The proof in Errors of Prediction in Multiple Regression with Stochastic Regressor Variables - D. Kerridge and Sample size and the accuracy predictions made from multiple regression equations - Richard Sawyer.

Consider p-dimensional vectors $x_1, ..., x_n$ are given as a random sample from multivariate normal population $N(\mu, \Sigma)$, corresponding to observations $y_1, y_2, ..., y_n$. We assume that for each y_i, we have

$$y_i = \beta + X_i'A + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

β is the constant coefficient and A is the coefficient matrix. Let $\bar{x}$ be the mean over random sample $i = 1, 2, ..., n$, and let

$$S = \sum_{i,j}(x_i - \bar{x})(x_i - \bar{x})'$$

Let $(y^*, x^{*'})$ be an additional independent observation and y^* to be predicted by

$$\hat{y} = \beta + x^{*'}\widehat{A} = \bar{y} - \bar{x}'\widehat{A} + x^{*'}\widehat{A} = \bar{y} + (x^* - \bar{x})'\widehat{A}$$

where A is estimated by least square estimation

$$\widehat{A} = S^{-1}\sum_i y_i(x_i - \bar{x})$$

The prediction error is

$$
\begin{aligned}
y^* - \hat{y} &= \beta + X^{*'}A + \epsilon_* - \bar{y} - (x^* - \bar{x})'\widehat{A} \\
&= (x^* - \bar{x})'(A - \widehat{A}) + \epsilon^* + \bar{\epsilon}
\end{aligned}
\tag{1}
$$

We argue that the conditional distribution is $A - \widehat{A}|x_1, ..., x_n, x^* \sim N(0, \sigma^2 S^{-1})$. Thus the unconditional distribution is like (the exact distribution of $y^* - \hat{y}$ cannot be obtained)

$$y^* - \hat{y}|x_1, ..., x_n, x^* \sim N(0, [(x^* - \bar{x})'S^{-1}(x^* - \bar{x}) + 1 + \frac{1}{n}]\sigma^2).$$

But $(1 + \frac{1}{n})^{-1}(n-1)(x^* - \bar{x})'S^{-1}(x^* - \bar{x})$ is distributed like $Hotelling - T^2(p, n-1) = (n-1)\frac{p}{n-p}F_{p,n-p}$. Denote $T^2 = [x^* - \bar{x}]'S^{-1}[x^* - \bar{x}]$, the expactation of PMSE is

$$
\begin{aligned}
E[(y^* - \hat{y})^2] &= E\{E[(y^* - \hat{y})^2|x_1, ..., x_n, x^*\} \\
&= E\{E[z^2\sigma^2(T^2 + 1 + \frac{1}{n})|T^2]\} \\
&= \sigma^2 E[(T^2 + 1 + \frac{1}{n})] \\
&= \sigma^2[(1 + \frac{1}{n})\frac{k}{n-p}E(F_{p,n-p}) + 1 + \frac{1}{n}] \\
&= \sigma^2[(1 + \frac{1}{n})(\frac{k}{n-p}\frac{n-p}{n-p-2} + 1)] \\
&= \sigma^2(1 + \frac{1}{n})(\frac{n-2}{n-p-2}) \\
&= \sigma^2\frac{(n+1)(n-2)}{n(n-p-2)}
\end{aligned}
\tag{2}
$$

Proporsition: Let M be a positive integer. Then

$$E[(y^* - \hat{y})^{2M-1}] = 0$$

when $2M \leq n - p$;

$$E[(y^* - \hat{y})^{2M}] = \frac{\sigma^{2M} \frac{(2M)!}{M!} (\frac{n+1}{2n})^M \Pi_{j=1}^M (n - 2j)}{\prod_{j=1}^M (n - p - 2j)}$$

when $2M \leq n - p - 1$.

Proof: The conditional distribution of $y^* - \hat{y}$ given $\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}$ is

$$N(0, \sigma^2(1 + \frac{1}{n})[1 + (\frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}))])$$

by KERRIDGE(1967).

The unconditional distribution of $y^* - \hat{y}$ is asymptotically normal with mean 0 and variance σ^2 By symmetry of normal distribution, the odd moments in the unconditional distribution of $y^* - \hat{y}$ are 0. The unconditional even moments,

$$E[(y^* - \hat{y})^{2M}] = E\{E[(y^* - \hat{y})^{2M} | \boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}]\}$$

By properties of the normal distribution, for any non-negative integer M, $E[(x - \mu)^{2M}] = \sigma^p (p-1)!!$,

$$
\begin{aligned}
& E\{E[(y^* - \hat{y})^{2M} | \boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}]\} \\
=& E\{[\sigma^2(1 + \frac{1}{n})(1 + \frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*})]^M (2M - 1)!!\} \\
=& \sigma^{2M}(1 + \frac{1}{n})^M (2M - 1)!! E[(1 + \frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}))^M] \\
=& \sigma^{2M}(1 + \frac{1}{n})^M \frac{(2M - 1)!}{2^{M-1}(M - 1)!} E[(1 + \frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}))^M] \\
=& \sigma^{2M}(\frac{n+1}{2n})^M \frac{2(2M - 1)!}{(M - 1)!} E[(1 + \frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}))^M] \\
=& \sigma^{2M}(\frac{n+1}{2n})^M \frac{2M(2M - 1)...M(M - 1)...1}{M(M - 1)...1} E[(1 + \frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}))^M] \\
=& \sigma^{2M}(\frac{n+1}{2n})^M \frac{(2M)!}{M!} E[(1 + \frac{p}{n-p} F(\boldsymbol{x_1}, ..., \boldsymbol{x_n}, \boldsymbol{x^*}))^M]
\end{aligned}
\tag{3}
$$

By binomial theorem, the latter expected value is equal to

$$
\sum_{j=0}^{M} \binom{M}{j} \left(\frac{p}{n-p}\right)^j E(F^j)
$$

$$
= \sum_{j=0}^{M} \binom{M}{j} \frac{\Gamma(\frac{p}{2}+j)\Gamma(\frac{n-p}{2}-j)}{\Gamma(\frac{p}{2})\Gamma(\frac{n-p}{2})}
$$

$$
= \sum_{j=0}^{M} \binom{M}{j} \frac{(\frac{p}{2}+j-1)!(\frac{n-p}{2}-j-1)!}{(\frac{p}{2}-1)!(\frac{n-p}{2}-1)!}
$$

$$
= \sum_{j=0}^{M} \binom{M}{j} \frac{(\frac{p}{2}+j-1)...(\frac{p}{2}+j-j)(\frac{p}{2}-1)...1}{(\frac{p}{2}-1)(\frac{p}{2}-2)...1} \frac{(\frac{n-p}{2}-j-1)(\frac{n-p}{2}-j-2)...1}{(\frac{n-p}{2}-1)...(\frac{n-p}{2}-j)(\frac{n-p}{2}-j-1)...1} \quad (4)
$$

$$
= \sum_{j=0}^{M} \binom{M}{j} \frac{\prod_{k=1}^{j}(\frac{p}{2}+k-1)}{\prod_{k=1}^{j}(\frac{n-p}{2}-k)}
$$

$$
= \sum_{j=0}^{M} \binom{M}{j} \frac{\prod_{k=1}^{j} p+2(k-1)}{\prod_{k=1}^{j}(n-p-2k)}
$$

$$
= 1 + \sum_{j=1}^{M} \binom{M}{j} \frac{\prod_{k=1}^{j} p+2(k-1)}{\prod_{k=1}^{j}(n-p-2k)}
$$

Take $M = 1$, it follows that

$$
MSE = E[(y^* - \hat{y})^2] = \sigma^2 \frac{(n+1)(n-2)}{n(n-p-2)}
$$

same as Kerridge paper.