Appendix 1
Methodology

We assume that the linear parameters are of the form
\[
\beta_p = \beta + \gamma_p \quad \text{where} \quad \gamma_p \sim N(0, \Sigma)
\]  
(A.1)

Thus,
\[
\beta_p X_{jt} = \beta X_{jt} + \gamma_p X_{jt}
\]  
(A.2)

We rewrite the stochastic combination of mean efficacy terms as follows
\[
\bar{Q}_{pjt} = E[\bar{Q}_{pjt}] + (\bar{Q}_{pjt} - E[\bar{Q}_{pjt}])
\]  
(A.3)

Let
\[
\mu_{pjt} = \delta_{jt} + \lambda_{pjt}
\]  
(A.4)

where
\[
\delta_{jt} = E[\bar{Q}_{pjt}] + \beta X_{jt} + \xi_{jt}
\]  
(A.5)

\[
\lambda_{pjt} = (\bar{Q}_{pjt} - E[\bar{Q}_{pjt}]) + \gamma_p X_{jt}
\]  
(A.6)

Thus,
\[
s_{jt} = \int \frac{\exp(\delta_{jt} + \lambda_{pjt})}{1 + \sum_{k=1} \exp(\delta_{kj} + \lambda_{pkt})} d\Phi(\lambda)
\]  
(A.7)

where \( \Phi(\cdot) \) is the distribution of \( \lambda \).

The first step in the methodology is to obtain the mean utility vector, \( \delta \) that equates predicted shares to the observed shares. We use the contraction mapping developed by BLP(1995).

In finding the predicted share, we need to integrate out the distributions of efficacy and heterogeneity. Since this integration is not analytically tractable, we do this by simulation using the method described by Pakes and Pollard (1989). We generate a set of draws for this distribution, taking into account the serially correlated nature of the efficacy
distributions. We then find the simulation equivalent of the predicted share by doing a Monte-Carlo integration over these draws.

We note that the expression for $\delta_{jt}$ in equation (A.5) is linear in the price coefficient. Thus, we can use standard instrumental variables methods to consistently estimate the parameters that enter this linear expression. However, this is not directly possible yet, since the efficacy term also enters this linear expression.

We note that given a set of guesses for the primitive learning parameters

- $Q$ the initial prior mean efficacy
- $\{Q_j\}_{j=1}^J$' the set of true efficacies of the $J$ drugs
- $\sigma^2_\nu$ the feedback (experienced efficacy) signal variance
- $\sigma^2_\omega$ the detailing signal variance, and

and given the set of serially correlated draws for $\lambda_{pjt}$, the simulation equivalent of the term $E[\tilde{Q}_{pjt}] (= \tilde{Q}_{jt},$ say), can be evaluated.

Thus, we can regress the vector $\{\delta_{jt} - \tilde{Q}_{jt}\}$ on the vector $\{X_{jt}\}$ and use linear instrumental variables methods to obtain consistent estimates of $\beta$ and the vector of residuals. We note that once the mean utility vector has been obtained by the contraction mapping procedure, there is no serial correlation in the error term. Thus we can apply standard GMM methods to obtain the primitive learning parameters. The moment conditions are $E[Z'\xi] = 0$ where $Z$ is a matrix of exogenous instruments. Our estimation methodology comprises the following steps:

1. Make guesses for the set of learning parameters $(Q, \{Q_j\}_{j=1}^J, \sigma^2_\nu, \sigma^2_\omega)$ and the variance covariance matrix (heterogeneity parameters) $\Sigma$.
2. Given these set of parameters, and for each set of draws for the detailing and feedback signals for all the time periods, compute the value of $\tilde{Q}_{jt}$ by simulation.
3. Integrate out the distribution of $\lambda_{pjt}$ using the set of draws for signals and those for heterogeneity.
4. By contraction mapping, find the mean utility vector $\delta$.

5. Estimate the linear parameters (vector $\beta$) by an IV regression of $\{\delta_{jt} - \hat{Q}_{jt}\}$ on the vector $\{X_{jt}\}$.

6. Compute the residuals of the above regression and find the GMM criterion function.

7. Repeat steps 2 to 7 and find the vector of learning and heterogeneity parameters that minimize this criterion function.