}_2$.

b. This is approximated by $100 \cdot \partial \log[\hat{\mathrm{E}}(y|\mathbf{z})]/\partial z_2 = 100 \cdot \hat{\theta}_3$.

c. Since $\partial \hat{\mathrm{E}}(y|\mathbf{z})/\partial z_2 = \exp[\hat{\theta}_1 + \hat{\theta}_2 \log(z_1) + \hat{\theta}_3 z_2 + \hat{\theta}_4 z_2^2] \cdot (\hat{\theta}_3 + 2\hat{\theta}_4 z_2)$, the estimated turning point is $\hat{z}_2^* = \hat{\theta}_3/(-2\hat{\theta}_4)$. This is a consistent estimator of $z_2^* \equiv \theta_3/(-2\theta_4)$.

d. Since $\nabla_{\boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}) = \exp(\mathbf{x}_1\boldsymbol{\theta}_1 + \mathbf{x}_2\boldsymbol{\theta}_2)\mathbf{x}$, the gradient of the mean function evaluated under the null is

$$\nabla_{\boldsymbol{\theta}} \tilde{m}_i = \exp(\mathbf{x}_{i1}\tilde{\boldsymbol{\theta}}_1)\mathbf{x}_i \equiv \tilde{m}_i \mathbf{x}_i,$$

where $\tilde{\boldsymbol{\theta}}_1$ is the restricted NLS estimator. From regression (12.72), we can compute the usual LM statistic as NR_u^2 from the regression $\tilde{u}_i$ on $\tilde{m}_i\mathbf{x}_{i1}, \tilde{m}_i\mathbf{x}_{i2}$, $i = 1, \ldots, N$, where $\tilde{u}_i = y_i - \tilde{m}_i$. For the robust test, we first regress $\tilde{m}_i\mathbf{x}_{i2}$ on $\tilde{m}_i\mathbf{x}_{i1}$ and obtain the $1 \times K_2$ residuals, $\tilde{\mathbf{r}}_i$. Then we compute the statistic as in regression (12.75).

12.4. a. Write the objective function as $(1/2)\sum_{i=1}^N [y_i - m(\mathbf{x}_i, \boldsymbol{\theta})]^2/h(\mathbf{x}_i, \hat{\boldsymbol{\gamma}})$. The objective function, for any value of $\boldsymbol{\gamma}$, is

b. We know from Chapter 6 that if we were using OLS in both stages then we could ignore the first-stage estimation of $\boldsymbol{\pi}_2$ under $H_0 : \rho_1 = 0$. That seems very likely the case here, too, but it does not follow from the results presented in the text (which assume smooth objective functions with nonsingular expected Hessians).

d. As mentioned in part c, an analytical calculation requires an extended set of tools, such as those in Newey and McFadden (1994). A computationally intensive solution is to bootstrap the two-step estimation method (being sure to recompute $\hat{\boldsymbol{\pi}}_2$ with every bootstrap sample in order to account for its sampling distribution).

12.17. a. We use a mean value expansion, similar to the delta method from Chapter 3 but now allowing for the randomness of $\mathbf{w}_i$. By a mean value expansion, we can write

$$N^{-1/2}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i,\hat{\boldsymbol{\theta}}) = N^{-1/2}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i,\boldsymbol{\theta}_o) + \left(N^{-1}\sum_{i=1}^{N}\ddot{\mathbf{G}}_i\right)\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o),$$

where $\ddot{\mathbf{G}}_i$ is the $M \times P$ Jacobian of $\mathbf{g}(\mathbf{w}_i,\boldsymbol{\theta})$ evaluated at mean values between $\boldsymbol{\theta}_o$ and $\hat{\boldsymbol{\theta}}$. Now, because $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o) \overset{a}{\sim} \text{Normal}(\mathbf{0},\mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1})$, it follows that $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o) = O_p(1)$. Further, by Lemma 12.1, $N^{-1}\sum_{i=1}^{N}\ddot{\mathbf{G}}_i \xrightarrow{p} \mathrm{E}[\nabla_{\boldsymbol{\theta}}\mathbf{g}(\mathbf{w},\boldsymbol{\theta}_o)] \equiv \mathbf{G}_o$ (the mean values all converge in probability to $\boldsymbol{\theta}_o$). Therefore,

$$\left(N^{-1}\sum_{i=1}^{N}\ddot{\mathbf{G}}_i\right)\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o) = \mathbf{G}_o\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o) + o_p(1),$$

and so

$$N^{-1/2}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i,\hat{\boldsymbol{\theta}}) = N^{-1/2}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i,\boldsymbol{\theta}_o) + \mathbf{G}_o\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o) + o_p(1).$$

Because $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o) = -N^{-1/2}\sum_{i=1}^{N}\mathbf{A}_o^{-1}\mathbf{s}_i(\boldsymbol{\theta}_o) = o_p(1)$, we can write

$$\sqrt{N}\hat{\boldsymbol{\delta}} = N^{-1/2}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i,\hat{\boldsymbol{\theta}}) = N^{-1/2}\sum_{i=1}^{N}[g(\mathbf{w}_i,\boldsymbol{\theta}_o) - \mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i(\boldsymbol{\theta}_o)] + o_p(1)$$

or, subtracting $\sqrt{N}\boldsymbol{\delta}_o$ from both sides,

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_o) = N^{-1/2}\sum_{i=1}^{N}[\mathbf{g}(\mathbf{w}_i,\boldsymbol{\theta}_o) - \boldsymbol{\delta}_o - \mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i(\boldsymbol{\theta}_o)] + o_p(1).$$

Now

$$\begin{aligned}\mathrm{E}[\mathbf{g}(\mathbf{w}_i,\boldsymbol{\theta}_o) - \boldsymbol{\delta}_o - \mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i(\boldsymbol{\theta}_o)] &= \mathrm{E}[\mathbf{g}(\mathbf{w}_i,\boldsymbol{\theta}_o)] - \boldsymbol{\delta}_o - \mathbf{G}_o\mathbf{A}_o^{-1}\mathrm{E}[\mathbf{s}_i(\boldsymbol{\theta}_o)]\\ &= \boldsymbol{\delta}_o - \boldsymbol{\delta}_o = \mathbf{0}.\end{aligned}$$

Therefore, by the CLT for i.i.d. sequences,

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_o) \overset{a}{\sim} \mathrm{Normal}(0,\mathbf{D}_o)$$

where

$$\mathbf{D}_o = \mathrm{Var}(\mathbf{g}_i - \boldsymbol{\delta}_o - \mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i),$$

where hopefully the shorthand is clear. This differs from the usual delta method result because the randomness in $\mathbf{g}_i = \mathbf{g}_i(\boldsymbol{\theta}_o)$ must be accounted for.

b. We assume we have $\hat{\mathbf{A}}$ consistent for $\mathbf{A}_o$. By the usual arguments, $\hat{\mathbf{G}} = N^{-1}\sum_{i=1}^{N}\nabla_{\boldsymbol{\theta}}\mathbf{g}(\mathbf{w}_i,\hat{\boldsymbol{\theta}})$ is consistent for $\mathbf{G}_o$. Then

$$\hat{\mathbf{D}} = N^{-1}\sum_{i=1}^{N}(\hat{\mathbf{g}}_i - \hat{\boldsymbol{\delta}} - \hat{\mathbf{G}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{s}}_i)(\hat{\mathbf{g}}_i - \hat{\boldsymbol{\delta}} - \hat{\mathbf{G}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{s}}_i)'$$

is consistent for $\mathbf{D}_o$, where the "^" denotes evaluation at $\hat{\boldsymbol{\theta}}$.

c. Using the shorthand notation, if $\mathrm{E}(\mathbf{s}_i|\mathbf{x}_i) = \mathbf{0}$ then $\mathbf{g}_i$ is uncorrelated with $\mathbf{s}_i$ because the premise of the problem is that $\mathbf{g}_i$ is a function of $\mathbf{x}_i$. Thefore, $(\mathbf{g}_i - \boldsymbol{\delta}_o)$ is uncorrelated with $\mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i$, which means

$$
\begin{aligned}
\mathbf{D}_o &= \mathrm{Var}(\mathbf{g}_i - \boldsymbol{\delta}_o - \mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i) \\
&= \mathrm{Var}(\mathbf{g}_i - \boldsymbol{\delta}_o) + \mathrm{Var}(\mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{s}_i) \\
&= \mathrm{Var}(\mathbf{g}_i) + \mathbf{G}_o\mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}\mathbf{G}_o' \\
&= \mathrm{Var}(\mathbf{g}_i) + \mathbf{G}_o[\mathrm{Avar}\,\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)]\mathbf{G}_o',
\end{aligned}
$$

which is what we wanted to show.